MODELOS DE REGRESSÃO E APLICACÕES

Gilberto A. Paula
Instituto de Matemática e Estatística
Universidade de São Paulo
e-mail:giapaula@ime.usp.br

Prefácio

A área de modelagem estatística de regressão recebeu um grande impulso desde a criação dos modelos lineares generalizados (MLGs) no início da década de 1970. O crescente interesse pela área motivou a realização de vários encontros informais no início dos anos 1980, a maioria deles na Inglaterra, até que em 1986 foi realizado na cidade de Innsbruck na Áustria o "1st International Workshop on Statistical Modelling"(1st IWSM). Esse encontro tem sido realizado anualmente no continente Europeu sendo que o último (39th IWSM) aconteceu em julho de 2025 em Limerick, Irlanda. O 40th IWSM será realizado em junho-julho de 2024 em Oslo, Noruega. No Brasil a área começou efetivamente a se desenvolver a partir de meados da década de 1980 e em particular após a 1ª Escola de Modelos de Regressão (1EMR) realizada na Universidade de São Paulo em 1989. As demais escolas ocorreram desde então a cada dois anos sendo que a 18EMR foi realizada em novembro de 2023 em Belém do Pará e a 19EMR será realizada em outubro de 2025 em João Pessoa, Paraíba.

Este texto começou a ser desenvolvido a partir de 1994 quando a disciplina `Modelos Lineares Generalizados` passou a ser ministrada regularmente no programa de pós-graduação em Probabilidade e Estatística do IME-USP. Mais de 450 alunos cursaram essa disciplina desde 1994. O texto é direcionado para alunos que tenham cursado um semestre de inferência estatística e que tenham conhecimentos de regressão linear. Portanto, trata-

se de um segundo curso de modelagem estatística de regressão com um enfoque inferencial básico e várias aplicações. O texto tem sido também utilizado nas disciplinas `Análise de Dados Categorizados` e `Modelos de Regressão II` (antiga `Tópicos de Regressão`) ministradas aos alunos do Bacharelado em Estatística do IME-USP.

O texto apresenta no Capítulo 1 uma revisão de regressão linear múltipla com algumas extensões. Inicialmente é discutido a solução de mínimos quadrados com interpretação geométrica, métodos de estimação e teste de hipóteses. Procedimentos de diagnóstico, conceito de interação, comparações múltiplas, regressão ponderada, multicolinearidade e seleção de modelos são discutidos em seguida com 5 exemplos ilustrativos apresentados. Tópicos especiais, tais como regressão por partes, métodos robustos, reamostragem, regressão não linesar e regressão com erros autoregressivos são também discutidos com mais 5 exemplos ilustrativos. Finalmente, conexão entre máxima verossimilhança e mínimos quadrados é apresetnada.

No Capítulo 2 a classe dos modelos lineares generalizados (MLGs) é descrita com alguns conceitos básicos. Em seguida são discutidos a estimação dos parâmetros, propriedades assintóticas dos estimadores de máxima verossimilhança e a aplicação de alguns testes estatísticos mais conhecidos. Uma revisão de métodos de diagnóstico é apresentada na sequência com extensões dos procedimentos apresentados no Capítulo 1 para a classe dos MLGs. Discute-se também a seleção de modelos e o capítulo é concluído com a análise de 5 conjuntos de dados através de MLGs apropriados.

O Capítulo 3 é dedicado aos modelos com resposta gama e resposta normal inversa para a análise de dados assimétricos positivos. Inicialmente aborda-se os modelos com resposta gama sendo apresentados alguns resultados inferenciais e técnicas de diagnóstico. Três conjuntos de dados são analisados. Em seguida, modelos usualmente aplicados em Econometria são discutidos e um exemplo é apresentado. Modelos com resposta normal in-

versa são discutidos e 2 conjuntos de dados são analisados. No final do capítulo são apresentados os MLGs duplos, em que a média e a dispersão são ajustados conjuntamente. Apresenta-se o processo de estimação conjunto, alguns procedimentos de diagnóstico e um exemplo ilustrativo.

No Capítulo 4 discute-se modelos para a análise de dados binários, com ênfase para os modelos logísticos lineares. Inicialmente uma revisão de procedimentos tradicionais para a análise de tabelas de contigência 2×2 é apresentada. Duas aplicações são descritas nesta primeira parte do capítulo. Em seguida o modelo logístico linear é apresentado. Alguns procedimentos são revisados, tais como seleção de modelos, análise de dados retrospectivos, qualidade do ajuste e técnicas de diagnóstico. Quatro conjuntos de dados são analisados. Discute-se no final do capítulo modelos de dose-resposta, sobredispersão e modelos logísticos aplicados na análise de dados emparelhados e mais cinco aplicações são apresentadas.

No Capítulo 5 são discutidos alguns modelos para a análise de dados de contagem, com destaque para modelos com resposta de Poisson e modelos com resposta binomial negativa. Inicialmente apresenta-se uma revisão de metodologias tradicionais para a análise da tabelas de contingência do tipo 2×2 com dados de contagem. Duas aplicações são apresentadas. Em seguida discute-se modelos de Poisson para a análise de dados de seguimento e modelos log-lineares de Poisson. Dois exemplos são apresentados. Na sequência são derivados modelos com resposta binomial negativa para a análise de dados de contagem com sobredispersão. Um processo iterativo para a estimação dos parâmetros, resultados assintóticos e metodologias de diagnóstico são apresentados, bem como 2 aplicações. Modelos log-lineares com resposta de Poisson são comparados com modelos log-lineares com resposta multinomial, sendo 3 conjuntos de dados analisados. Finalmente, uma breve resenha dos modelos com excesso de zeros é apresentada, em particular, os modelos ajustados em zero e os modelos inflacionados de zeros.

O Capítulo 6 é dedicado aos modelos de quase-verossimilhança e às equações de estimação generalizadas. Inicia-se o capítulo com a introdução do conceito de quase-verossimilhança. Em seguida são apresentados os modelos de quase-verossimilhança para respostas independentes juntamente com o processo de estimação, alguns resultados assintóticos e técnicas de diagnóstico. Três aplicações são apresentadas. Na sequência deriva-se as equações de estimação generalizadas para a análise de dados correlacionados não gaussianos, tais como dados agrupados, medidas repetidas e dados longitudinais. Apresenta-se o processo de estimação, alguns resultados assintóticos e metodologias de diagnóstico. Esse subtópico é ilustrado com 3 aplicações.

No Apêndice A são descritos os conjuntos de dados usados nas aplicações e nos exercícios propostos e no Apêndice B são descritos alguns códigos em linguagem R. No final de cada capítulo são propostos exercícios teóricos e aplicados num total de 150 exercícios conjuntamente com códigos e subrotinas em `R`. Procura-se diversificar as aplicações com conjuntos de dados das diversas áreas do conhecimento, tais como Agricultura, Biologia, Ciências Atuariais, Ciências Sociais, Economia, Engenharia, Geografia, Medicina, Nutrição, Pecuária, Pesca e Odontologia. Alguns conjuntos de dados são oriundos de trabalhos desenvolvidos no Centro de Estatística Aplicada (CEA) do IME-USP. Página na `Web` onde estão disponíveis informações sobre este texto: `http://www.ime.usp.br/` ∼`giapaula/textoregressao.htm`.

Finalizando, fica um agradecimento aos alunos que cursaram as disciplinas e contribuiram com suas sugestões para o aprimoramento dos primeiros manuscritos.

São Paulo, agosto de 2025

Gilberto A. Paula
e-mail:giapaula@ime.usp.br

Sumário

Capítulo 1

Regressão Linear Múltipla

1.1 Introdução

O principal objetivo deste capítulo é apresentar uma síntese dos principais tópicos relacionados com regressão linear múltipla, tais como estimação por mínimos quadrados e máxima verossimilhança, procedimentos inferenciais e de teste de hipóteses, além de métodos de diagnóstico, conceito de interação, comparação de médias, regressão ponderada, multicolinearidade, seleção de modelos, regressão por partes e métodos robustos e de reamostragem com extensões para regressão não linear. Exemplos ilustrativos são apresentados ao longo do capítulo e vários exercícios teóricos e aplicados são propostos no final do capítulo. Uma abordagem mais completa pode ser encontrada, por exemplo, no livro de Montgomery et al.(2021).

1.2 Modelo de Regressão Linear Múltipla

Denote por $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$ uma amostra aleatória de tamanho n de uma determinada população, em que $y_1, \ldots, y_n$ representam os valores observados da variável resposta (assumida contínua), enquanto $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ denotam valores observados de variáveis explicativas, para $i = 1, \ldots, n$. O

principal objetivo da regressão linear múltipla é tentar explicar o valor esperado da variável resposta dados os valores das variáveis explicativas. A formulação mais usual é a seguinte:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i, \tag{1.1}$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$. Muitas vezes tem-se um intercepto em (1.1), sendo nesse caso assumido que $x_{i1} = 1 \ \forall i$.

A suposição de normalidade para os erros pode ser relaxada para amostras grandes, contudo para amostras pequenas e moderadas essa suposição é crucial para fazer inferência. De (1.1) segue que $Y_i|\mathbf{x}_i \overset{\text{ind}}{\sim} \text{N}(\mu_i, \sigma^2)$ com $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ e $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, para $i = 1, \ldots, n$.

Em forma matricial o modelo (1.1) fica expresso na forma

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.2}$$

em que $\mathbf{y} = (y_1, \ldots, y_n)^\top$, $\mathbf{X}$ é a matriz modelo de dimensão $n \times p$ dada por

$$\mathbf{X} = \begin{bmatrix} x_{i1} & x_{12} & \ldots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix},$$

$\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ com $\boldsymbol{\epsilon} \sim \text{N}_n(0, \sigma^2 \mathbf{I}_n)$ e $\mathbf{I}_n$ a matriz identidade de ordem n.

1.3 Solução de Mínimos Quadrados

A estimativa de mínimos quadrados de $\boldsymbol{\beta}$ é obtida minimizando a função objetivo $S(\boldsymbol{\beta})$ que corresponde a minimizar a soma dos quadrados dos erros

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

A derivada parcial de $S(\boldsymbol{\beta})$ com relação a β_j fica dada por

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \beta_j} = -2\sum_{i=1}^{n} x_{ij}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}),$$

para $j = 1, \ldots, p$. Assim, a derivada de $S(\boldsymbol{\beta})$ com relação a $\boldsymbol{\beta}$ é um vetor de dimensão $p \times 1$ expresso na forma

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

A estimativa de mínimos quadrados $\widehat{\boldsymbol{\beta}}$ é obtida igualando-se a primeira derivada a zero

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0} \Rightarrow -2\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{0}.$$

Se $\mathbf{X}$ é uma matriz de posto coluna completo então tem-se uma solução única

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}.$$

Na Figura 1.1 é apresentada uma representação geométrica da solução de mínimos quadrados, em que $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{y}$ corresponde à projeção ortogonal de $\mathbf{y}$ através do projetor linear $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$, no subespaço gerado pelas colunas da matriz $\mathbf{X}$, denotado por $\mathrm{C}(\mathbf{X})$. Por outro lado, $r = \mathbf{y} - \widehat{\mathbf{y}} = (\mathbf{I}_n - \mathbf{H})\mathbf{y}$ definido como vetor de resíduos ordinários, corresponde à projeção ortogonal de $\mathbf{y}$ através do projetor linear $(\mathbf{I}_n - \mathbf{H})$, no subespaço complementar $\mathrm{C}^c(\mathbf{X})$, denominado ortocomplemento de $\mathrm{C}(\mathbf{X})$.

É preciso verificar se a raiz da primeira derivada é de fato um ponto de mínimo da superfície formada por $(S(\boldsymbol{\beta}), \boldsymbol{\beta}^\top)^\top$. Deriva-se então novamente $S(\boldsymbol{\beta})$ com relação a β_ℓ, obtendo-se

$$\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_\ell} = 2\sum_{i=1}^{n} x_{ij}x_{i\ell},$$

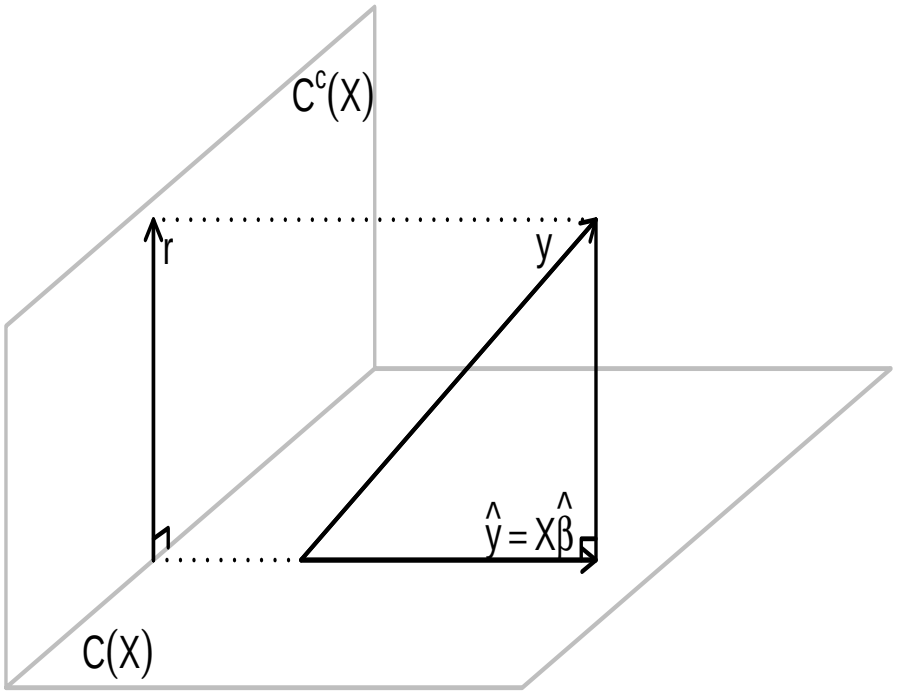

Figura 1.1: Representação geométrica da solução de mínimos quadrados referente ao modelo de regressão linear múltipla (1.2), em que $\mathbf{r} = \mathbf{y} - \widehat{\mathbf{y}}$ é o resíduo ordinário e $\mathrm{C}(\mathbf{X})$ denota o subespaço gerado pelas colunas da matriz $\mathbf{X}$ e $\mathrm{C}^c(\mathbf{X})$ o ortocomplemento.

para $j, \ell = 1, \ldots, p$. Assim, a matriz de segundas derivadas de $S(\boldsymbol{\beta})$ com relação a $\boldsymbol{\beta}$ tem dimensão $p \times p$ e fica expressa na forma

$$\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = 2\mathbf{X}^\top \mathbf{X}.$$

Como é assumido que $\mathbf{X}$ tem posto coluna completo então $\mathbf{X}^\top \mathbf{X}$ é uma matriz

positiva definida, logo $S(\boldsymbol{\beta})$ é uma superfície convexa e $\widehat{\boldsymbol{\beta}}$ é ponto de mínimo.

Resumindo, tem-se que $\mathbf{Y}|\mathbf{X} \sim \mathrm{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$ e como consequências $\mathrm{E}(\mathbf{Y}|\mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$ e $\mathrm{Var}(\mathbf{Y}|\mathbf{X}) = \sigma^2\mathbf{I}_n$, em que $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$. Seguem as seguintes propriedades do estimador de mínimos quadrados:

$$\mathrm{E}(\widehat{\boldsymbol{\beta}}) = \mathrm{E}\{(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}\} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathrm{E}(\mathbf{Y}|\mathbf{X}) = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}.$$

Ou seja, $\widehat{\boldsymbol{\beta}}$ é um estimador não tendencioso de $\boldsymbol{\beta}$. A matriz de variância-covariância de $\widehat{\boldsymbol{\beta}}$ fica dada por

$$\begin{aligned}\mathrm{Var}(\widehat{\boldsymbol{\beta}}) &= \mathrm{Var}\{(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}\} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathrm{Var}(\mathbf{Y}|\mathbf{X})\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}.\end{aligned}$$

Logo, $\widehat{\boldsymbol{\beta}} \sim \mathrm{N}_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1})$ e conforme mostrado em Montgomery et al. (2021, Apêndice C.4) $\widehat{\boldsymbol{\beta}}$ tem a menor variância entre todos os estimadores lineares não viesados de $\boldsymbol{\beta}$.

Pelo Teorema de Pitágoras aplicado ao triângulo retângulo da Figura 1.1, tem-se que

$$\begin{aligned}|| \mathbf{y} ||^2 &= || \hat{\mathbf{y}} ||^2 + || \mathbf{y} - \hat{\mathbf{y}} ||^2 \\ \sum_{i=1}^{n} y_i^2 &= \sum_{i=1}^{n} \widehat{y}_i^2 + \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2,\end{aligned}$$

em que $|| \mathbf{v} ||= \sqrt{v_1^2 + \cdots + v_n^2}$ denota norma ou comprimento do vetor $\mathbf{v} = (v_1, \ldots, v_n)^\top$. Se o modelo tem intercepto segue da solução de mínimos quadrados $\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{0}$ que $\sum_{i=1}^{n} r_i = 0$. Logo, obtém-se a decomposição de somas de quadrados

$$\mathrm{SQT} = \mathrm{SQReg} + \mathrm{SQRes},$$

em que $\mathrm{SQT} = \sum_{i=1}^{n}(y_i - \bar{y})^2$ é a soma de quadrados total, $\mathrm{SQReg} = \sum_{i=1}^{n}(y_i - \bar{y})^2$ é a soma de quadrados devido à regressão, enquanto $\mathrm{SQRes} =$

$\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$ é a soma de quadrados de resíduos. Uma maneira de avaliar a qualidade do ajuste é comparar SQReg com SQT através do coeficiente de determinação

$$\mathrm{R}^2 = \frac{\mathrm{SQReg}}{\mathrm{SQT}} = 1 - \frac{\mathrm{SQRes}}{\mathrm{SQT}},$$

em que $0 \leq \mathrm{R}^2 \leq 1$. Quanto mais próximo R^2 está de 1 melhor a qualidade do ajuste. Contudo, como o coeficiente de determinação cresce à medida que o número p de parâmetros aumenta, recomenda-se a utilização do coeficiente de determinação ajustado

$$\overline{\mathrm{R}}^2 = 1 - \frac{\mathrm{QMRes}}{\mathrm{QMT}},$$

em que $\mathrm{QMRes} = \frac{\mathrm{SQRes}}{n-p}$ e $\mathrm{QMT} = \frac{\mathrm{SQT}}{p-1}$ e $0 \leq \overline{\mathrm{R}}^2 \leq 1$. É possível estabelecer a seguinte relação:

$$\overline{\mathrm{R}}^2 = 1 - (1 - \mathrm{R}^2)\frac{(n-1)}{(n-p)}.$$

Portanto, segue que $\overline{\mathrm{R}}^2 \leq \mathrm{R}^2$.

1.3.1 Regressão Linear Simples

Considere agora o modelo de regressão linear simples definido por

$$y_i = \beta_1 + \beta_2 x_i + \epsilon_i,$$

em que $y_1, \ldots, y_n$ são valores observados da variável resposta, $x_1, \ldots, x_n$ são valores observados da variável explicativa X e $\epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, para $i = 1, \ldots, n$. A matriz modelo de dimensão $n \times 2$ fica dada por

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

Assim, obtém-se

$$\mathbf{X}^\top\mathbf{X} = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum x_i^2 \end{bmatrix} \quad \text{e} \quad \mathbf{X}^\top\mathbf{y} = (n\bar{y}, \sum x_i y_i)^\top.$$

em que $\bar{x} = \frac{\sum x_i}{n}$ e $\bar{y} = \frac{\sum y_i}{n}$. Logo,

$$(\mathbf{X}^\top\mathbf{X})^{-1} = \frac{1}{n\mathrm{S}_{xx}} \begin{bmatrix} \sum x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix},$$

em que $\mathrm{S}_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$. O estimador de mínimos quadrados fica dado por

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \bar{y} - \widehat{\beta}_2\bar{x} \\ \frac{\mathrm{S}_{xy}}{\mathrm{S}_{xx}} \end{bmatrix}$$

com $\mathrm{S}_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$. A matriz de variância-covariância assume a forma

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} = \frac{\sigma^2}{n\mathrm{S}_{xx}} \begin{bmatrix} \sum x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}.$$

Daí segue que $\mathrm{Var}(\widehat{\beta}_1) = \frac{\sigma^2 \sum x_i^2}{n\mathrm{S}_{xx}}$, $\mathrm{Var}(\widehat{\beta}_2) = \frac{\sigma^2}{\mathrm{S}_{xx}}$ e $\mathrm{Cov}(\widehat{\beta}_1, \widehat{\beta}_2) = -\frac{\sigma^2\bar{x}}{\mathrm{S}_{xx}}$.

Supondo que X é uma variável quantitativa contínua, o coeficiente de correlação linear amostral de Pearson entre X e Y é expresso na forma

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2\}^{\frac{1}{2}}},$$

em que $-1 \leq r \leq 1$. Aternativamente, tem-se que

$$r_{xy} = \frac{\mathrm{S}_{xy}}{\{\mathrm{S}_{xx}\mathrm{SQT}\}^{\frac{1}{2}}} = \frac{\mathrm{S}_{xy}}{\mathrm{S}_{xx}}\sqrt{\frac{\mathrm{S}_{xx}}{\mathrm{SQT}}} = \widehat{\beta}_2\sqrt{\frac{\mathrm{S}_{xx}}{\mathrm{SQT}}}.$$

Por outro lado, obtém-se

$$\widehat{y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 x_i = (\bar{y} - \bar{x}\widehat{\beta}_2) + \widehat{\beta}_2 x_i = \bar{y} + (x_i - \bar{x})\widehat{\beta}_2.$$

Logo $(\widehat{y}_i - \bar{y}) = (x_i - \bar{x})\widehat{\beta}_2$ e portanto $\sum_{i=1}^n (\widehat{y}_i - \bar{y})^2 = \widehat{\beta}_2^2 \sum_{i=1}^n (x_i - \bar{x})^2$. Então, segue que $\mathrm{SQReg} = \widehat{\beta}_2^2\mathrm{S}_{xx}$. E desde que $\widehat{\beta}_2 = \frac{\mathrm{S}_{xy}}{\mathrm{S}_{xx}}$ obtém-se

$$\mathrm{SQReg} = \widehat{\beta}_2\mathrm{S}_{xy} \rightarrow \widehat{\beta}_2 = \frac{\mathrm{SQReg}}{\mathrm{S}_{xy}}.$$

Finalmente, segue a relação

$$r_{xy}^2 = \widehat{\beta}_2^2 \frac{\mathrm{S}_{xx}}{\mathrm{SQT}} = \frac{\mathrm{S}_{xy}}{\mathrm{S}_{xx}} \frac{\mathrm{SQReg}}{\mathrm{S}_{xy}} \frac{\mathrm{S}_{xx}}{\mathrm{SQT}} = \frac{\mathrm{SQReg}}{\mathrm{SQT}} = \mathrm{R}^2.$$

Ou seja, o coeficiente de determinação R^2 coincide com o quadrado do coeficiente de correlação linear amostral de Pearson entre X e Y na regressão linear simples.

1.4 Teste de Hipóteses

Inicialmente, supor que o interesse é avaliar se os coeficientes da regressão são nulos, que corresponde a testar as hipóteses

$$\mathrm{H}_0 : \beta_2 = \cdots = \beta_p = 0 \ \text{ contra } \ \mathrm{H}_1 : \beta_j \neq 0,$$

para pelo menos algum $j = 2, \ldots, p$. A estatística F fica expressa na forma

$$\mathrm{F} = \frac{\mathrm{SQReg}/(p-1)}{\mathrm{SQRes}/(n-p)} = \frac{\mathrm{QMReg}}{\mathrm{QMRes}} \overset{\mathrm{H}_0}{\sim} \mathrm{F}_{(p-1),(n-p)}.$$

Para um nível de significância $0 < \alpha < 1$, rejeita-se H_0 se $\mathrm{F} > \mathrm{F}_{(1-\alpha),(p-1),(n-p)}$, em que $\mathrm{F}_{(1-\alpha),(p-1),(n-p)}$ denota o quantil $(1-\alpha)$ da distribuição F com $(p-1)$ e $(n-p)$ graus de liberdade. É usual construir a tabela de análise de variância (ANOVA), conforme descrito na Tabela 1.1.

Tabela 1.1: Descrição da tabela de Análise de Variância (ANOVA).

F. Variação	S.Quadrados	G.L.	Q. Médio	F
Regressão	SQReg	$p-1$	QMReg	$\frac{\mathrm{QMReg}}{\mathrm{QMRes}}$
Resíduos	SQRes	$n-p$	QMRes	
Total	SQT	$n-1$		

Denote $\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2 \mathbf{C}$, em que $\mathbf{C} = (\mathbf{X}^\top \mathbf{X})^{-1}$. Então, pode-se expressar as variâncias e covariâncias dos estimadores $\widehat{\beta}_1, \ldots, \widehat{\beta}_p$ nas formas $\mathrm{Var}(\widehat{\beta}_j) =$

$\sigma^2 \mathrm{C}_{jj}$ e $\mathrm{Cov}(\widehat{\beta}_j, \widehat{\beta}_\ell) = \sigma^2 \mathrm{C}_{j\ell}$, em que $\mathrm{C}_{j\ell}$ denota o elemento (j, ℓ) da matriz $\mathbf{C}$, para $j, \ell = 1, \ldots, p$. Supor então que o interesse é testar as hipóteses $\mathrm{H}_0 : \beta_j = 0$ contra $\mathrm{H}_1 : \beta_j \neq 0$, para algum $j = 1, \ldots, p$. A estatística t-Student fica expressa na forma

$$\mathrm{t} = \frac{\widehat{\beta}_j}{\widehat{\mathrm{EP}}(\widehat{\beta}_j)} \overset{\mathrm{H}_0}{\sim} \mathrm{t}_{(n-p)},$$

em que $\widehat{\mathrm{EP}}(\widehat{\beta}_j) = s\sqrt{\mathrm{C}_{jj}}$. Para um nível de significância $0 < \alpha < 1$, rejeita-se H_0 se $|\mathrm{t}| > \mathrm{t}_{(1-\alpha/2),(n-p)}$, em que $\mathrm{t}_{(1-\alpha/2),(n-p)}$ denota o quantil $(1 - \alpha/2)$ de uma distribuição t-Student com $(n - p)$ graus de liberdade. Em particular, pode-se mostrar que t^2 segue sob H_0 distribuição $\mathrm{F}_{1,(n-p)}$.

Generalizando, supor que o interesse agora é testar $\mathrm{H}_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{0}$ contra contra $\mathrm{H}_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{0}$, com pelo menos uma desigualdade estrita em H_1, em que $\mathbf{R}$ é uma matriz $r \times p$ com posto linha $r \leq p$. O acréscimo na soma de quadrados de resíduos devido à restrição $\mathbf{R}\boldsymbol{\beta} = \mathbf{0}$ (vide Montgomery et al., 2021, Cap. 3) é dado por

$$\mathrm{ASQ}(\mathbf{R}\boldsymbol{\beta} = \mathbf{0}) = (\mathbf{R}\widehat{\boldsymbol{\beta}})^\top \{\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top\}^{-1} \mathbf{R}\widehat{\boldsymbol{\beta}}.$$

Portanto, tem-se que

$$\mathrm{F} = \frac{\mathrm{ASQ}(\mathbf{R}\boldsymbol{\beta} = \mathbf{0})/r}{\mathrm{SQRes}/(n-p)} \overset{\mathrm{H}_0}{\sim} \mathrm{F}_{r,(n-p)}.$$

Logo, para um nível de significância $0 < \alpha < 1$, rejeita-se H_0 se $\mathrm{F} > \mathrm{F}_{(1-\alpha),r,(n-p)}$.

Um caso particular é considerar a regressão linear múltipla (1.2) com efeitos particionados

$$\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \tag{1.3}$$

em que $\mathbf{X}_1$ e $\mathbf{X}_2$ são matrizes de dimensões $n \times p_1$ e $n \times p_2$, respectivamente, enquanto $\boldsymbol{\beta}_1$ tem dimensão $p_1 \times 1$ e $\boldsymbol{\beta}_2$ tem dimensão $p_2 \times 1$. Logo,

$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$ e $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$. Supor que o interesse seja testar $\mathrm{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ contra $\mathrm{H}_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$, com pelo menos uma desigualdade estrita em H_1. A soma de quadrados de resíduos correspondente ao modelo (1.3) com p parâmetros será denotada por $\mathrm{SQRes}(\boldsymbol{\beta}) = \mathbf{y}^\top(\mathbf{I}_n - \mathbf{H})\mathbf{y}$, enquanto que a soma de quadrados de resíduos sob o modelo em H_0 com p_1 parâmetros será denotada por $\mathrm{SQRes}(\boldsymbol{\beta}|\boldsymbol{\beta}_2 = \mathbf{0}) = \mathbf{y}^\top(\mathbf{I}_n - \mathbf{H}_1)\mathbf{y}$, em que $\mathbf{H}_1 = \mathbf{X}_1(\mathbf{X}_1^\top\mathbf{X}_1)^{-1}\mathbf{X}_1^\top$. Note que $\mathrm{SQRes}(\boldsymbol{\beta}|\boldsymbol{\beta}_2 = \mathbf{0}) \geq \mathrm{SQRes}(\boldsymbol{\beta})$. Assim, o acréscimo na soma de quadrados de resíduos devido à restrição $\boldsymbol{\beta}_2 = \mathbf{0}$ pode ser expresso na forma

$$\mathrm{ASQ}(\boldsymbol{\beta}_2 = \mathbf{0}) = \mathrm{SQRes}(\boldsymbol{\beta}|\boldsymbol{\beta}_2 = \mathbf{0}) - \mathrm{SQRes}(\boldsymbol{\beta}) = \mathbf{y}^\top(\mathbf{H}_1 - \mathbf{H})\mathbf{y},$$

e consequentemente a estatística F para testar $\mathrm{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ contra $\mathrm{H}_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$ fica dada por

$$\mathrm{F} = \frac{\mathbf{y}^\top(\mathbf{H}_1 - \mathbf{H})\mathbf{y}/p_2}{\mathbf{y}^\top(\mathbf{I}_n - \mathbf{H})\mathbf{y}/(n-p)} \overset{\mathrm{H}_0}{\sim} \mathrm{F}_{p_2,(n-p)}.$$

Logo, para um nível de significância $0 < \alpha < 1$, rejeita-se H_0 se $\mathrm{F} > \mathrm{F}_{(1-\alpha),p_2,(n-p)}$.

1.5 Estimativa Intervalar

Um estimativa intervalar de coeficiente de confiança $(1-\alpha)$ para β_j fica dada por

$$[\widehat{\beta}_j \pm \mathrm{t}_{(1-\alpha/2),(n-p)}\widehat{\mathrm{EP}}(\widehat{\beta}_j)],$$

em que $j = 1, \ldots, p$. Como para n grande a t-Student se aproxima da normal, pode-se usar o quantil $(1 - \alpha/2)$ da $\mathrm{N}(0, 1)$ no lugar de $\mathrm{t}_{(1-\alpha/2),(n-p)}$.

É possível mostrar que

$$\frac{\mathrm{SQRes}}{\sigma^2} \overset{\text{modelo}}{\sim} \chi^2_{(n-p)}.$$

Logo, segue que $\mathrm{E}\big(\frac{\mathrm{SQRes}}{\sigma^2}\big) = (n-p)$ e portanto $s^2 = \frac{\mathrm{SQRes}}{(n-p)}$ é um estimador não tendencioso de σ^2. Após algumas manipulações com a distribuição $\chi^2_{(n-p)}$

tem-se que

$$P\left\{\frac{(n-p)s^2}{\chi^2_{(1-\alpha/2),(n-p)}} \leq \sigma^2 \leq \frac{(n-p)s^2}{\chi^2_{(\alpha/2),(n-p)}}\right\} = (1-\alpha),$$

em que $\chi^2_{(\alpha/2),(n-p)}$ e $\chi^2_{(1-\alpha/2),(n-p)}$ denotam, respectivamente, os quantis $\alpha/2$ e $(1-\alpha/2)$ da distribuição $\chi^2_{(n-p)}$. Assim, uma estimativa intervalar de coeficiente de confiança $(1-\alpha)$ para σ^2 fica dada por

$$\left[\frac{(n-p)s^2}{\chi^2_{(1-\alpha/2),(n-p)}}; \frac{(n-p)s^2}{\chi^2_{(\alpha/2),(n-p)}}\right].$$

Alternativamente, é possível encontrar uma estimativa intervalar de menor comprimento para σ^2 dada por

$$\left[\frac{(n-p)s^2}{a}; \frac{(n-p)s^2}{b}\right],$$

em que a e b são constantes tais que $a^2 g_{(n-p)}(a) = b^2 g_{(n-p)}(b)$ e $\int_a^b g_{(n-p)}(t)dt = (1-\alpha)$, com $g_{(n-p)}(t)$ denotando a função densidade de probabilidade da distribuição $\chi^2_{(n-p)}$ (vide Exercício 1.8).

1.6 Bandas de Confiança

Supor uma nova observação que não pertence à amostra com valores para as variáveis explicativas representados por $\mathbf{z} = (z_1, z_2, \ldots, z_p)^\top$. Portanto, tem-se que

$$y(\mathbf{z}) = \mathbf{z}^\top \boldsymbol{\beta} + \epsilon(\mathbf{z})$$

e valor esperado $\mathrm{E}\{Y(\mathbf{z})\} = \mu(\mathbf{z})$. Logo $\widehat{\mu}(\mathbf{z}) = \mathbf{z}^\top \widehat{\boldsymbol{\beta}}$ e

$$\mathrm{Var}\{\widehat{\mu}(\mathbf{z})\} = \mathrm{Var}(\mathbf{z}^\top \widehat{\boldsymbol{\beta}}) = \mathbf{z}^\top \mathrm{Var}(\widehat{\boldsymbol{\beta}})\mathbf{z} = \sigma^2 \mathbf{z}^\top (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{z}.$$

Desde que $\widehat{\mathrm{Var}}\{\widehat{\mu}(\mathbf{z})\} = s^2 \mathbf{z}^\top (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{z}$, uma estimativa intervalar de coeficiente de confiança $(1-\alpha)$ para $\mu(\mathbf{z})$ fica dada por

$$[\mathbf{z}^\top \widehat{\boldsymbol{\beta}} \pm \mathrm{t}_{(1-\alpha/2),(n-p)} s\{\mathbf{z}^\top (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{z}\}^{\frac{1}{2}}],$$

em que $t_{(1-\alpha/2),(n-p)}$ denota o quantil $(1-\alpha/2)$ de uma distribuição t-Student com $(n-p)$ graus de liberdade. A banda de coeficiente de confiança $(1-\alpha)$ para $\mu(\mathbf{z})$ assume a forma

$$[\mathbf{z}^\top \widehat{\boldsymbol{\beta}} \pm \sqrt{c_\alpha}\sigma\{\mathbf{z}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{z}\}^{\frac{1}{2}},\ \forall \mathbf{z} \in I\!R^p],$$

em que c_α é tal que $\mathrm{P}\{\chi^2_p \leq c_\alpha\} = 1-\alpha$ (vide, por exemplo, Rao, 1973).

Por outro lado, o valor predito de $Y(\mathbf{z})$ pode ser representado por $\widehat{y}(\mathbf{z}) = \mathbf{z}^\top\widehat{\boldsymbol{\beta}} + \epsilon(\mathbf{z})$ e portanto

$$\begin{aligned}
\mathrm{Var}\{\widehat{Y}(\mathbf{z})\} &= \mathrm{Var}\{\mathbf{z}^\top\widehat{\boldsymbol{\beta}} + \epsilon(\mathbf{z})\} = \mathrm{Var}\{\mathbf{z}^\top\widehat{\boldsymbol{\beta}}\} + \mathrm{Var}\{\epsilon(\mathbf{z})\} \\
&= \mathbf{z}^\top\mathrm{Var}(\widehat{\boldsymbol{\beta}})\mathbf{z} + \mathrm{Var}\{\epsilon(\mathbf{z})\} = \sigma^2\mathbf{z}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{z} + \sigma^2 \\
&= \sigma^2\{1 + \mathbf{z}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{z}\}.
\end{aligned}$$

Tem-se que $\widehat{\mathrm{Var}}\{\widehat{Y}(\mathbf{z})\} = s^2\{1 + \mathbf{z}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{z}\}$.

Assim, estimativa intervalar e banda de confiança de coeficiente de confiança $(1-\alpha)$ para $y(\mathbf{z})$ ficam, respectivamente, dadas por

$$[\mathbf{z}^\top\widehat{\boldsymbol{\beta}} \pm t_{(1-\alpha/2),(n-p)}s\{1 + \mathbf{z}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{z}\}^{\frac{1}{2}}]$$

e

$$[\mathbf{z}^\top\widehat{\boldsymbol{\beta}} \pm \sqrt{c_\alpha}\sigma\{1 + \mathbf{z}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{z}\}^{\frac{1}{2}},\ \forall \mathbf{z} \in I\!R^p].$$

Na prática deve-se substituir σ^2 por s^2 e c_α é obtido tal que $\mathrm{P}\{F_{p,(n-p)} \leq c_\alpha\} = 1-\alpha$. Em particular, para regressão linear simples é possível mostrar que $\mathbf{z}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{z} = 1/n + (z-\bar{x})^2/S_{xx}$.

1.7 Métodos de Diagnóstico

Procedimentos de diagnóstico devem ser aplicados após o ajuste do modelo linear normal e têm como principais objetivos:

(i) avaliar se há afastamentos importantes das suposições feitas para o modelo, tais como independência, normalidade, homocedasticidade dos erros e linearidade da média com relação aos valores das variáveis explicativas;

(ii) avaliar se há presença de observações atípicas ou discrepantes. Essas observações podem ser classificadas como pontos de alavanca, pontos aberrantes ou pontos influentes.

Abaixo segue descrição dos três tipos de observações atípicas.

Pontos de alavanca: observações em que o vetor $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^\top$ está remoto no subespaço C($\mathbf{X}$) gerado pelas colunas da matriz $\mathbf{X}$. Essas observações têm influência desproporcional no próprio valor ajustado.

Pontos aberrantes: observações com resíduo alto, posicionadas fora da banda de confiança. Ou seja, observações mal ajustadas pelo modelo. Em geral essas observações têm influência desproporcional na predição das respostas.

Pontos influentes: observações com peso desproporcional nas estimativas dos coeficientes do componente sistemático do modelo. Em geral são pontos de alavanca mas a recíproca nem sempre é verdadeira.

Na Figura 1.2 há uma descrição gráfica de observações atípicas. No primeiro gráfico (acima à esquerda) tem-se uma regressão hipotética com a reta ajustada passando pelas 5 observações, no segundo gráfico (acima à direita) a $3^{\underline{a}}$ observação é deslocada verticalmente de forma a tornar- se aberrante, enquanto no terceiro e quarto gráficos (abaixo à esquerda e à direita) a $5^{\underline{a}}$ observação é deslocada em direções diferentes de modo a tornar-se de alavanca e influente, respectivamente.

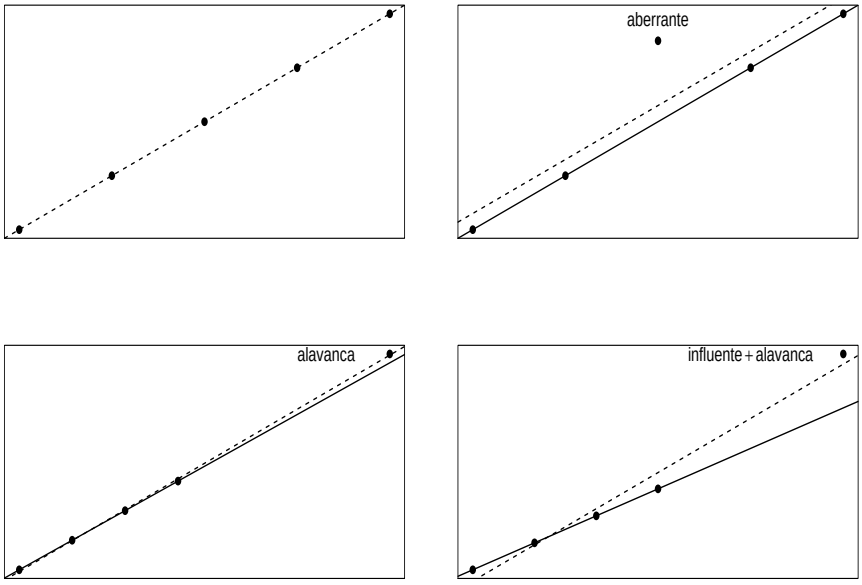

Figura 1.2: Representação gráfica para um conjunto de dados hipotéticos de pontos de alavanca, aberrantes e influentes. Reta ajustada com todos as observações (----) e sem a observação deslocada (——).

1.7.1 Pontos de Alavanca

Uma observação é definida como ponto de alavanca se tem uma alta influência no próprio valor ajustado. Essa influência é medida através da derivada $\partial\widehat{y}/\partial y$. Ou seja, mede o impacto que uma variação infinitesimal na resposta causa no valor ajustado. Da relação $\widehat{\mathbf{y}} = \mathbf{H}\mathbf{y}$ obtém-se $\widehat{y}_i = h_{ii}y_i + \sum_{j\neq i} h_{ij}y_j$, em que h_{ij} denota o elemento (i, j) da matriz $\mathbf{H}$ que é simétrica de dimensão $n \times n$. Daí segue que $\partial\widehat{y}_i/\partial y_i = h_{ii}$ e ainda pode-se mostrar que $h_{ii} = \mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i$.

Como a matriz $\mathbf{H}$ é idempotente ($\mathbf{H} = \mathbf{HH}$) segue que

$$\sum_{j=1}^{n} h_{ij}^2 = h_{ii} \rightarrow \sum_{j \neq i} h_{ij}^2 = h_{ii} - h_{ii}^2 = h_{ii}(1 - h_{ii}),$$

então $h_{ii} \geq 0$ e $h_{ii}(1 - h_{ii}) \geq 0$ e portanto $0 \leq h_{ii} \leq 1$. Note que se $h_{ii} = 1$ então $h_{ij} = 0 \quad \forall j \neq i$ e logo $\widehat{y}_i = y_i$. Hoaglin e Welsch (1978) propõem classificar pontos de alavanca segundo o critério $h_{ii} \geq 2\bar{h}$, em que $\bar{h} = \frac{\sum_{i=1}^{n} h_{ii}}{n}$. Assim, desde que

$$\sum_{i=1}^{n} h_{ii} = \text{tr}(\mathbf{H}) = \text{tr}\{\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\} = \text{tr}\{(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}\} = \text{tr}(\mathbf{I}_p) = p,$$

o critério fica dado por $h_{ii} \geq \frac{2p}{n}$. Para amostras grandes sugere-se $h_{ii} \geq \frac{3p}{n}$.

1.7.2 Limites para a Predição

Supor uma nova observação com valores para as variáveis explicativas representados por $\mathbf{z} = (z_1, z_2, \ldots, z_p)^\top$. Qual a condição para obter $\widehat{y}(\mathbf{z})$? Segundo Montgomery et al.(2021, p.110) pode-se fazer predição (interpolação) no modelo de regressão linear múltipla com segurança se a seguinte condição for satisfeita:

$$\mathbf{x}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x} \leq h_{\max} \ \forall \mathbf{x} \in I\!R^p,$$

em que $h_{\max} = \max\{h_{11}, \ldots, h_{nn}\}$. Logo, uma condição para predição de $y(\mathbf{z})$ é que $\mathbf{z}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{z} \leq h_{\max}$.

Na Figura 1.3 tem-se a representação geométrica da "região conjunta dos dados"para a qual recomenda-se fazer as predições do modelo linear $y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$, em que $a \leq x_1 \leq b$ e $c \leq x_2 \leq d$. Nota-se que há vários pares de valores (x_1, x_2) para os quais não é recomendado fazer interpolação.

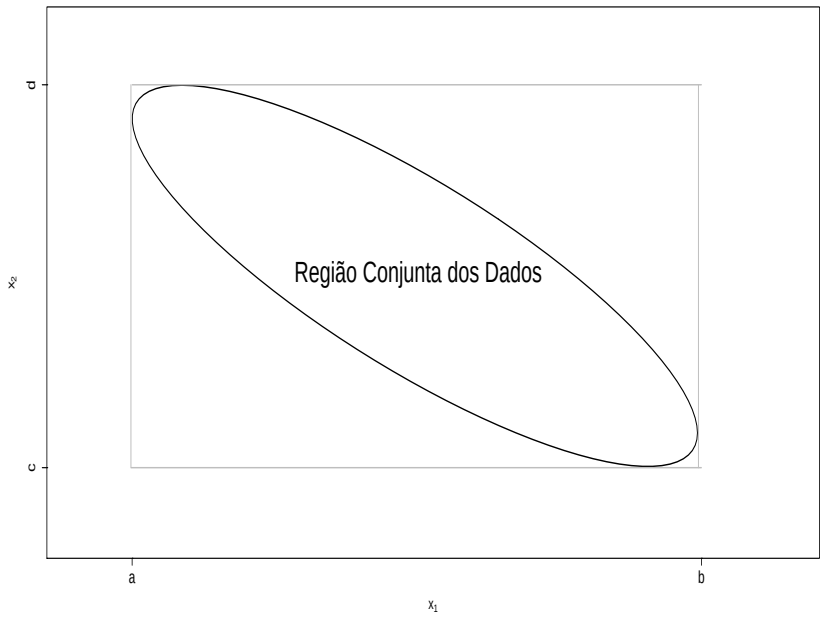

Figura 1.3: Representação geométrica para os limites de predição de um modelo de regressão (sem intercepto) com duas variáveis explicativas, com valores tais que $a \leq x_1 \leq b$ e $c \leq x_2 \leq d$.

1.7.3 Análise de Resíduos

Como visto anteriormente, o vetor de resíduos ordinários é definido por $\mathbf{r} = (\mathbf{I}_n - \mathbf{H})\mathbf{y}$, em que $\mathbf{r} = (r_1, \ldots, r_n)^\top$ com $r_i = y_i - \widehat{y}_i$, para $i = 1, \ldots, n$. Tem-se que

$$\begin{aligned} \mathrm{E}(\mathbf{r}) &= \mathrm{E}(\mathbf{Y}|\mathbf{X}) - \mathbf{H}\mathrm{E}(\mathbf{Y}|\mathbf{X}) \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} = \mathbf{0}. \end{aligned}$$

A matriz de variância-covariância de $\mathbf{r}$ fica dada por

$$\begin{aligned}
\mathrm{Var}(\mathbf{r}) &= \mathrm{Var}\{(\mathbf{I}_n - \mathbf{H})\mathbf{Y}|\mathbf{X}) \\
&= (\mathbf{I}_n - \mathbf{H})\mathrm{Var}(\mathbf{Y}|\mathbf{X})(\mathbf{I}_n - \mathbf{H}) \\
&= \sigma^2(\mathbf{I}_n - \mathbf{H})(\mathbf{I}_n - \mathbf{H}) \\
&= \sigma^2(\mathbf{I}_n - \mathbf{H}).
\end{aligned}$$

Portanto, segue que $\mathbf{r} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2(\mathbf{I}_n - \mathbf{H}))$, e consequentemente

(i) $r_i \sim \mathrm{N}(0, \sigma^2(1 - h_{ii}))$;

(ii) $\mathrm{Cov}(r_i, r_j) = -\sigma^2 h_{ij},\ i \neq j$ e

(iii) $\mathrm{Corr}(r_i, r_j) = \frac{-h_{ij}}{\sqrt{(1-h_{ii})(1-h_{jj})}},\ i \neq j,$

para $i, j = 1, \ldots, n$. Ou seja, os resíduos têm distribuição marginal normal de média zero, variâncias não constantes e são correlacionados.

Para que os resíduos sejam comparáveis é preciso padronizá-los. Uma padronização natural seria o resíduo normalizado

$$t_{r_i} = \frac{r_i}{\sigma\sqrt{1 - h_{ii}}} \sim \mathrm{N}(0, 1),\ i = 1, \ldots, n.$$

Porém, é preciso estimar σ^2. Sabe-se que a estatística t-Student é construída da seguinte forma:

$$t = \frac{Z}{\sqrt{U/\nu}} \sim t_\nu,$$

em que $Z \sim \mathrm{N}(0, 1)$, $U \sim \chi^2_\nu$ e Z e U são variáveis aleatórias independentes. Tem-se que $t_{r_i} \sim \mathrm{N}(0, 1)$ e é possível mostrar que $(n - p)s^2/\sigma^2 \sim \chi^2_{(n-p)}$, porém t_{r_i} e s^2 não são independentes. Logo, o resíduo

$$t_i = \frac{r_i}{s\sqrt{1 - h_{ii}}} \nsim t_{(n-p)}.$$

Cook e Weisberg (1982) mostram que $\frac{t_i^2}{(n-p)} \sim \text{Beta}(\frac{1}{2}, \frac{(n-p-1)}{2})$. A sugestão é substituir s^2 por $s^2_{(i)}$, o erro quadrático médio do modelo sem a i-ésima observação. Agora, tem-se que $t_{r_i} \sim \text{N}(0,1)$, $(n-p-1)s^2_{(i)}/\sigma^2 \sim \chi^2_{(n-p-1)}$ e ainda t_{r_i} e $s^2_{(i)}$ são independentes. Então, tem-se o resíduo Studentizado

$$t_i^* = \frac{r_i}{s_{(i)}\sqrt{1-h_{ii}}} \sim t_{(n-p-1)},$$

para $i = 1, \ldots, n$. É possível mostrar que

$$s^2_{(i)} = s^2 \left(\frac{n-p-t_i^2}{n-p-1} \right).$$

Ou seja, $s^2_{(i)}$ pode ser obtido sem a necessidade de fazer o ajuste sem a i-ésima observação.

Abaixo são descritos alguns gráficos sugeridos com o resíduo t_i^*.

(i) Gráfico entre os quantis observados $t^*_{1)} < \cdots < t^*_{(n)}$ do resíduo t_i^* contra os quantis da distribuição $\text{N}(0,1)$. Esse gráfico é equivalente ao gráfico normal de probabilidades sugerido em Montgomery et al. (2021, Cap.4). Sugere-se a inclusão de banda de confiança empírica, denominada envelope (Atkinson, 1981). Essa banda é recomendada em virtude dos resíduos serem correlacionados. Espera-se os pontos distribuídos de forma aleatória dentro da banda de confiança. Distorções no gráfico podem ser causadas por observações aberrantes e outras formas para o gráfico são indícios de afastamentos da normalidade dos erros.

(ii) Gráfico de t_i^* contra valores ajustados $\widehat{y}_i$. Desde que $\text{Cov}(\mathbf{r}, \widehat{\mathbf{y}}) = \mathbf{0}$, espera-se distribuição uniforme dos pontos conforma varia o valor ajustado. Afastamentos dessa tendência são indícios de que a variância dos erros não deve ser constante.

(iii) Gráfico de t_i^* contra a ordem das observações para detectar (quando fizer sentido) correlação temporal dos dados. Pode-se também aplicar o teste de Durbin-Watson para avaliar se há correlação autoregressiva positiva nos erros. Esse teste será discutido na Seção 1.19.1.

(iv) Gráfico de t_i^* contra valores de variáveis explicativas contínuas para avaliar se há algum termo que não foi incluído no componente sistemático do modelo. Alternativamente, tem-se o gráfico da variável adicionada (Seção 1.7.7).

A suposição de normalidade dos erros é crucial para fazer inferências quando o tamanho amostral n é pequeno ou moderado, contudo para n grande tem-se pelo Teorema Central do Limite (TCL) a normalidade assintótica de $\widehat{\boldsymbol{\beta}}$ desde que os erros tenham média zero e variância constante. Assim, quando há indícios de afastamentos importantes da suposição de normalidade dos erros pode-se tentar aplicar alguma transformaçao apropriada $g(Y)$ a fim de alcançar a normalidade mesmo que aproximadamente (vide Exercícios 1.16 e 1.17). O incoveniente desse procedimento é que o novo modelo estará explicando $\mathrm{E}\{g(Y)\}$ ao invés de $\mathrm{E}(Y)$. Outra opção seria aplicar modelos lineares generalizados, em que procura-se uma distribuição apropriada para Y, porém tem-se em contrapartida a modelagem de $\mathrm{E}(Y)$. No caso da violação da suposição de variância constante para os erros, uma primeira opção seria aplicar regressão linear ponderada (Seção 1.10) que flexibiliza a variância dos erros sem comprometer os resultados da regressão linear. Alternativamente, pode-se aplicar a modelagem dupla em que $\mathrm{E}(Y)$ e $\mathrm{Var}(Y)$ são modelados conjuntamente.

Para amostras pequenas e moderadas, quando há violação da suposição de erros normais, pode-se aplicar procedimentos de reamostragem para estimação e inferência dos coeficientes da regressão (vide, por exemplo, Fox e

Weisberg, 2019).

1.7.4 Outra Interpretação para t_i^*

Supor que o i-ésimo ponto é suspeito de ser aberrante. Essa hipótese pode ser testada através do modelo

$$y_j = \mathbf{x}_j^\top \boldsymbol{\beta} + \omega_j \gamma + \epsilon_j, \tag{1.4}$$

em que $\mathbf{x}_j = (x_{j1}, \ldots, x_{jp})^\top$ e $\epsilon_j \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$ para $j = 1, \ldots, n$, com $\omega_j = 1$ para $j = i$ e $\omega_j = 0$ em caso contrário. Usando resultados da Seção 1.4 pode-se mostrar que sob a hipótese $\text{H}_0 : \gamma = 0$ o acréscimo na soma de quadrados de resíduos fica dado por

$$\text{ASQ}(\gamma = 0) = \widehat{\gamma}^2(1 - h_{ii}),$$

em que $\widehat{\gamma} = r_i(1 - h_{ii})^{-1}$ com $r_i = y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}$ e $h_{ii} = \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i$. Logo, a soma de quadrados de resíduos correspondente ao modelo (1.4) fica dada por $(n-p)s^2 - \widehat{\gamma}^2(1 - h_{ii}) = (n-p)s^2 - \frac{r_i^2}{1-h_{ii}}$ e a estatística F para testar $\text{H}_0 : \gamma = 0$ contra $\text{H}_1 : \gamma \neq 0$ assume a forma

$$\text{F} = \frac{\widehat{\gamma}^2(1 - h_{ii})}{\left\{(n-p)s^2 - \frac{r_i^2}{(1-h_{ii})}\right\}/(n-p-1)} \overset{\text{H}_0}{\sim} \text{F}_{1,(n-p-1)}.$$

Trabalhando um pouco a expressão acima chega-se ao seguinte resultado:

$$\text{F} = \frac{r_i^2(n-p-1)}{s^2(1-h_{ii})(n-p-t_i^2)} = t_i^{*2}.$$

Portanto, para um nível de significância α, rejeita-se H_0 se $|t_i^*| > \text{t}_{(1-\alpha/2),(n-p-1)}$.

1.7.5 Análise de Influência

O objetivo principal da análise de influência em regressão é avaliar o impacto de perturbações no modelo e/ou dados nos coeficientes da regressão, sendo

esse impacto avaliado através de alguma medida de influência. A medida de influência mais conhecida, denominada distância de Cook (Cook, 1977), procura avaliar o impacto da retirada de cada observação nas estimativas dos coeficientes. Uma vez detectadas as observações com maior variação para essa medida, deve-se proceder algum tipo de análise confirmatória a fim de avaliar a influência das observações destacadas e também o tipo de influência. Variações numéricas nas estimativas dos coeficientes são esperadas quando elimina-se observações, contudo quando essas variações são desproporcionais, muito acima $\frac{1}{n} \times 100\%$, as observações podem ser consideradas influentes. O mais grave é quando a eliminação individual de uma observação leva a mudanças inferenciais, ou seja, determinados coeficientes deixam ou passam a ser significativos. No primeiro caso a observação induz o efeito do coeficiente enquanto que no segundo caso há mascaramento do efeito pela observação.

Transformações dos valores das variáveis explicativas, inclusão de interação ou mesmo ponderação na regressão, dentre outros procedimentos, são comumente aplicados para reduzir a influência de observações na regressão. Contudo, quando esses procedimentos não levam a soluções satisfatórias recomenda-se a aplicação de procedimentos de estimação robusta. Na Seção 1.16 são apresentados alguns procedimentos usuais de estimação robusta para regressão linear múltipla. Uma discussão mais abrangente pode ser encontrada em Montogomery et al. (2021, Cap.15).

Nesta seção será discutida a distância de Cook aplicada ao modelo de regressão linear múltipla (1.2). Essa medida pode ser motivada através da região de confiança de coeficiente $(1-\alpha)$ para $\boldsymbol{\beta}$, dada por

$$\frac{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top (\mathbf{X}^\top \mathbf{X})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{ps^2} \leq F_{(1-\alpha),p,(n-p)},$$

em que $F_{(1-\alpha),p,(n-p)}$, como definido anteriormente, denota o quantil $(1-\alpha)$ de uma distribuição F com p e $(n-p)$ graus de liberdade. Essa região de

confiança é construída usando o resultado abaixo

$$P\left\{\frac{(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top(\mathbf{X}^\top\mathbf{X})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{ps^2}\leq F_{(1-\alpha),p,(n-p)}\right\}=1-\alpha.$$

Na Figura 1.4 tem-se a representação gráfica da superfície correspondente à região de confiança para os coeficientes de uma regressão hipotética com $p=2$.

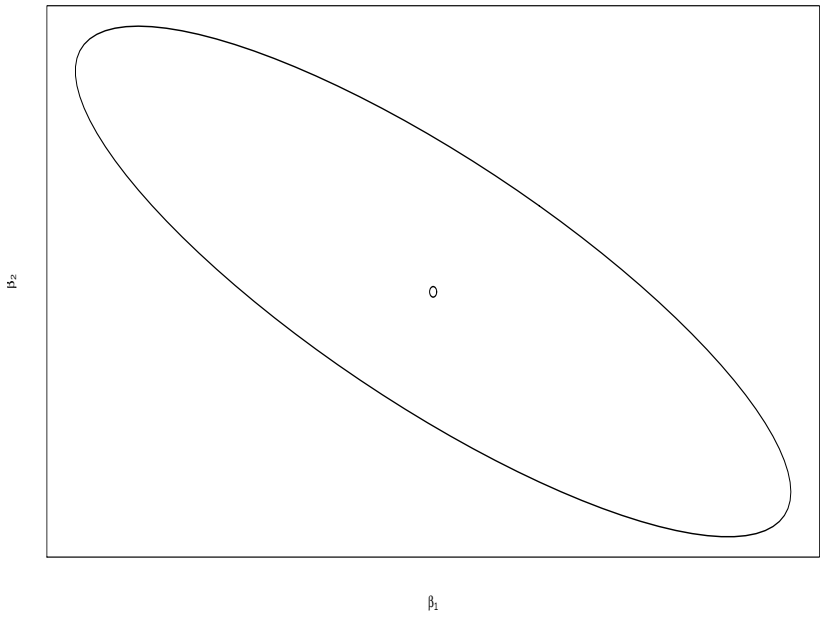

Figura 1.4: Representação geométrica para a região de confiança de 95% para os coeficientes de um modelo de regressão hipotético com $p=2$.

A distância de Cook é definida por

$$\mathrm{D}_i=\frac{(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{(i)})^\top(\mathbf{X}^\top\mathbf{X})(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{(i)})}{ps^2},$$

em que $\widehat{\boldsymbol{\beta}}_{(i)}$ denota a estimativa de mínimos quadrados quando a i-ésima observação não é considerada no modelo. Após manipulações algébricas obtém-

se

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{(i)} &= \{\mathbf{X}_{(i)}^\top \mathbf{X}_{(i)}\}^{-1}\mathbf{X}_{(i)}^\top \mathbf{y}_{(i)} \\
&= \{\mathbf{X}^\top\mathbf{X} - \mathbf{x}_i\mathbf{x}_i^\top\}^{-1}\{\mathbf{X}^\top\mathbf{y} - \mathbf{x}_i y_i\} \\
&= \left\{(\mathbf{X}^\top\mathbf{X})^{-1} + \frac{(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1}}{1-h_{ii}}\right\}\{\mathbf{X}^\top\mathbf{y} - \mathbf{x}_i y_i\} \\
&= \widehat{\boldsymbol{\beta}} - \frac{r_i}{(1-h_{ii})}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i,
\end{aligned}$$

para $i = 1, \ldots, n$. Portanto, tem-se que

$$\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(i)} = \frac{r_i}{(1-h_{ii})}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i.$$

Assim, a distância de Cook fica dada

$$\mathrm{D}_i = \frac{1}{p}t_i^2\frac{h_{ii}}{(1-h_{ii})}.$$

Como $h_{ii}/(1-h_{ii})$ é uma função crescente de h_{ii}, então D_i será grande se $|t_i|$ e/ou h_{ii} forem (for) grande(s). Uma proposta de pontos suspeitos de serem influentes, baseada na região de confiança para $\boldsymbol{\beta}$, é destacar as observações tais que $\mathrm{D}_i \geq F_{(1-\alpha),p,(n-p)}$. Outras sugestões se baseiam em obter limites superiores para a distância de Cook com base nas variações dos valores amostrais da distância e que levem em conta o tamanho amostral. Sugere-se destacar as observações tais que $\mathrm{D}_i \geq \bar{\mathrm{D}} + k\mathrm{DP}(\mathrm{D}_i)$, para $k = 2, 3, 4$. Deve-se aumentar o valor k à medida que aumenta o tamanho amostral.

Outra medida de influência proposta por Belsley et al. (1980), que é derivada da distância de Cook com s^2 substuído por $s_{(i)}^2$, é definida por

$$\begin{aligned}
\mathrm{DFFITS}_i &= \frac{|r_i|}{s_{(i)}\sqrt{1-h_{ii}}}\left\{\frac{h_{ii}}{1-h_{ii}}\right\}^{\frac{1}{2}} \\
&= |t_i^*|\left\{\frac{h_{ii}}{1-h_{ii}}\right\}^{\frac{1}{2}}.
\end{aligned}$$

Sugere-se destacar as observações tais que $\text{DFFITS}_i \geq 2\{p/(n-p)\}^{\frac{1}{2}}$. Essa medida leva também em conta a influência das observações na estimativa de σ^2. Contudo, quando o interesse está apenas nos coeficientes da regressão sugere-se utilizar apenas a distância de Cook.

Finalmente, pode haver interesse em estudar a influência das observações em coeficientes específicos da regressão. Por exemplo, se há interesse em avaliar a influência da eliminação da i-ésima observação no j-ésimo coeficiente estimado da regressão, utiliza-se a seguinte medida de influência:

$$\begin{aligned}
\text{DFBETAS}_{ji} &= \frac{(\widehat{\beta}_j - \widehat{\beta}_{j(i)})}{s_{(i)}\sqrt{C_{jj}}} \\
&= \frac{\mathbf{C}_j^\top \mathbf{x}_i r_i}{s_{(i)}(1-h_{ii})\sqrt{C_{jj}}} \\
&= \frac{p_{ji}}{\sqrt{\mathbf{p}_j^\top \mathbf{p}_j}} \frac{t_i^*}{\sqrt{1-h_{ii}}},
\end{aligned}$$

em que $\mathbf{C} = (\mathbf{X}^\top\mathbf{X})^{-1}$, $\mathbf{C}_j$ denota a j-ésima coluna de $\mathbf{C}$, p_{ji} e $\mathbf{p}_j^\top$ denotam, respectivamente, o (j,i)-ésimo elemento e a j-ésima linha de $\mathbf{P} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$, $i = 1, \ldots, n$ e $j = 1, \ldots, p$. Recomenda-se dar atenção àquelas observações tais que $\text{DFBETAS}_{ji} > \frac{2}{\sqrt{n}}$ (vide Montgomery et al., 2021, Cap.6).

1.7.6 Análise Confirmatória

O procedimento mais tradicional de verificação das observações suspeitas de serem discrepantes é através da deleção individual de cada observação suspeita, computando-se a variação percentual de cada coeficiente da regressão e o respectivo valor-P. Para ilustrar alguns procedimentos, denote o conjunto das m observações supeitas por $\text{S} = \{\text{S}_1, \ldots, \text{S}_m\}$.

Variação Percentual

A variação percentual do j-ésimo coeficiente da regressão quando a i-ésima observação não é considerada no ajuste é definido por

$$\Delta_{ij} = \left| \frac{\hat{\beta}_{(i)j} - \hat{\beta}_j}{\hat{\beta}_j} \right| \times 100\%,$$

para $j = 1, \ldots, p$ e $i \in$ S. Deve-se associar a cada observação deletada o novo valor-P de cada coeficiente. Variações percentuais desproporcionais (muito acima de $(1/n) \times 100\%$) são esperadas, porém deve-se dar atenção quando ocorrerem mudanças inferenciais.

Comparação com Observações não Destacadas

Um outro procedimento usual é comparar alguma medida resumo das observações suspeitas com a mesma medida resumo obtida de r amostras aleatórias de tamanho m das observações não suspeitas. Por exemplo, pode-se computar a medida

$$\text{MRC}_S = \max_{1 \leq j \leq p} \left| \frac{\hat{\beta}_{(S)j} - \hat{\beta}_j}{\hat{\beta}_j} \right|.$$

Comparar MRC_S com as r medidas, $\text{MRC}_{NS_1}, \ldots, \text{MRC}_{NS_r}$, das r amostras aleatórias de tamanho m extraídas do grupo de observações não suspeitas. Se MRC_S for muito maior que $\max_{1 \leq j \leq r} \text{MRC}_{NS_j}$ é um indício de que as observações em S são discrepantes. Sugere-se utilizar que $r \geq 10$.

Tratamentos de Observações Discrepantes

Os procedimentos descritos abaixo são usuais para acomodar observações discrepantes.

- Aplicar transformações nas variáveis explicativas, por exemplo padronização, raiz quadrada e logarítmica.

- Incluir termos não lineares em variáveis explicativas contínuas.
- Incluir (ou retirar) interações.
- Aplicar regressão linear ponderada.
- Aplicar métodos robustos.
- Mudar a distribuição dos erros. Por exemplo, erros com caudas mais leves ou mais pesadas do que as caudas da distribuição normal padrão ou erros assimétricos.

1.7.7 Gráfico da Variável Adicionada

Supor que uma variável explicativa é adicionada no modelo (1.2) obtendo-se o seguinte modelo de regressão linear:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{w}\gamma + \boldsymbol{\epsilon}$$

em que $\mathbf{X}$ denota a matriz modelo $n \times p$ do modelo reduzido, $\mathbf{w}$ denota vetor $n \times 1$ dos valores observados da variável adicionada, $\mathbf{y}$ é o vetor $n \times 1$ dos valores observados da variável resposta, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ e $\boldsymbol{\epsilon} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2\mathbf{I_n})$. Definindo $\mathbf{Z} = (\mathbf{X}, \boldsymbol{\omega})$ como matriz do modelo ampliado, mostra-se facilmente que a estimativa de mínimos quadrados de $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma})^\top$ fica expressa na forma $\hat{\boldsymbol{\theta}} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}$. Após algumas manipulações algébricas a estimativa de mínimos quadrados do coeficiente da variável adicionada fica dada por

$$\begin{aligned} \hat{\gamma} &= \frac{\boldsymbol{\omega}^\top(\mathbf{I}_n - \mathbf{H})\mathbf{y}}{\boldsymbol{\omega}^\top(\mathbf{I}_n - \mathbf{H})\boldsymbol{\omega}} \\ &= \frac{\boldsymbol{\omega}^\top\mathbf{r}}{\boldsymbol{\omega}^\top(\mathbf{I}_n - \mathbf{H})\boldsymbol{\omega}}. \end{aligned}$$

Ou seja, $\hat{\gamma}$ pode ser expresso como sendo o coeficiente da regressão linear passando pela origem do vetor de resíduos $\mathbf{r} = (\mathbf{I}_n - \mathbf{H})\mathbf{y}$ sobre o novo

resíduo $\boldsymbol{\upsilon} = (\mathbf{I}_n - \mathbf{H})\boldsymbol{\omega}$, dado por

$$\begin{aligned}\hat{\gamma} &= (\boldsymbol{\upsilon}^\top \mathbf{v})^{-1}\boldsymbol{\upsilon}^\top \mathbf{r} \\ &= \{\boldsymbol{\omega}^\top(\mathbf{I}_n - \mathbf{H})(\mathbf{I}_n - \mathbf{H})\boldsymbol{\omega}\}^{-1}\boldsymbol{\omega}^\top(\mathbf{I}_n - \mathbf{H})(\mathbf{I}_n - \mathbf{H})\mathbf{y} \\ &= \frac{\boldsymbol{\omega}^\top(\mathbf{I}_n - \mathbf{H})\mathbf{y}}{\boldsymbol{\omega}^T(\mathbf{I}_n - \mathbf{H})\boldsymbol{\omega}}.\end{aligned}$$

Portanto, o gráfico de $\mathbf{r}$ contra $\boldsymbol{\upsilon}$ pode fornecer informações sobre a evidência dessa regressão, indicando quais observações que estão contribuindo para a relação linear e quais observações que estão se desviando da mesma. Esse gráfico, conhecido como gráfico da variável adicionada (ver, por exemplo, Atkinson, 1985) pode revelar quais observações que estão influenciando (e de que maneira) a inclusão da nova variável explicativa no modelo.

A sugestão é que seja construído para cada variável explicativa contínua incluída de forma linear no modelo um gráfico da variável adicionada.

1.7.8 Aplicação

Para ilustrar um exemplo de regressão linear simples considere parte dos dados descritos em Neter et al. (1996, p.449) referentes à venda no ano anterior de um tipo de telhado de madeira em $n = 26$ filiais de uma rede de lojas de construção civil (arquivo **vendas.txt**). Apenas duas variáveis serão consideradas:

(i) Telhados: total de telhados vendidos (em mil metros quadrados) e

(ii) Nclientes: número de clientes cadastrados na loja (em milhares).

O interesse é explicar o número médio de telhados vendidos dado o número de clientes cadastrados. Na Tabela 1.2 são apresentadas algumas medidas resumo referentes às duas variáveis observadas.

Tabela 1.2: Medidas resumo referentes ao exemplo sobre venda de telhados.

Medida	Telhados	Nclientes
Média	170,20	51,85
D.Padrão	84,55	14,21
CV(em %)	49,68	27,41
Mínimo	30,90	26,00
1º Quartil	102,00	49,50
Mediana	159,80	51,50
3º Quartil	217,50	61,50
Máximo	339,40	75,00

Na Figura 1.5 tem-se o boxplot robusto (Hubert e Vandervierin, 2008) e a densidade estimada do total de telhados vendidos. Nota-se ausência de observações aberrantes e uma ligeira assimetria à direita. O diagrama de dispersão entre o total de telhados vendidos e o número de clientes cadastrados na loja (Figura 1.6) apresenta uma tendência aproximadamente linear e positiva. À medida que aumenta o número de clientes aumenta o total de telhados vendidos.

Tabela 1.3: Estimativas dos parâmetros referentes ao modelo de regressão linear simples ajustado aos dados sobre venda de telhados.

Efeito	Estimativa	E.Padrão	valor-t	valor-P
Constante	-71,208	40,558	-1,76	0,092
Nclientes	4,656	0,756	6,16	0,000
s	53,69			
R^2	0,61			
$\overline{R}^2$	0,60			

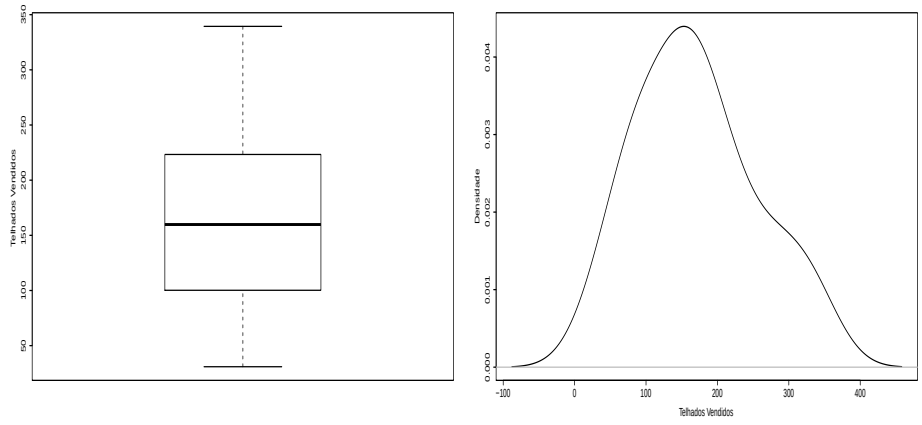

Figura 1.5: Boxplot robusto e densidade estimada do total de telhados vendidos.

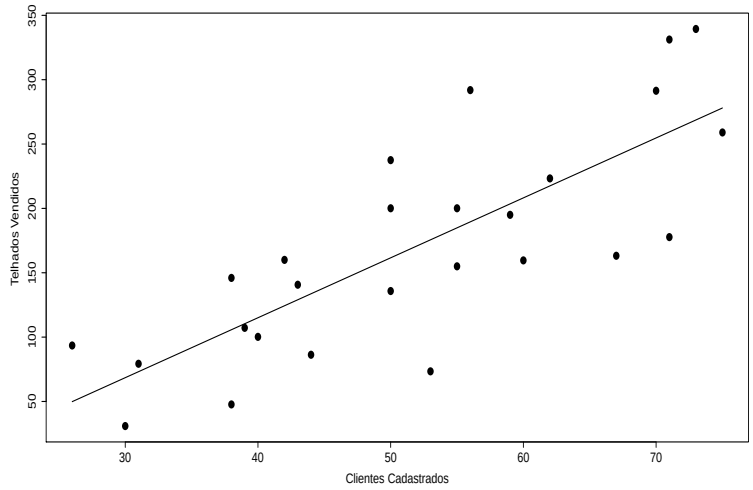

Figura 1.6: Diagrama de dispersão (com tendência) entre o total de telhados vendidos e o número de clientes cadastrados na loja.

Portanto, sugere-se o seguinte modelo de regressão linear simples:

$$y_i = \beta_1 + \beta_2 \text{Nclientes}_i + \epsilon_i,$$

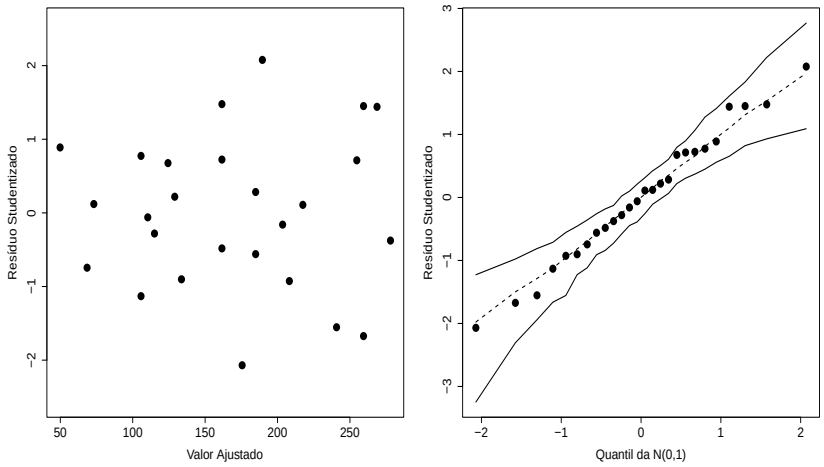

Figura 1.7: Gráficos de resíduos referentes ao modelo de regressão linear simples ajustado aos dados sobre venda de telhados.

em que y_i denota o total de telhados vendidos na i-ésima filial e $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, 26$. Nota-se pela Tabela 1.3 que o coeficiente estimado do número de clientes é altamente significativo e o intercepto é significativo ao nível de 10%. Assim, para um aumento de 1000 clientes em qualquer filial espera-se aumento de 4656 mil m^2 de telhados vendidos.

Pela Figura 1.7, em que são apresentados o gráfico do resíduo t_i^* contra o valor ajustado $\widehat{y}_i$ e o gráfico normal de probabilidades para t_i^* com banda empírica de confiança (envelope) de 95%, não há indícios de variância não constante nem de afastamentos da normalidade dos erros. Nota-se também ausência de observações aberrantes. O gráfico da distância de Cook com $k = 2$ (Figura 1.8) contra a ordem das observações destaca como possivelmente influentes as observações #6 e #10. O ajuste sem cada uma das observações traz variações nas estimativas dos coeficientes, como pode ser notado pela Figura 1.9, porém não há mudanças inferencias. Finalmente,

tem-se na Figura 1.10 as bandas de confiança de 95% para o número esperado de telhados vendidos e para o número de telhados vendidos de uma filial qualquer, dado o número de clientes cadastrados.

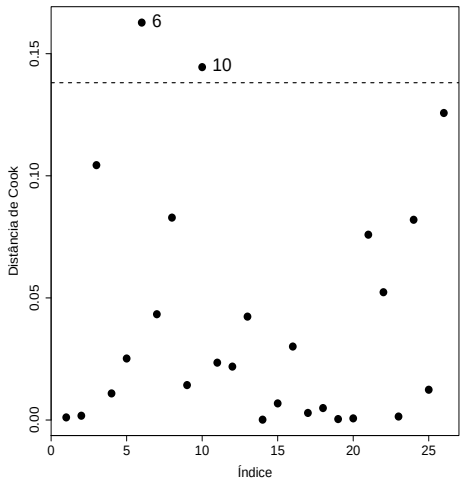

Figura 1.8: Distância de Cook contra a ordem das observações referente ao modelo de regressão linear simples ajustado aos dados sobre venda de telhados.

1.8 Variável Binária e Interação

Supor o seguinte modelo de regressão linear múltipla:

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i,$$

em que $y_1, \ldots, y_n$ são valores observados da variável resposta, x_{i2} representa os valores de uma variável aleatória binária tal que

$$x_{i2} = \begin{cases} 1 & \text{grupo A} \\ 0 & \text{grupo B}, \end{cases}$$

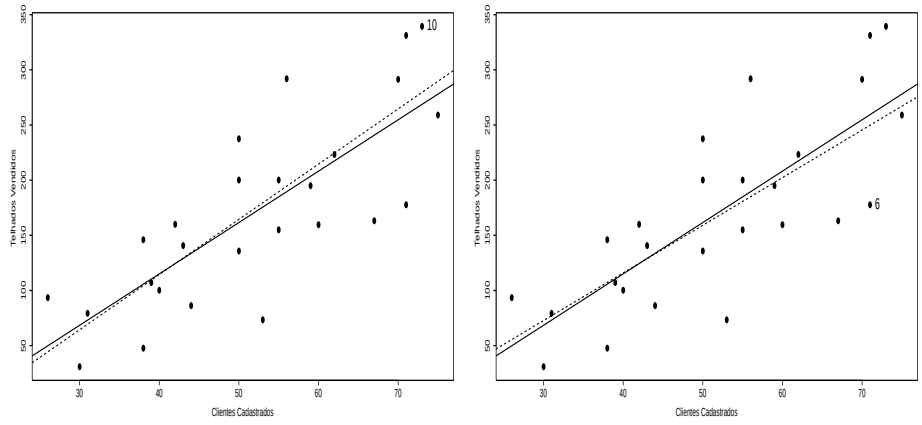

Figura 1.9: Retas ajustadas com todos os pontos (----) e sem as observações destacatadas pela distância de Cook (——).

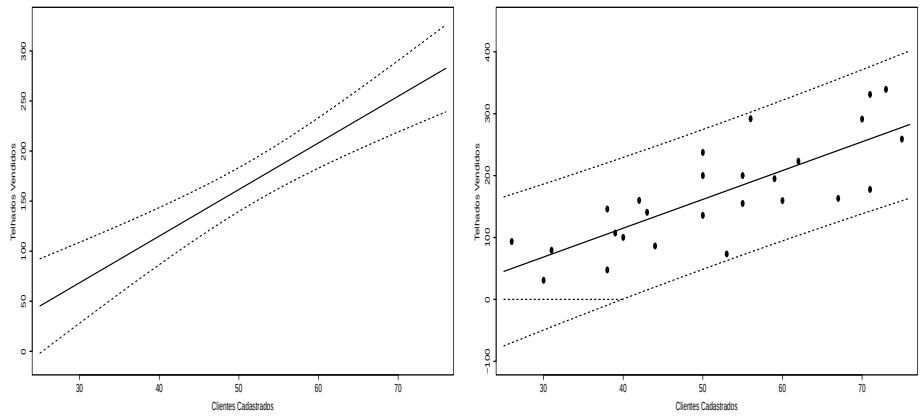

Figura 1.10: Bandas de confiança de 95% para o número esperado de telhados vendidos (esquerda) e para o número de telhados vendidos de uma filial qualquer (direita), dado o número de clientes cadastrados.

enquanto x_{i3} representa valores observados de uma variável contínua e $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$.

Portanto, tem-se dois submodelos de regressão

- (Grupo A) $y_i = \beta_1 + \beta_2 + \beta_3 x_{i3} + \epsilon_i$
- (Grupo B) $y_i = \beta_1 + \beta_3 x_{i3} + \epsilon_i$

com valores esperados

- $\mathrm{E_A}(Y_i|x_{i3}) = \beta_1 + \beta_2 + \beta_3 x_{i3}$
- $\mathrm{E_B}(Y_i|x_{i3}) = \beta_1 + \beta_3 x_{i3}$,

para $i = 1, \ldots, n$. Assim, $\mathrm{E_A}(Y_i|x_{i3}) - \mathrm{E_B}(Y_i|x_{i3}) = \beta_2$, que indica ausência de interação (paralelismo) entre as variáveis explicativas X_2 e X_3 (vide ilustração na Figura 1.11).

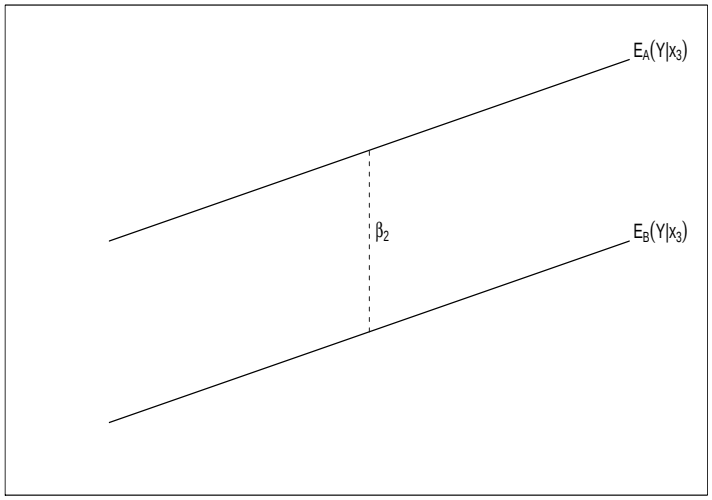

Figura 1.11: Descrição gráfica de ausência de interação (paralelismo) entre as variáveis explicativas X_2 e X_3.

Supor agora a inclusão de interação entre as variáveis explicativas X_2 e X_3, resultando no seguinte modelo de regressão linear múltipla:

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i2} x_{i3} + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$. Tem-se os seguintes submodelos:

- (Grupo A) $y_i = \beta_1 + \beta_2 + \beta_3 x_{i3} + \beta_4 x_{i3} + \epsilon_i$
- (Grupo B) $y_i = \beta_1 + \beta_3 x_{i3} + \epsilon_i$

com valores esperados expressos por

- $\text{E}_\text{A}(Y_i|x_{i3}) = \beta_1 + \beta_2 + \beta_3 x_{i3} + \beta_4 x_{i3}$
- $\text{E}_\text{B}(Y_i|x_{i3}) = \beta_1 + \beta_3 x_{i3}$,

para $i = 1, \ldots, n$. Assim, a diferença entre os valores esperados, $\text{E}_\text{A}(Y_i|x_{i3}) - \text{E}_\text{B}(Y_i|x_{i3}) = \beta_2 + \beta_4 x_{i3}$, não é mais constante dependendo dos valores da variável explicativa X_3. Isso indica presença de interação (ausência de paralelismo) entre as variáveis explicativas X_2 e X_3 (vide Figura 1.12).

Supor agora variável explicativa categórica com três níveis

$$X = \begin{cases} 1 & \text{grupo A} \\ 2 & \text{grupo B} \\ 3 & \text{grupo C.} \end{cases}$$

Um maneira de representar essa variável explicativa num modelo de regressão é atribuindo a cada grupo uma variável binária da seguinte forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i,$$

em que $y_1, \ldots, y_n$ denotam os valores observados da variável resposta, x_{i1}, x_{i2} e x_{i3} são os valores observados das variáveis binárias representando os grupos e $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $j = 1, \ldots, n$.

Supondo que os grupos A, B e C têm n_1, n_2 e n_3 elementos, respectivamente, o modelo pode ser expresso na forma matricial $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, em que $\mathbf{y} = (\mathbf{y}_1^\top, \mathbf{y}_2^\top, \mathbf{y}_3^\top)^\top$ com $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$, para $i = 1, 2, 3$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top$ e matriz $\mathbf{X}$ de dimensão $(n_1 + n_2 + n_3) \times 4$ dada por

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

Note que a matriz $\mathbf{X}$ não tem posto coluna completo, a $1^{\underline{a}}$ coluna é a soma das outras três colunas. Uma solução é reduzir o número de colunas da matriz modelo impondo alguma restrição nos parâmetros.

Os seguintes procedimentos são mais utilizados:

- Restrição nos parâmetros: $\beta_1 + \beta_2 + \beta_3 = 0$, que implica em $\beta_1 = -\beta_2 - \beta_3$.

- Casela de referência: um dos coeficientes é fixado como sendo zero. Por exemplo, fazendo $\beta_1 = 0$ o grupo A será denominado casela de referência.

Nesses dois casos $\boldsymbol{\beta} = (\beta_0, \beta_2, \beta_3)^\top$ e a matriz modelo terá dimensão $n \times 3$ com posto coluna completo.

Como exemplo, o modelo com casela de referência no grupo A pode ser expresso na forma

$$y_i = \beta_0 + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i,$$

em que $y_1, \ldots, y_n$ denotam os valores observados da variável resposta, x_{i2} e x_{i3} são valores de variáveis binárias representando os grupos B e C, respectivamente, e $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$. Quando $x_{i2} = x_{i3} = 0$ tem-se o grupo A. A matriz modelo nesse caso fica dada por

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{bmatrix}.$$

Supor agora a inclusão de uma variável explicativa contínua na parte sistemática do modelo

$$y_i = \beta_0 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i,$$

em que x_{i4}, para $i = 1, \ldots, n$, representa os valores observados da variável explicativa contínua. Portanto, tem-se três submodelos

- (Grupo A) $y_i = \beta_0 + \beta_4 x_{i4} + \epsilon_i$
- (Grupo B) $y_i = \beta_0 + \beta_2 + \beta_4 x_{i4} + \epsilon_i$
- (Grupo C) $y_i = \beta_0 + \beta_3 + \beta_4 x_{i4} + \epsilon_i$

com diferenças de valores esperados

- $\text{E}_\text{B}(Y_i|x_{i4}) - \text{E}_\text{A}(Y_i|x_{i4}) = \beta_2$
- $\text{E}_\text{C}(Y_i|x_{i4}) - \text{E}_\text{A}(Y_i|x_{i4}) = \beta_3$,

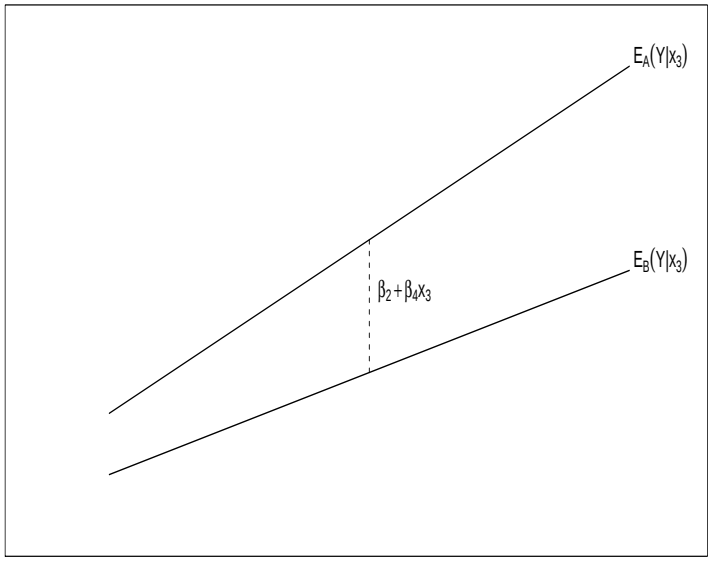

Figura 1.12: Descrição gráfica de presença de interação (ausência de paralelismo) entre as variáveis explicativas X_2 e X_3.

para $i = 1, \ldots, n$. Assim, os efeitos β_2 e β_3 são incrementos nos valores esperados dos grupos B e C, respectivamente, com relação ao grupo A (vide ilustração na Figura 1.13).

Em forma matricial o modelo com ausência de interação fica dado por $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, em que $\mathbf{y} = (\mathbf{y}_1^\top, \mathbf{y}_2^\top, \mathbf{y}_3^\top)^\top$ com $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$, para $i = 1, 2, 3$, $\boldsymbol{\beta} = (\beta_0, \beta_2, \beta_3, \beta_4)^\top$ e a matriz modelo $\mathbf{X}$ terá adicionada a coluna $(x_1, \ldots, x_{n_1}, x_{n_1+1}, \ldots, x_{n_1+n_2}, x_{n_1+n_2+1}, \ldots, x_n)^\top$.

O modelo com interação entre a variável categórica X e a variável contínua X_4 pode ser expresso na seguinte forma:

$$y_i = \beta_0 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i2} x_{i4} + \beta_6 x_{i3} x_{i4} + \epsilon_i,$$

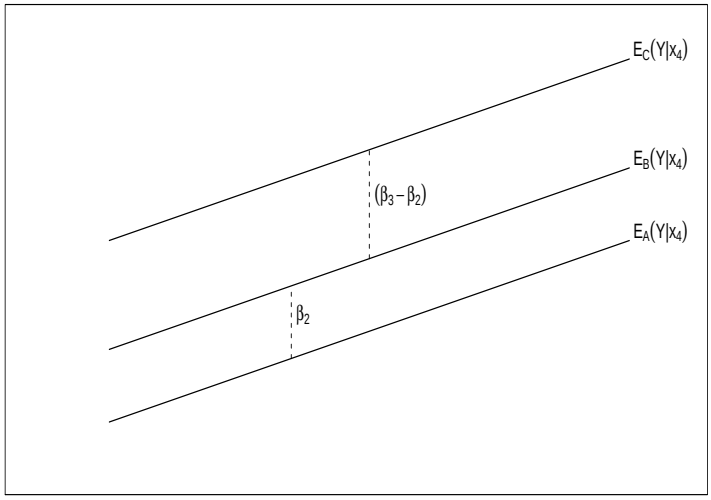

Figura 1.13: Descrição gráfica de ausência de interação (paralelismo) entre a variável categórica X e a variável contínua X_4.

em que $y_1, \ldots, y_n$ denotam os valores observados da variável resposta, x_{i2} e x_{i3} são valores de variáveis binárias representando os grupos B e C, respectivamente, enquanto x_{i4} representa os valores observados de uma variável contínua e $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$.

Portanto, tem-se três submodelos

- (Grupo A) $y_i = \beta_0 + \beta_4 x_{i4} + \epsilon_i$
- (Grupo B) $y_i = \beta_0 + \beta_2 + \beta_4 x_{i4} + \beta_5 x_{i4} + \epsilon_i$
- (Grupo C) $y_i = \beta_0 + \beta_3 + \beta_4 x_{i4} + \beta_6 x_{i4} + \epsilon_i$,

com diferenças de valores esperados

- $\text{E}_\text{B}(Y_i|x_{i4}) - \text{E}_\text{A}(Y_i|x_{i4}) = \beta_2 + \beta_5 x_{i4}$

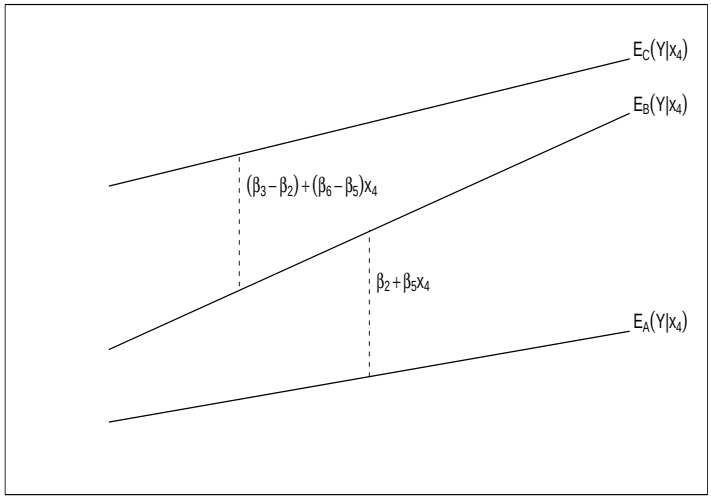

Figura 1.14: Descrição gráfica de interação entre a variável categórica X e a variável contínua X_4.

- $\mathrm{E_C}(Y_i|x_{i4}) - \mathrm{E_A}(Y_i|x_{i4}) = \beta_3 + \beta_6 x_{i4}$,

para $i = 1, \ldots, n$. Assim, nota-se que as diferenças entre os valores esperados dependem dos valores da variável explicativa X_4 (vide Figura 1.14). A matriz modelo $\mathbf{X}$ terá duas colunas adicionais com relação à matriz modelo sob ausência de interação.

O conceito de interação pode ser estendido para quaisquer tipos de variáveis explicativas e para mais do que duas variáveis explicativas. Contudo, devido a dificuldades na interpretação, em geral considera-se apenas interações de $1^{\underline{a}}$ ordem (entre duas variáveis eplicativas). Na Seção 1.14.2 é apresentado um exemplo ilustrativo em que a interação entre duas variáveis quantitativas agrega informações relavante nas interpretações do problema.

1.9 Comparação de Médias

Uma aplicação de modelos de regressão linear com variáveis binárias é na comparação das médias de k grupos. O modelo pode ser expresso na forma

$$y_{ij} = \alpha + \beta_i + \epsilon_{ij},$$

em que $\epsilon_{ij} \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, k$ e $j = 1, \ldots, n_i$, com a restrição $\beta_1 = 0$. O Grupo 1 é denominado casela de referência. Assim, tem-se os valores esperados

- $\text{E}(Y_{1j}) = \alpha$ para $j = 1, \ldots, n_1$
- $\text{E}(Y_{ij}) = \alpha + \beta_i$, para $i = 2, \ldots, k$ e $j = 1, \ldots, n_i$,

e daí segue que β_i é o incremento no valor médio do i-ésimo grupo com relação ao valor médio do grupo 1, para $i = 2, \ldots, k$. Testar a igualdade de médias equivale a testar $\text{H}_0 : \beta_2 = \cdots = \beta_k$ contra $\text{H}_1 : \beta_j \neq 0$ para pelo menos algum $j = 2, \ldots, k$.

Em forma matricial o modelo fica dado por

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

em que $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_k^\top)^\top$ com $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$, para $i = 1, \ldots, k$, $\boldsymbol{\beta} = (\alpha, \beta_2, \ldots, \beta_k)^\top$ e matriz $\mathbf{X}$ de dimensão $(\sum_{i=1}^k n_i) \times k$ dada abaixo.

A solução de mínimos quadrados leva às estimativas $\widehat{\alpha} = \bar{y}_1$ e $\widehat{\beta}_i = \bar{y}_i - \bar{y}_1$ para $i = 1, \ldots, k$, com variâncias e covariâncias

$$\text{Var}(\widehat{\alpha}) = \frac{\sigma^2}{n_1}, \ \text{Var}(\widehat{\beta}_j) = \sigma^2 \left\{ \frac{1}{n_j} + \frac{1}{n_1} \right\}, \ \text{Cov}(\widehat{\alpha}, \widehat{\beta}_j) = -\frac{\sigma^2}{n_1} \ \text{e} \ \text{Cov}(\widehat{\beta}_j, \widehat{\beta}_\ell) = \frac{\sigma^2}{n_1},$$

para $j \neq \ell = 2, \ldots, k$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \dots & 1 \end{bmatrix}.$$

Tem-se a seguinte decomposição das somas de quadrados:

$$\begin{aligned} \text{SQT} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2, \\ \text{SQReg} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\widehat{y}_{ij} - \bar{y})^2 = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 \quad \text{e} \\ \text{SQRes} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \widehat{y}_{ij})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2. \end{aligned}$$

Daí segue que a estatística F para testar a homogeneidade de médias H_0 : $\beta_2 = \cdots = \beta_k = 0$ contra H_1 : pelo menos duas médias diferentes fica expressa na forma

$$F = \frac{(n-k+1)}{(k-1)} \frac{\sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2} \overset{\text{H}_0}{\sim} F_{(k-1),(n-k+1)}.$$

Rejeita-se H_0 se $\text{F} > \text{F}_{(1-\alpha),(k-1),(n-k+1)}$, em que $\text{F}_{(1-\alpha),(k-1),(n-k+1)}$ denota o quantil $(1-\alpha)$ da distribuição F com $(k-1)$ e $(n-k+1)$ graus de liberdade e $n = n_1 + \cdots + n_k$.

1.9.1 Comparações Múltiplas

Quando rejeita-se a hipótese nula deseja-se saber onde estão as diferenças entre as médias dos k grupos. As propostas mais conhecidas consistem em construir $m = \binom{k}{2}$ estimativas intervalares para as diferenças de médias, de modo que cada estimativa intervalar tenha coeficiente de confiança $(1 - \alpha^*)$ sendo o coeficiente de confiança global $(1 - \alpha)$.

Pelo método de Bonferroni (recomendado para m pequeno) cada estimativa intervalar deve ter coeficiente de confiança $(1 - \alpha^*)$, sendo dadas por

$$(\bar{y}_i - \bar{y}_j) \pm t_{(1-\alpha^*/2),(n-k)} \sqrt{s^2 \left\{ \frac{1}{n_i} + \frac{1}{n_j} \right\}},$$

para $i \neq j$, em que $\alpha^* = \frac{\alpha}{m}$, de modo que o coeficiente global de confiança seja de pelo menos $(1 - \alpha)$.

O método de Tukey é o mais utilizado na prática por ter um nível de significância global mais próximo de $(1 - \alpha)$. As estimativas intervalares são expressas na forma

$$(\bar{y}_i - \bar{y}_j) \pm q(k, n-k) \sqrt{\frac{s^2}{2} \left\{ \frac{1}{n_i} + \frac{1}{n_j} \right\}},$$

para $i \neq j$, em que $q(k, n-k)$ é o quantil de uma distribuição denominada amplitude Studentizada.

1.9.2 Aplicação

Como ilustração serão considerados os dados referentes ao tempo de deslocamento (em minutos) antes de decolar de 184 aeronaves de 8 Cias Aéreas no aeroporto EWR (Newark) no período 1999-2001 (Venzani, 2004, Exemplo 11.7), descritas abaixo e no arquivo **takeoff.txt**

- AA, American Airlines

- CO, Continental Airlines
- DL, Delta Airlines
- HP, American West Airlines
- NW, North West Airlines
- TW, Trans World Airlines
- UA, United Airlines
- US, US Airways.

Na Figura 1.15 tem-se os boxplots robustos dos tempos para a decolagem das Cias Aéreas. Nota-se tempos medianos distintos, porém em geral variabilidades similares. As Cias Aéreas NW e US apresentam os menores tempos medianos enquanto CO apresenta o maior tempo mediano. A fim de comparar os tempos médios supondo variabilidades homogêneas considere o modelo $y_{ij} = \alpha + \beta_i + \epsilon_{ij}$, em que $\epsilon_{ij} \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, 8$ e $j = 1, \ldots, 23$, com a restrição $\beta_1 = 0$. AA como casela de referência.

É bastante razoável esperar pelo TCL que $\widehat{\alpha}$ e $\widehat{\beta}_i$ estejam bem aproximadas pela distribuição normal levando-se em conta o número de réplicas para cada Cia Aérea. Assim, como não há indícios pela Figura 1.15 de afastamentos importantes da suposição de variâncias contantes para os erros, pode-se esperar uma boa aprocimação da distribuiçao nula da estatística F para testar a homogeneidade de médias.

Pela Tabela 1.4 nota-se que o tempo de deslocamento médio de algumas Cias Aéreas é significativamente diferente do tempo médio da Cia AA. Por exemplo, o tempo médio de NW é significativamente menor enquanto o tempo médio de CO é significativamente maior. Porém, para algumas

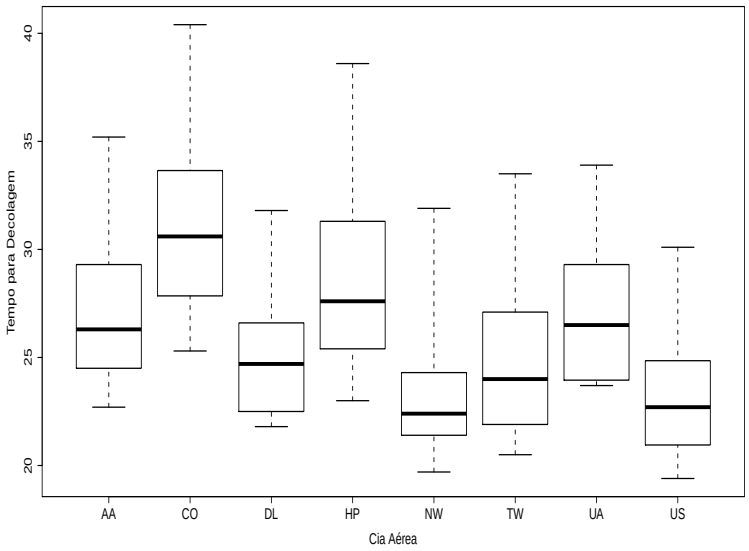

Figura 1.15: Boxplot do tempo de deslocamento segundo a Cia Aérea.

Tabela 1.4: Estimativas dos parâmetros referentes ao modelo de comparação dos tempos médios de deslocamento das Cias Aéreas.

Efeito	Estimativa	valor-t	valor-P
AA	27,056	37,56	0,000
CO	3,835	3,76	0,000
DL	-2,052	-2,01	0,045
HP	1,526	1,50	0,136
NW	-4,061	-3,99	0,000
TW	-1,652	-1,62	0,107
UA	-0,039	-0,04	0,969
US	-3,830	-3.76	0,000
s	3,455		
R^2	0,355		
$\overline{R}^2$	0,329		

Cias Aéreas não foi possível detectar diferença significativa com AA. Isso é confirmado pelo teste F de homogeneidade de médias (vide Tabela 1.5), em que a hipótese nula é fortemente rejeitada. Logo, há tempos médios de deslocamento diferentes e resta saber entre quais Cias Aéreas.

Tabela 1.5: Tabela ANOVA referente à comparação dos tempos médios de deslocamento das Cias Aéreas.

F.Variação	S.Q.	G.L.	Q.M.	F	valor-P
Cia Aérea	1155,0	7	165,01	13,82	0,000
Resíduos	2100,9	176	11,94		
Total	3255,9	183			

Como há $m = \binom{8}{2} = 28$ pares de Cias Aéreas o método de Tukey é o mais adequado para construir as estimativas intervalares para as diferenças das médias. Na Figura 1.16 tem-se um resumo das 28 estimativas intervalares com coeficiente global de confiança de 95%, construída através da biblioteca `UsingR` do R. Nota-se que 15 dessas estimativas intervalares cobrem o valor zero indicando que não foi possível detectar diferença significativa entre os deslocamentos médios das Cias Aéreas correspondentes. Por outro lado, há 13 estimativas intervalares que não cobrem o valor zero. Observando essas estimativas intervalares nota-se que as Cias Aéreas NW e US são aqueleas que mais diferem das demais no sentido de terem um tempo médio de deslocamento menor do que as demais. Isso vai ao encontro dos resultados da Tabela 1.4.

1.10 Regressão Linear Ponderada

Quando há indícios fortes de afastamentos da suposição de variâncias constantes dos erros (homocedasticidade), uma maneira de correção é através da

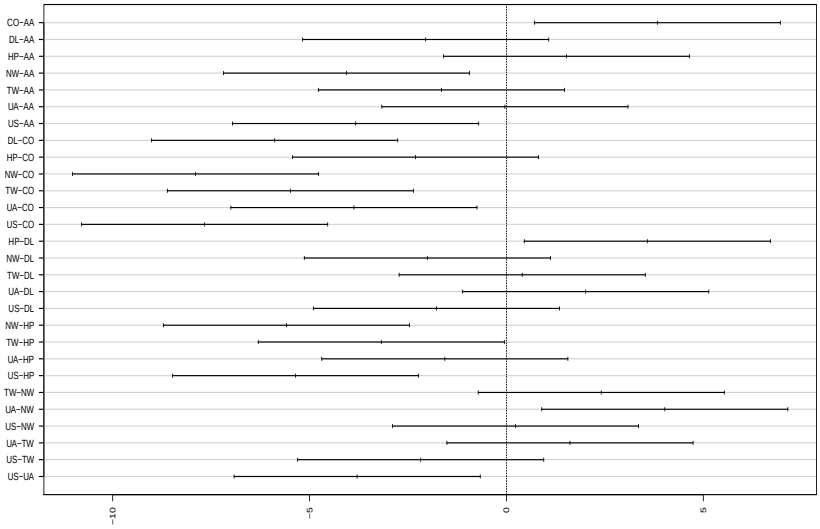

Figura 1.16: Estimativas intervalares para as diferenças entre os deslocamentos médios das Cias Aérea pelo método de Tukey com coeficiente global de confiança de 95%.

regressão linear ponderada em que a variância de cada erro é flexibilizada. A forma mais usual de regressão linear ponderada é a seguinte:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i, \tag{1.5}$$

em que $y_1, \ldots, y_n$ são valores observados da variável resposta, $x_{i1}, \ldots, x_{ip}$ são valores observados de variáveis explicativas e $\epsilon_i \overset{\text{ind}}{\sim} \text{N}(0, \sigma_i^2)$, com $\sigma_i^2 = \frac{\sigma^2}{\omega_i}$ e $\omega_i > 0$ (conhecido), para $i = 1, \ldots, n$. A soma dos quadrados dos erros (função objetivo) fica nesse caso expressa na forma

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \omega_i (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2,$$

em que em que $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ e $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$. Matricialmente tem-se que

$$S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

em que $\mathbf{y} = (y_1, \ldots, y_n)^\top$, $\mathbf{W} = \text{diag}\{\omega_1, \ldots, \omega_n\}$ e $\mathbf{X}$ é a matriz modelo.

Derivando a função objetivo $S(\boldsymbol{\beta})$ em relação a $\boldsymbol{\beta}$ obtém-se

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

que igualando a zero leva à seguinte solução de de mínimos quadrados ponderados:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{W}\mathbf{y}.$$

Denotando $\widehat{\boldsymbol{\beta}} = \mathbf{A}\mathbf{Y}$, em que $\mathbf{A} = (\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{W}$, tem-se a seguinte propriedade:

$$\begin{aligned} \text{E}(\widehat{\boldsymbol{\beta}}) &= \text{E}(\mathbf{A}\mathbf{Y}|\mathbf{X}) = \mathbf{A}\text{E}(\mathbf{Y}|\mathbf{X}) \\ &= (\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{W}\mathbf{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta}. \end{aligned}$$

Logo, $\widehat{\boldsymbol{\beta}}$ é um estimador não tendencioso de $\boldsymbol{\beta}$. Desde que $\text{Var}(\mathbf{Y}|\mathbf{X}) = \sigma^2\mathbf{W}^{-1}$, segue a propriedade

$$\begin{aligned} \text{Var}(\widehat{\boldsymbol{\beta}}) &= \text{Var}(\mathbf{A}\mathbf{Y}|\mathbf{X}) = \mathbf{A}\text{Var}(\mathbf{Y}|\mathbf{X})\mathbf{A}^\top \\ &= \sigma^2(\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{W}\mathbf{W}^{-1}\mathbf{W}\mathbf{X}(\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1}, \end{aligned}$$

e portanto $\widehat{\boldsymbol{\beta}} \sim \text{N}_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1})$.

As somas de quadrados ponderadas ficam expressas nas formas

$$\text{SQT} = \sum_{i=1}^{n} \omega_i(y_i - \bar{y})^2, \;\; \text{SQReg} = \sum_{i=1}^{n} \omega_i(\widehat{y}_i - \bar{y})^2 \;\; \text{e} \;\; \text{SQRes} = \sum_{i=1}^{n} \omega_i(y_i - \widehat{y}_i)^2.$$

Similarmente ao caso homocedástico é possível mostrar que $s^2 = \frac{\text{SQRes}}{(n-p)}$ é um estimador não tendencioso de σ^2. Continuam valendo a decomposição das somas de quadrados e as interpretações do R^2 e $\overline{\text{R}}^2$.

É possível mostrar que o acréscimo na soma de quadrados de resíduos no modelo linear ponderado (5), devido às restrições lineares $\mathbf{R}\boldsymbol{\beta} = \mathbf{0}$, pode ser expresso na forma

$$\text{ASQ}(\mathbf{R}\boldsymbol{\beta} = \mathbf{0}) = (\mathbf{R}\widehat{\boldsymbol{\beta}})^\top \{\mathbf{R}(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{R}^\top\}^{-1}\mathbf{R}\widehat{\boldsymbol{\beta}},$$

em que $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{W}\mathbf{y}$. Assim, se o interesse é testar $\text{H}_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{0}$ contra $\text{H}_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{0}$, a estatística F fica dada por

$$\text{F} = \frac{\text{ASQ}(\mathbf{R}\boldsymbol{\beta} = \mathbf{0})/r}{\text{SQRes}/(n-p)} \overset{\text{H}_0}{\sim} \text{F}_{r,(n-p)}.$$

Rejeita-se H_0 se $\text{F} > \text{F}_{(1-\alpha),r,(n-p)}$, em que $\text{F}_{(1-\alpha),r,(n-p)}$ denota o quantil $(1-\alpha)$ da distribuição F com r e $(n-p)$ graus de liberdade.

1.10.1 Forma Equivalente

Os resultados da regressão linear ponderada (1.5) podem ser obtidos de forma equivalente através de uma regressão linear homocedástica aplicando as seguintes transformações:

- $z_i = y_i\sqrt{\omega_i}$,
- $u_{ij} = x_{ij}\sqrt{\omega_i}$,

para $i = 1, \ldots, n$ e $j = 1, \ldots, p$. Então, considere o modelo

$$z_i = \beta_1 u_{i1} + \beta_2 u_{i2} + \cdots + \beta_p u_{ip} + e_i,$$

com $e_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$. Esse modelo em forma matricial fica dado por

$$\mathbf{z} = \mathbf{U}\boldsymbol{\beta} + \mathbf{e},$$

em que $\mathbf{z} = \mathbf{W}^{\frac{1}{2}}\mathbf{y}$, $\mathbf{U} = \mathbf{W}^{\frac{1}{2}}\mathbf{X}$ é a matriz modelo, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, enquanto $\mathbf{e} = \mathbf{W}^{\frac{1}{2}}\boldsymbol{\epsilon}$. Note que $\mathbf{e} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2\mathbf{I_n})$. Mostra-se facilmente que $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{W}\mathbf{y}$. Assim, todos os resultados descritos nas seções anteriores podem ser estendidos facilmente para o modelo (1.5) através das transformações acima.

1.10.2 Aplicação

Como ilustração considere parte dos dados de um experimento desenvolvimento em 2006 nas Faculdades de Medicina e de Filosofia, Letras e Ciências Humanas da USP e analisado no Centro de Estatística Aplicada do IME-USP (CEA0P16) para avaliar o fluxo da fala de falantes do Português Brasileiro segundo o gênero, idade e escolaridade. Uma amostra de 595 indivíduos residentes na cidade de São Paulo com idade entre 2 e 99 anos foi avaliada segundo a fala auto-expressiva. O indivíduo era apresentado a uma figura e orientado a discorrer sobre a mesma durante um tempo mínimo de 3 minutos e máximo de 6 minutos. Para crianças de 2 e 3 anos, as amostras foram obtidas com a colaboração dos pais. As variáveis consideradas no estudo foram as seguintes: (i) idade (em anos), (ii) gênero (1:feminino, 2:masculino), (iii) interj (número de interjeições durante o discurso), (iv) fpm (fluxo de palavras por minuto) e (v) fsm (fluxo de sílabas por minuto). Descrição no arquivo **fluxo.txt**.

Como aplicação de regressão linear ponderada considere apenas duas variáveis, fpm e fsm. Na Figura 1.17 tem-se o diagrama de dispersão entre fpm e fsm e nota-se uma forte relação linear positiva e variabilidade não constante da resposta fpm à medida que aumenta fsm. Isso sugere um modelo linear simples entre fpm e fsm. Nas Tabelas 1.6 e 1.7 tem-se as estimativas

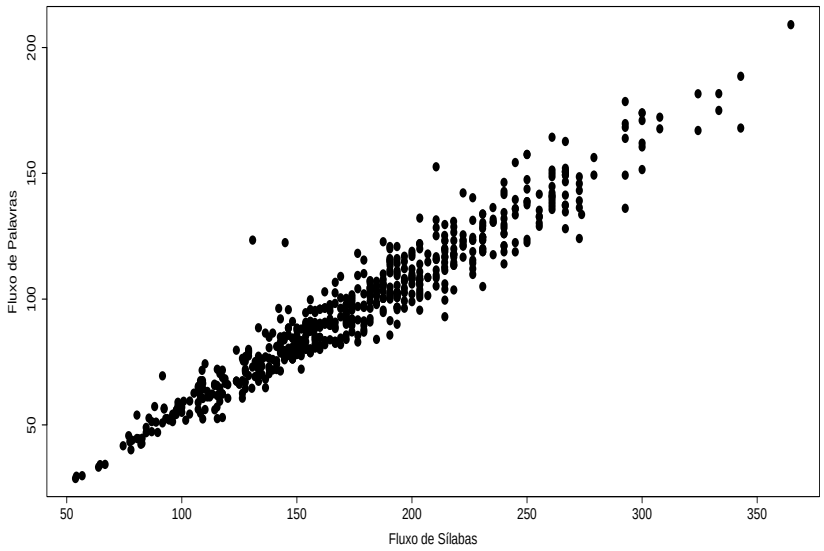

Figura 1.17: Diagrama de dispersão entre o fluxo de palavras por minuto e o fluxo de sílabas por minuto.

dos parâmetros do modelo

$$\text{fpm}_i = \beta_1 + \beta_2 \text{fsm}_i + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$ ou $\epsilon_i \overset{\text{ind}}{\sim} \text{N}(0, \omega_i^{-1}\sigma^2)$ com $\omega_i = 1/\text{fsm}_i$, respectivamente, para $i = 1, \ldots, 594$. Nota-se uma redução na estimativa do intercepto e aumento do coeficiente de determinação sob o modelo linear ponderado. Há também um controle melhor da variabilidade sob esse modelo (Figura 1.18) e melhora na qualidade do ajuste (Figura 1.19). As três observações que aparecem destacadas como pontos aberrantes afetam muito pouco as estimativas quando são excluídas. Outros procedimentos para aprimoramento do controle da variabilidade poderiam ser aplicados, como por exemplo a modelagem dupla da média e variância.

Tabela 1.6: Estimativas dos parâmetros referentes ao modelo de regressão linear simples ajustado aos dados sobre fluxo da fala de falantes do Português Brasileiro.

Efeito	Estimativa	E.Padrão	valor-t	valor-P
Constante	4,198	1,172	3,74	0,00
fsm	0,527	0,006	88,10	0,00
s	7,98			
R^2	0,93			
$\overline{\mathrm{R}}^2$	0,93			

Tabela 1.7: Estimativas dos parâmetros referentes ao modelo de regressão linear simples ponderado ajustado aos dados sobre fluxo da fala de falantes do Português Brasileiro.

Efeito	Estimativa	E.Padrão	valor-t	valor-P
Constante	3,663	0,974	3,76	0,00
fsm	0,530	0,006	92,57	0,00
s	0,59			
R^2	0,99			
$\overline{\mathrm{R}}^2$	0,99			

1.11 Ortogonalidade

Supor novamente o modelo de regressão linear múltipla

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i,$$

em que $y_1, \ldots, y_n$ são valores observados da variável resposta, $x_{i1}, \ldots, x_{ip}$ são valores observados de variáveis explicativas e $\epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$. Tem-se ortogonalidade entre as colunas da matriz modelo $\mathbf{X}$ se

$$\sum_{i=1}^{n} x_{ij} x_{i\ell} = 0, \ \forall j \neq \ell = 1, \ldots, p,$$

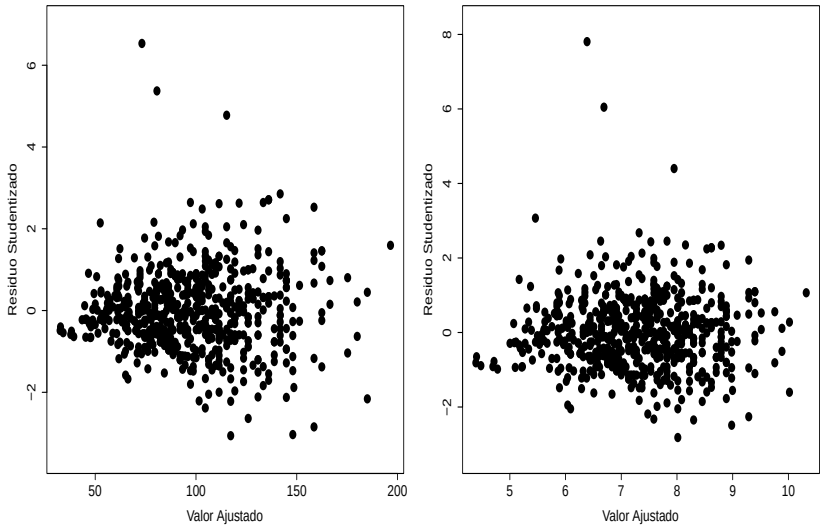

Figura 1.18: Gráficos entre o resíduo Studentizado e o valor ajustado referentes aos modelos homocedástico (esquerdo) e ponderado (direito) ajustados aos dados spbre fluxo da fala de falantes do Português Brasileiro.

ou seja, a matriz $\mathbf{X}^\top\mathbf{X}$ é bloco diagonal.

Quando a matriz modelo $\mathbf{X}$ tem posto coluna completo tem-se sob ortogonalidade que

$$\mathbf{X}^\top\mathbf{X} = \text{diag}\{\sum_{i=1}^{n} x_{i1}^2, \ldots, \sum_{i=1}^{n} x_{ip}^2\} \quad \text{e} \quad \mathbf{X}^\top\mathbf{y} = (\sum_{i=1}^{n} x_{i1}y_i, \ldots, \sum_{i=1}^{n} x_{ip}y_i)^\top,$$

em que $\mathbf{y} = (y_1, \ldots, y_n)^\top$, e consequentemente

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_1 \\ \vdots \\ \widehat{\beta}_p \end{bmatrix} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} = \begin{bmatrix} \frac{\sum_{i=1}^{n} x_{i1}y_i}{\sum_{i=1}^{n} x_{i1}^2} \\ \vdots \\ \frac{\sum_{i=1}^{n} x_{ip}y_i}{\sum_{i=1}^{n} x_{ip}^2} \end{bmatrix}.$$

Logo, $\widehat{\beta}_j = \frac{\sum_{i=1}^{n} x_{ij}y_i}{\sum_{i=1}^{n} x_{ij}^2}$ depende apenas dos valores $y_1, \ldots, y_n$ e de $x_{1j}, \ldots, x_{nj}$,

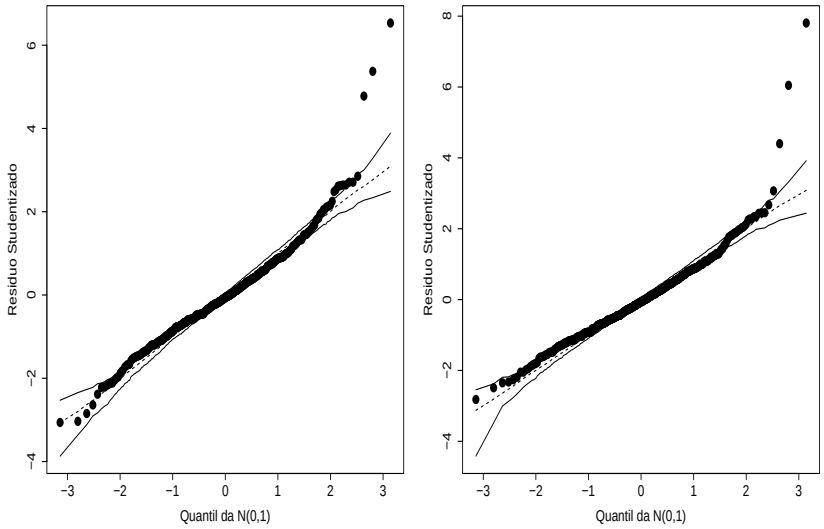

Figura 1.19: Gráficos normais de probabilidade com banda empírica de 95% referentes aos modelos homocedástico (esquerdo) e ponderado (direito) ajustados aos dados spbre fluxo da fala de falantes do Português Brasileiro.

para $j = 1, \ldots, p$. Ou seja, dos valores da variável resposta e da variável explicativa X_j.

Além disso, a matriz de variância-covariância para $\widehat{\boldsymbol{\beta}}$ fica dada por

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} = \begin{bmatrix} \frac{\sigma^2}{\sum_{i=1}^n x_{i1}^2} & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & \frac{\sigma^2}{\sum_{i=1}^n x_{ip}^2} \end{bmatrix}.$$

Portanto, $\mathrm{Var}(\widehat{\beta}_j) = \frac{\sigma^2}{\sum_{i=1}^n x_{ij}^2}$ e $\mathrm{Cov}(\widehat{\beta}_j, \widehat{\beta}_\ell) = 0$, para $j \neq \ell$ e $j, \ell = 1, \ldots, p$. Tem-se, portanto, ausência de correlação linear entre os estimadores dos coeficientes, que sob normalidade implica na independência probabilística mútua entre os estimadores dos coeficientes.

1.12 Multicolinearidade

Multicolinearidade é o oposto da ortogonalidade. Ocorre quando há uma alta correlação linear entre variáveis explicativas e consequentemente entre os estimadores dos coeficientes da regressão linear múltipla. Uma consequência prática é que $\det(\mathbf{X}^\top\mathbf{X}) \cong 0$. Agumas fontes de multicolinearidade são as seguintes:

- Método empregado na coleta de dados
  Os dados são coletados de um estrato da população onde há uma alta correlação linear entre duas variáveis explicativas. Por exemplo, num estudo de regressão em que tem-se como variáveis explicativas o consumo de um produto alimentício e o preço do produto alimentício. É razoável esperar nos estratos de renda mais baixa uma correlação mais alta entre as duas variáveis explicativas.

- Restrições no modelo ou na população
  Duas variáveis explicativas que têm uma correlação linear alta são incluídas no modelo. Por exemplo, consumo de energia elétrica e renda percapita. Notas referentes às avaliações sobre qualidade e clareza das aulas de um instrutor.

- Especificação do modelo
  No modelo são incluídos vários termos que estão em função de uma mesma variável explicativa. Por exemplo, numa regressão polinomial em que são incluídos termos $x + x^2 + x^3 + \cdots$.

- Modelo superdimensionado
  Estudos com amostras pequenas e uma grande quantidade de variáveis

explicativas. Por exemplo, na área médica em geral tem-se amostras pequenas com uma grande quantidade de informações por paciente.

1.12.1 Efeitos da Multicoliearidade

Para ilustrar considere o seguinte modelo de regressão linear múltipla:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i,$$

em que $y_1, \ldots, y_n$ são valores observados da variável resposta com comprimento unitário, x_{i1} e x_{i2} são valores observados de variáveis explicativas com comprimento unitário, em que $\sum_{i=1}^n x_{ij} = 0$ e $\sum_{i=1}^n x_{ij}^2 = 1$ para $j = 1, 2$, e $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$.

Para esse exemplo tem-se que

$$\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1} x_{i2} \\ \sum_{i=1}^n x_{i2} x_{i1} & \sum_{i=1}^n x_{i2}^2 \end{bmatrix} = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix},$$

em que r_{12} denota a correlação linear amostral entre X_1 e X_2. Além disso

$$\mathbf{X}^\top \mathbf{y} = \begin{bmatrix} \sum_{i=1}^n x_{i1} y_i \\ \sum_{i=1}^n x_{i2} y_i \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix},$$

em que r_{1y} e r_{2y} denotam, respectivamente, as correlações lineares amostrais entre X_1 e Y e X_2 e Y. Portanto, as estimativas de mínimos quadrados ficam dadas por

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \begin{bmatrix} \frac{r_{1y} - r_{12} r_{2y}}{(1 - r_{12}^2)} \\ \frac{r_{2y} - r_{12} r_{1y}}{(1 - r_{12}^2)} \end{bmatrix},$$

e dependem das correlações lineares r_{12}, r_{1y} e r_{2y}. Além disso, a matriz de variância-covariância para $\widehat{\boldsymbol{\beta}}$ assume a forma

$$\text{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} = \begin{bmatrix} \frac{\sigma^2}{(1 - r_{12}^2)} & -\frac{\sigma^2 r_{12}}{(1 - r_{12}^2)} \\ -\frac{\sigma^2 r_{12}}{(1 - r_{12}^2)} & \frac{\sigma^2}{(1 - r_{12}^2)} \end{bmatrix}.$$

Ou seja, $\text{Var}(\widehat{\beta}_1) = \text{Var}(\widehat{\beta}_2) = \frac{\sigma^2}{(1 - r_{12}^2)}$ e $\text{Cov}(\widehat{\beta}_1, \widehat{\beta}_2) = -\frac{\sigma^2 r_{12}}{(1 - r_{12}^2)}$. E tem-se as seguintes consequências:

- Se $|r_{12}| \to 1$ então $\text{Var}(\widehat{\beta}_1)$ e $\text{Var}(\widehat{\beta}_2)$ ficam grandes.
- Se $r_{12} \to 1$ então $\text{Cov}(\widehat{\beta}_1, \widehat{\beta}_2) \to -\infty$.
- Se $r_{12} \to -1$ então $\text{Cov}(\widehat{\beta}_1, \widehat{\beta}_2) \to \infty$.

1.12.2 Procedimentos para Detectar Multicolinearidade

Fator de Inflação da Variância

É possível mostrar que

$$\text{Var}(\widehat{\beta}_j) = \sigma^2 \text{C}_{jj} = \sigma^2 (1 - \text{R}_j^2)^{-1},$$

em que $\text{C}_{j\ell}$ denota o (j, ℓ)-ésimo elemento da matriz $\mathbf{C} = (\mathbf{X}^\top \mathbf{X})^{-1}$ e R_j^2 denota o coeficiente de determinação da regressão linear da variável explicativa X_j contra as demais variáveis explicativas X_ℓ, em que $j \neq \ell$, para $j, \ell = 1, \ldots, p$. O fator de inflação de variância da j-ésima variável explicativa é definido por

$$\text{VIF}_j = (1 - \text{R}_j^2)^{-1}.$$

Assim, se $\text{R}_j^2 \to 1$ então $\text{VIF}_j \to \infty$, para $j = 1, \ldots, p$. Para ilustrar, supor três variáveis explicativas X_1, X_2 e X_3 cujos valores amostrais têm comprimento unitário. Os VIFs saem das seguintes regressões:

- VIF_1: da regressão $x_{i1} = \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$
- VIF_2: da regressão $x_{i2} = \beta_1 x_{i1} + \beta_3 x_{i3} + \epsilon_i$
- VIF_3: da regressão $x_{i3} = \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$, para $i = 1, \ldots, n$.

Critério: $\text{VIF}_j \geq 10$ indica que $\widehat{\beta}_j$ está com variância inflacionada.

Número da Condição

Sejam $\lambda_1, \ldots, \lambda_p$ os autovalores da matrix $\mathbf{X}^\top\mathbf{X}$. Como é uma matriz simétrica positiva definida todos os seus autovalores são não negativos. Contudo, a existência de autovalores próximos de zero é indício de multicolinearidade. Uma medida resumo de multicolinearidade entre as colunas da matriz $\mathbf{X}$ é o número da condição definido por

$$k = \frac{\lambda_{\max}}{\lambda_{\min}}.$$

Portanto, se esta razão é muito grande há indícios de multicolinearidade com a matriz $\mathbf{X}^\top\mathbf{X}$.

Critério: (i) se $k \leq 100$ não há indícios de multicolinearidade, (ii) se $100 < k \leq 1000$ há indícios moderados de multicolinearidade e (iii) se $k > 1000$ há indícios fortes de multicolinearidade.

Índice da Condição

Quando há indícios de multicolinearidade através do número da condição, pode-se avaliar a contribuição de cada variável explicativa através do índice da condição definido por

$$k_j = \frac{\lambda_{\max}}{\lambda_j},$$

para $j = 1, \ldots, p$. Os mesmos critérios usados para o número da condição são usados para o índice da condição.

Determinante da Matrix $\mathbf{X}^\top\mathbf{X}$

Se as variáveis explicativas têm comprimento unitário, mostra-se que

$$0 \leq \det(\mathbf{X}^\top\mathbf{X}) \leq 1.$$

Logo, $\det(\mathbf{X}^\top\mathbf{X}) = 1$ indica ortogonalidade entre as colunas da matriz $\mathbf{X}$, enquanto $\det(\mathbf{X}^\top\mathbf{X}) = 0$ indica dependência linear entre as colunas da matrix $\mathbf{X}$. Valores próximos de zero são indícios de multicolinearidade.

1.12.3 Tratamentos da Multicolineridade

Alguns tratamentos para a multicolinearidade

- Coletar mais dados.
- Eliminação de variáveis explicativas.
- Transformação de variáveis explicativas.
- Regressão *ridge*.
- Regressão através de componentes principais.

Regressão *ridge*

O objetivo da regressão *ridge* é utilizar um estimador tendencioso que produza variâncias mais estáveis para os estimadores dos coeficientes da regressão. Assim, seja $\widehat{\boldsymbol{\beta}}^*$ um estimador tendencioso de $\boldsymbol{\beta}$. Mostra-se que o erro quadrático médio de $\widehat{\boldsymbol{\beta}}^*$ pode ser expresso na forma

$$\text{EQM}(\widehat{\boldsymbol{\beta}}^*) = \text{Var}(\widehat{\boldsymbol{\beta}}^*) + [\text{Viés}][\text{Viés}]^\top,$$

em que $\text{Viés} = \text{E}(\widehat{\boldsymbol{\beta}}^*) - \boldsymbol{\beta}$. A fim de estabilizar as estimativas dos coeficientes da regressão linear múltipla bem com as respectivas variâncias é proposto o seguinte estimador:

$$\widehat{\boldsymbol{\beta}}_R = (\mathbf{X}^\top\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}^\top\mathbf{y},$$

em que $k > 0$ é uma constante desconhecida que é estimada separadamente. Em particular quando $k = 0$ recupera-se o estimador de mínimos quadrados. Estima-se k até estabilizar as estimativas dos coeficientes. Na Figura 1.20 tem-se um exemplo ilustrativo em que quatro coeficientes estão sendo ajustados e nota-se uma estabilidade das estimativas a partir de $k = 0, 10$.

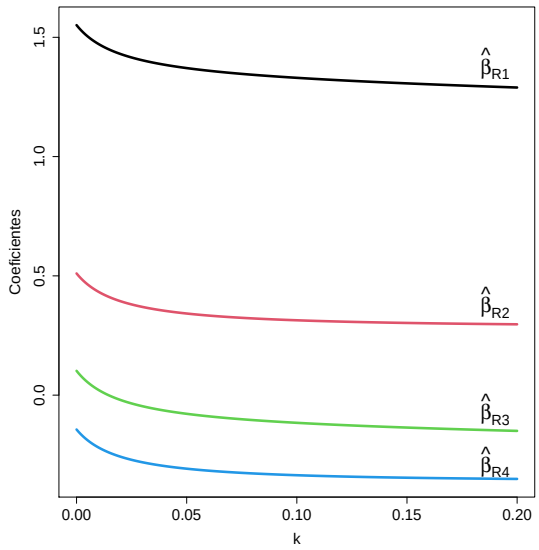

Figura 1.20: Ilustração dos coeficientes estimados através da regressão *ridge* variando-se o valor de k.

Denotando $\widehat{\boldsymbol{\beta}}_R = \mathbf{Z}_k\widehat{\boldsymbol{\beta}}$, em que $\mathbf{Z}_k = (\mathbf{X}^\top\mathbf{X} + k\mathbf{I}_p)^{-1}(\mathbf{X}^\top\mathbf{X})$, tem-se as seguintes propriedades:

- $\mathrm{E}(\widehat{\boldsymbol{\beta}}_R) = \mathrm{E}(\mathbf{Z}_k\widehat{\boldsymbol{\beta}}) = \mathbf{Z}_k\mathrm{E}(\widehat{\boldsymbol{\beta}}) = \mathbf{Z}_k\boldsymbol{\beta}$.
- $\mathrm{Var}(\widehat{\boldsymbol{\beta}}_R) = \mathrm{Var}(\mathbf{Z}_k\widehat{\boldsymbol{\beta}}) = \mathbf{Z}_k\mathrm{Var}(\widehat{\boldsymbol{\beta}})\mathbf{Z}_k^\top = \sigma^2\mathbf{Z}_k(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{Z}_k^\top$.

Em particular, se $\mathbf{X}^\top\mathbf{X} = \mathbf{I}_p$ tem-se que $\mathbf{Z}_k = (1+k)^{-1}\mathbf{I}_p$. Logo, $\mathrm{E}(\widehat{\boldsymbol{\beta}}_R) = (1+k)^{-1}\boldsymbol{\beta}$ e $\mathrm{Var}(\widehat{\boldsymbol{\beta}}_R) = \sigma^2(1+k)^{-2}\mathbf{I}_p$. Ou seja, à medida que k cresce o

estimador *ridge* fica mais tendencioso havendo um encolhimento com relação ao estimador de mínimos quadrados. A variância diminui com o aumento de k.

Tem-se ainda que $\widehat{\boldsymbol{\beta}}_R \sim \mathrm{N}_p(\mathrm{E}(\widehat{\boldsymbol{\beta}}_R), \mathrm{Var}(\widehat{\boldsymbol{\beta}}_R))$. Daí segue que $\widehat{\beta}_{R_j}$ são normais de média $\mathrm{E}(\widehat{\beta}_{R_j})$ e variância $\mathrm{Var}(\widehat{\beta}_{R_j})$, para $j = 1, \ldots, p$. É possível mostrar que

$$\begin{aligned} \mathrm{SQRes}(k) &= (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_R)^\top (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_R) \\ &= \mathrm{SQRes} + (\widehat{\boldsymbol{\beta}}_R - \widehat{\boldsymbol{\beta}})^\top (\mathbf{X}^\top \mathbf{X})(\widehat{\boldsymbol{\beta}}_R - \widehat{\boldsymbol{\beta}}), \end{aligned}$$

em que SQRes denota a soma de quadrados de resíduos da regressão de mínimos quadrados. Portanto, na regressão *ridge* há um aumento na soma de quadrados de resíduos, logo uma redução no valor de R^2.

A constante k pode ser estimada através do processo iterativo

$$k^{(m+1)} = \frac{p\widehat{\sigma}^2}{\widehat{\boldsymbol{\beta}}_R^\top(k^{(m)})\widehat{\boldsymbol{\beta}}_R(k^{(m)})},$$

para $m = 0, 1, \ldots$, em que $\widehat{\sigma}^2$ é obtido através do estimador de mínimos quadrados $\widehat{\boldsymbol{\beta}}$. Para valor inicial utiliza-se o estimador de HKB (Montgomery et al., 2021, Cap.9) dado por $k^{(0)} = p\widehat{\sigma}^2/\widehat{\boldsymbol{\beta}}^\top\widehat{\boldsymbol{\beta}}$.

Regressão dos Componentes Principais

A forma canônica da regressão linear múltipla $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ é definida por

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon},$$

em que $\boldsymbol{\epsilon} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$, $\mathbf{Z} = \mathbf{X}\mathbf{T}$, $\boldsymbol{\alpha} = \mathbf{T}^\top\boldsymbol{\beta}$ e $\mathbf{Z}^\top\mathbf{Z} = \mathbf{T}^\top\mathbf{X}^\top\mathbf{X}\mathbf{T} = \boldsymbol{\Lambda}$, com $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_p\}$ sendo a matriz diagonal $p \times p$ com os autovalores da matriz $\mathbf{X}^\top\mathbf{X}$ e $\mathbf{T}$ a matriz $p \times p$ cujas colunas são os autovetores ortonormais (ortogonais com comprimento unitário) correspondentes aos autovalores

$\lambda_1, \ldots, \lambda_p$. Como $\mathbf{T}$ é uma matriz ortonormal tem-se que $\mathbf{T}^\top = \mathbf{T}^{-1}$, e daí segue que $\boldsymbol{\beta} = \mathbf{T}\alpha$. Sugere-se que $\mathbf{y}$ e a matriz $\mathbf{X}$ sejam centralizadas, assim não precisa de intercepto.

Portanto, a estimativa de mínimos quadrados de $\boldsymbol{\alpha}$ fica dada por

$$\begin{aligned} \widehat{\boldsymbol{\alpha}} &= (\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{y} \\ &= \boldsymbol{\Lambda}^{-1}\mathbf{Z}^\top\mathbf{y}, \end{aligned}$$

com matriz de variância-covariância expressa na forma

$$\begin{aligned} \mathrm{Var}(\widehat{\boldsymbol{\alpha}}) &= \sigma^2(\mathbf{Z}^\top\mathbf{Z})^{-1} \\ &= \sigma^2\boldsymbol{\Lambda}^{-1}. \end{aligned}$$

Daí segue que $\mathrm{Var}(\widehat{\alpha}_j) = \sigma^2\lambda_j^{-1}$. Assim, λ_j próximo de zero inflaciona a variância de $\widehat{\alpha}_j$. Similarmente, segue que a matriz de variância-covariância de $\widehat{\boldsymbol{\beta}}$ pode ser expressa na forma

$$\begin{aligned} \mathrm{Var}(\widehat{\boldsymbol{\beta}}) &= \mathrm{Var}(\mathbf{T}\widehat{\boldsymbol{\alpha}}) \\ &= \mathbf{T}\mathrm{Var}(\widehat{\boldsymbol{\alpha}})\mathbf{T}^\top \\ &= \sigma^2\mathbf{T}\boldsymbol{\Lambda}^{-1}\mathbf{T}^\top. \end{aligned}$$

E daí pode-se mostrar que $\mathrm{Var}(\widehat{\beta}_j) = \sigma^2\sum_{\ell=1}^{p} t_{j\ell}^2/\lambda_\ell$, em que $t_{j\ell}$ denota o (j,ℓ)-ésimo elemento da matriz $\mathbf{T}$. Esse resultado confirma o efeito de autovalores próximos de zero na inflação da variância de $\widehat{\beta}_j$.

A partir da relação $\widehat{\boldsymbol{\beta}} = \mathbf{T}\widehat{\boldsymbol{\alpha}}$, a proposta da regressão dos componentes principais é considerar os coeficientes estimados

$$\widehat{\boldsymbol{\beta}}^{CP} = \mathbf{T}\widehat{\boldsymbol{\alpha}}^{CP},$$

em que $\widehat{\boldsymbol{\alpha}}^{CP}$ é um vetor $p \times 1$ que contém os coeficientes estimados correspondentes aos $p - s$ maiores autovalores da matriz $\mathbf{X}^\top\mathbf{X}$ e os demais s

coeficientes como sendo iguais a zero. Assim, os novos coeficientes estimados $\widehat{\beta}_1^{CP}, \ldots, \widehat{\beta}_p^{CP}$ irão depender apenas das variáveis explicativas com menor potencial de estarem causando multicolinearidade. Esses coeficients estimados são interpretados de forma similar aos coeficientes estimados por mínimos quadrados.

Da relação $\mathbf{Z} = \mathbf{XT}$ segue que $\mathbf{Z}_j = \sum_{\ell=1}^{p} \mathbf{X}_\ell t_{\ell j}$, em que $\mathbf{Z}_1, \ldots, \mathbf{Z}_p$ e $\mathbf{X}_1, \ldots, \mathbf{X}_p$ denotam, respectivamente, as colunas de $\mathbf{Z}$ e $\mathbf{X}$, enquanto $t_{1j}, \ldots, t_{pj}$ denotam os componentes do autovetor correspondente ao autovalor λ_j. Assim, se λ_j for próximo de zero os componentes de $\mathbf{Z}_j$ devem ser aproximadamente constantes. Deve-se portanto escolher os $p - s$ componentes principais $\mathbf{Z}_1, \ldots, \mathbf{Z}_{(p-s)}$ que correspondem aos $p - s$ maiores autovalores.

1.12.4 Aplicação

Como ilustração para o tópico de multicolinearidade será analisado um conjunto de dados proposto em Montgomery et al. (2021, Tabela B.21) em que o calor (em calorias por grama) de $n = 13$ amostras de cimento é relacionado com as seguintes variáveis explicativas referentes a ingredientes usados na mistura do cimento:

- X_1: aluminato tricálcico
- X_2: silicato tricálcico
- X_3: aluminato-ferrita tetracálcico
- X_4: silicato dicálcico.

Os dados estão descrito no arquivo **cimento.txt**. Nota-se pela Tabela 1.8 correlações lineares altas entre a resposta calor do cimento e as variáveis

Tabela 1.8: Matriz de correlações lineares amostrais de Pearson entre as variáveis do exemplo sobre o calor do cimento em amostras de cimento.

	Calor	X_1	X_2	X_3	X_4
Calor	1,00	0,73	0,82	-0,54	-0,82
X_1		1,00	0,23	-0,82	-0,25
X_2			1,00	-0,14	-0,97
X_3				1,00	0,03
X_4					1,00

explicativas X_2 e X_4, enquanto entre as variáveis explicativas nota-se correlação liner muito alta entre X_2 e X_4, indicando possível multicolineridade nos dados. Nota-se pelo boxplot robusto da Figura 1.21 que a distribuição da variável resposta é aproximadamente simétrica, enquanto os diagramas de dispersão da Figura 1.22 confirmam as correlações lineares descritas na Tabela 1.8.

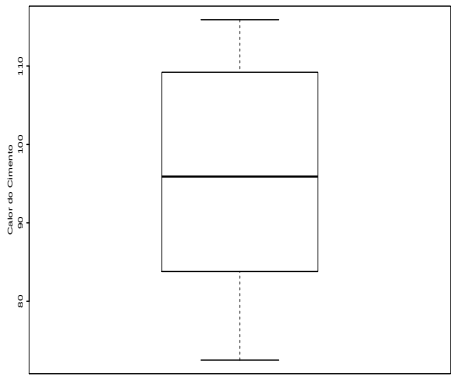

Figura 1.21: Boxplot robusto da variável resposta calor do cimento.

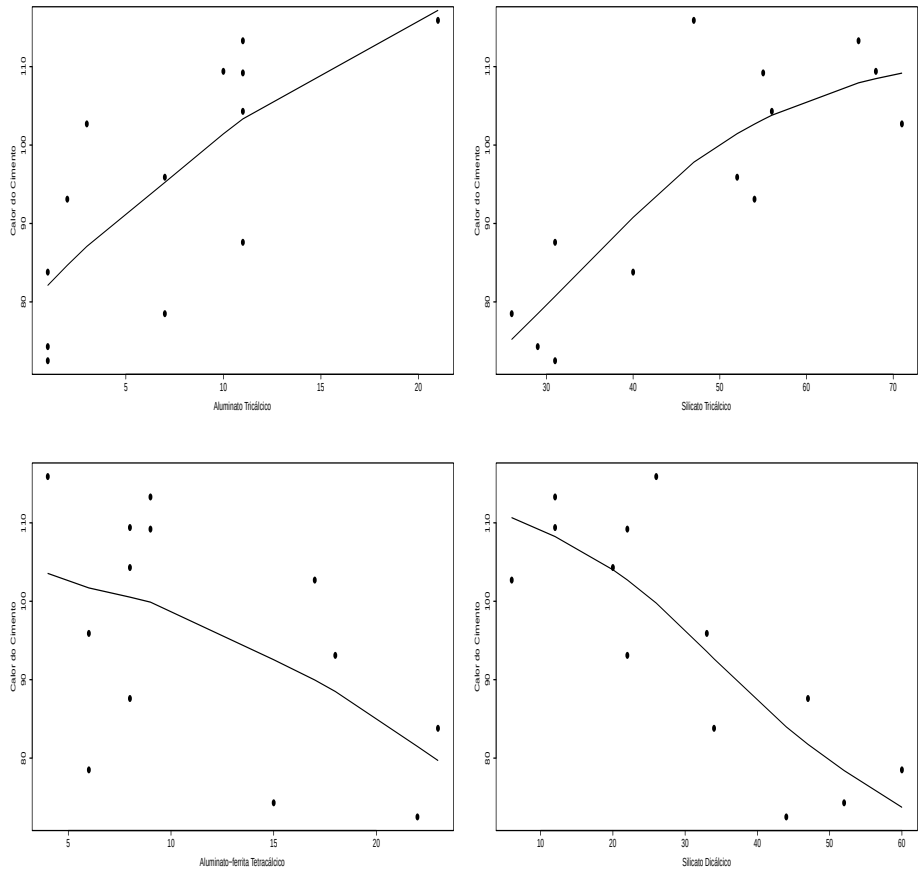

Figura 1.22: Diagramas de dispersão (com tendência) entre a variável resposta calor do cimento e as demais variáveis explicativas.

Com base nos diagramas de dispersão o seguinte modelo é proposto:

$$cy_i = \beta_1 \mathrm{cx}_{i1} + \beta_2 \mathrm{cx}_{i2} + \beta_3 \mathrm{cx}_{i3} + \beta_4 \mathrm{cx}_{i4} + \epsilon_i,$$

em que cy_i denota o calor da i-ésima amostra de cimento centralizada (subtraído da média amostral), bem como os valores das variáveis explicativas e $\epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, para $i = 1, \ldots, 13$. Dessa forma, não é necessário incluir o intercepto.

Pela Tabela 1.9 apenas a variável X_1 é marginalmente significativa. Os gráficos de resíduos são apresentados na Figura 1.23, não havendo indícios de afastamentos da normalidade, de presença de observações aberrantes e de variância não constante dos erros. Como a amostra é pequena a suposição de normalidade dos erros é crucial para fazer inferência. A observação #8 aparece como possivelmente influente no gráfico da distância de Cook com $k = 2$ (Figura 1.24). Quando essa observação não é considerada na regressão o valor-P correspondente à estimativa do coeficiente da variável X_1 reduz para 0,02, porém os demais coeficientes continuam não significativos e todos com sinal positivo.

Tabela 1.9: Estimativas dos parâmetros referentes ao modelo de regressão linear ajustado aos dados sobre o calor do cimento em amostras de cimento.

Efeito	Estimativa	E.Padrão	valor-t	valor-P
cx_1	1,551	0,702	2,21	0,06
cx_2	0,510	0,602	0,75	0,47
cx_3	0,102	0,716	0,14	0,89
cx_4	-0,144	0,669	-0,22	0,83
s	2,31			
R^2	0,98			
$\overline{R}^2$	0,97			

Na Tabela 1.10 tem-se os VIFs corresponentes às 4 variáveis explicativos, confirmando os indícios de multicolinearidade. As estimativas da regressão *ridge* com $k = 0,076$ (vide comportamento dos coeficientes etimados na Figura 1.20) apresenta estimativas mais coerentes com a análise descritiva, porém apenas a variável explicativa X_1 é marginalmente significativa. Os autovalores da matriz $\mathbf{X}^\top\mathbf{X}$ são respectivamente dados por $\lambda_1 = 6213,56$, $\lambda_2 = 809,96$, $\lambda_3 = 148,86$ e $\lambda_4 = 2,84$ com autovalores ortonormais dados

Tabela 1.10: Fator de inflação da variância das variáveis explicativas do modelo de regressão linear ajustado aos dados sobre o calor do cimento em amostras de cimento.

Variável	VIF
cx_1	38,49
cx_2	254,42
cx_3	46,87
cx_4	282,51

Tabela 1.11: Estimativas dos parâmetros referentes ao modelo de regressão *ridge* ajustado aos dados sobre o calor do cimento em amostras de cimento.

Efeito	Estimativa	Erro padrão	valor-z
cx_1	1,3460	0,6844	1,967
cx_2	0,3236	0,6651	0,486
cx_3	-0,1018	0,6934	-0,147
cx_4	-0,3263	0,6514	-0,501

abaixo.

$\mathbf{T}_1$	$\mathbf{T}_2$	$\mathbf{T}_3$	$\mathbf{T}_4$
-0,067800	0,646018	-0,567315	0,506180
-0,678516	0,019993	0,543969	0.493268
0,029021	-0,755310	-0,403554	0,515567
0,730874	0,108480	0,468398	0,484416

Considerando apenas o primeiro componente principal, que explica 86,60%, tem-se a seguinte relação:

$$z_1 = -0,067800cx_1 - 0,678516cx_2 + 0,029021cx_3 + 0,730874cx_4.$$

Com base nos diagramas de dispersão da Figura 1.22, o componente z_1 aumenta à medida que os valores das variáveis explicativas diminuem. O modelo

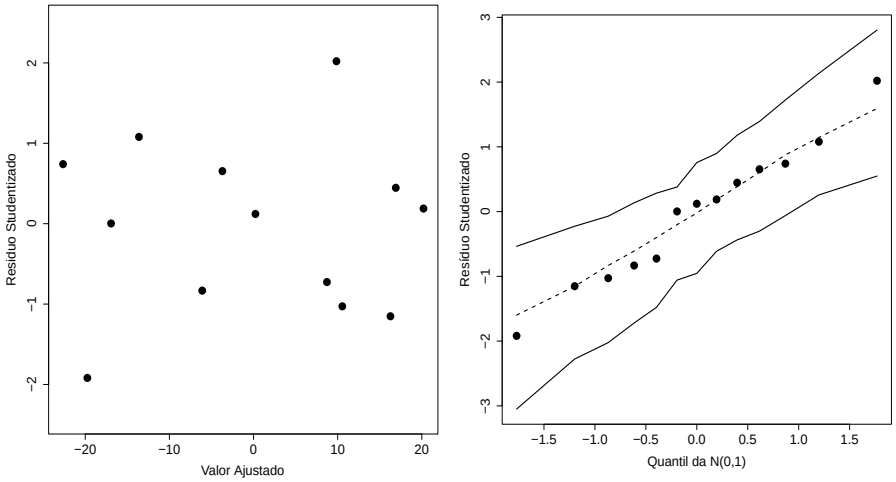

Figura 1.23: Gráficos de resíduos referentes ao ajuste do modelo de regressão linear aos dados sobre o calor do cimento em amostras de cimento.

na forma canônica fica dado por

$$cy_i = z_{i1}\alpha + \epsilon_i,$$

em que cy_i denota o calor da i-ésima amostra de cimento centralizado e $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, 13$. Desse ajuste obtém-se $\widehat{\alpha} = -0,5537(0,1043)$, que é altamente significativo. Assim, espera-se aumento do calor do cimento à medida que aumenta z_1.

1.13 Seleção de Modelos

A seleção de modelos consiste em uma etapa importante e também complexa na análise de regressão, principalmente quando há um grande número de variáveis explicativas candidatas a entrarem no modelo. O fato das variáveis explicativas em geral estarem correlacionadas dificulta a seleção de um sub-

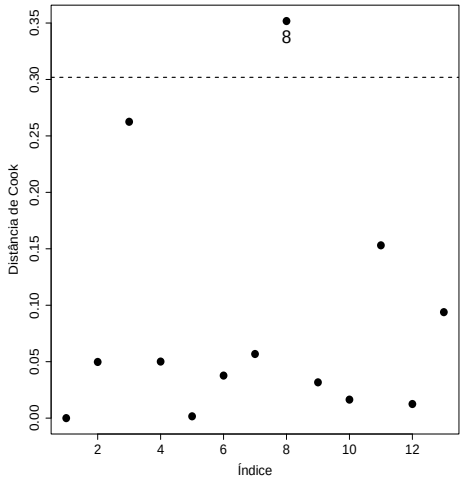

Figura 1.24: Gráfico da distância de Cook contra a ordem das observações referente ao ajuste do modelo de regressão linear aos dados sobre o calor do cimento em amostras de cimento.

conjunto de coeficientes que além de serem significativos sejam de fácil interpretação. Sabe-se que a omissão de coeficientes significativos pode levar a estimativas tendenciosas para os demais coeficientes da regressão. Assim, a seleção de modelos pode ser considerado um procedimentos que envolve técnica e bom senso. Nesta seção serão apresentados alguns procedimentos tradicionais de seleção de modelos em regressão linear múltipla.

1.13.1 Todas Regressões Possíveis

Supor um total de $(p-1)$ variáveis explicativas a serem selecionadas num modelo de regressão e seja T o total de regressões possíveis. Tem-se que

$$T = 1 + \binom{p-1}{1} + \binom{p-1}{2} + \cdots + \binom{p-1}{p-1} = 2^{(p-1)}.$$

Por exemplo, se $p = 4$ (3 variáveis explicativas), haverá um total de $T = 1 + 3 + 3 + 1 = 8$ regressões possíveis.

Maior $\mathbf{R}_k^2$

Seja R_k^2 o coeficiente de determinação de um submodelo com k coeficientes ($(k-1)$ variáveis explicativas + intercepto), definido por

$$\begin{aligned} \mathrm{R}_k^2 &= \frac{\mathrm{SQReg}(k)}{\mathrm{SQT}} \\ &= 1 - \frac{\mathrm{SQRes}(k)}{\mathrm{SQT}}. \end{aligned}$$

Esse critério procura um submodelo com R_k^2 alto e k pequeno (vide Figura 1.25). Alternativamente, denote por $\overline{\mathrm{R}}_k^2$ o coeficiente de determinação ajustado do submodelo com k coeficientes. Tem-se que

$$\overline{\mathrm{R}}_k^2 = 1 - (1 - \mathrm{R}_k^2)\frac{(n-1)}{(n-k)}.$$

Pode-se adotar como critério a escolha de um submodelo com $\overline{\mathrm{R}}_k^2$ alto e k pequeno. Contudo, $\overline{\mathrm{R}}_k^2$ não necessariamente cresce com k.

Menor s_k^2

Seja s_k^2 o erro quadrático médio de um submodelo com k, sendo denotado por

$$s_k^2 = \frac{\mathrm{SQRes}(k)}{n-k}.$$

Esse critério procura um submodelo com s_k^2 pequeno e k pequeno. Conforme descrito pela Figura 1.26 nem sempre o erro quadrático médio decresce com o aumento do número de coeficientes.

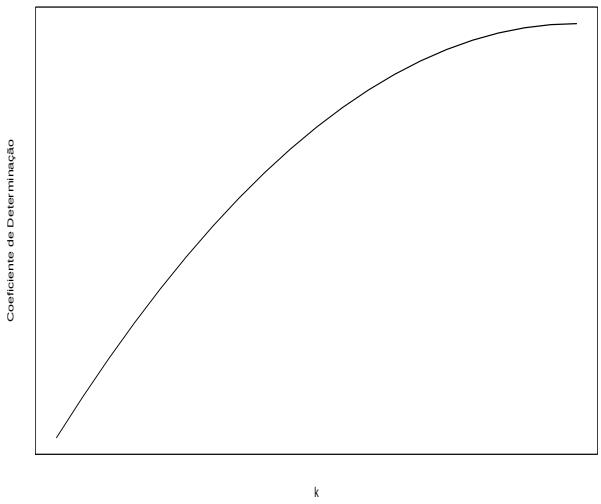

Figura 1.25: Comportamento do coeficiente de determinação R_k^2 com o número k de coeficientes.

Mostra-se que

$$\begin{aligned}
\overline{\mathrm{R}}_k^2 &= 1 - \frac{(n-1)}{(n-k)}(1-\mathrm{R}_k^2) \\
&= 1 - \frac{(n-1)}{(n-k)}\left\{1 - \frac{\mathrm{SQReg}(k)}{\mathrm{SQT}}\right\} \\
&= 1 - \frac{(n-1)}{(n-k)}\frac{\mathrm{SQRes}(k)}{\mathrm{SQT}} \\
&= 1 - \frac{(n-1)}{\mathrm{SQT}}s_k^2.
\end{aligned}$$

Assim, minimizar s_k^2 é equivalente a maximizar $\overline{\mathrm{R}}_k^2$.

Critério de Mallows

Um outro método, conhecido como critério de Mallows, está relacionado com o erro quadrático médio do i-ésimo valor ajustado $\widehat{Y}_i$ do submodelo com k

coeficientes

$$\mathrm{E}\{\widehat{Y}_i - \mathrm{E}(Y_i)\}^2 = \mathrm{Var}(\widehat{Y}_i) + \{\mathrm{E}(\widehat{Y}_i) - \mathrm{E}(Y_i)\}^2.$$

A soma dos vieses ao quadrado do submodelo com k coeficientes fica dada por

$$\{\mathrm{Viés}(k)\}^2 = \sum_{i=1}^{n} \{\mathrm{E}(\widehat{Y}_i) - \mathrm{E}(Y_i)\}^2,$$

em que $\mathrm{E}(Y_i)$ denota o valor esperado do modelo correto. Uma forma padronizada para o erro quadrático médio do submodelo com k coeficientes é expressa na forma

$$\mathrm{EQM}(k) = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n} \{\mathrm{E}(\widehat{Y}_i) - \mathrm{E}(Y_i)\}^2 + \sum_{i=1}^{n} \mathrm{Var}(\widehat{Y}_i)\right].$$

Usando o resultado $\sum_{i=1}^{n} \mathrm{Var}(\widehat{Y}_i) = k\sigma^2$ obtém-se

$$\mathrm{EQM}(k) = \frac{\{\mathrm{Viés}(k)\}^2}{\sigma^2} + k.$$

Por outro lado

$$\mathrm{E}\{\mathrm{SQRes}(k)\} = \{\mathrm{Viés}(k)\}^2 + (n-k)\sigma^2.$$

Portanto, o erro quadrático médio padronizado assume a forma

$$\mathrm{EQM}(k) = \frac{\mathrm{E}\{\mathrm{SQRes}(k)\}}{\sigma^2} - n + 2k.$$

Deve-se escolher submodelos com $\mathrm{EQM}(k)$ pequeno.

A estatística C_k de Mallows é definida por

$$\mathrm{C}_k = \frac{\mathrm{SQRes}(k)}{\widehat{\sigma}^2} - n + 2k,$$

em que $\widehat{\sigma}^2$ deve ser obtido de um modelo bem ajustado. Sob viés zero tem-se que

$$\mathrm{E}(\mathrm{C}_k|\mathrm{Viés} = 0) = \frac{(n-k)\sigma^2}{\sigma^2} - n + 2k = k.$$

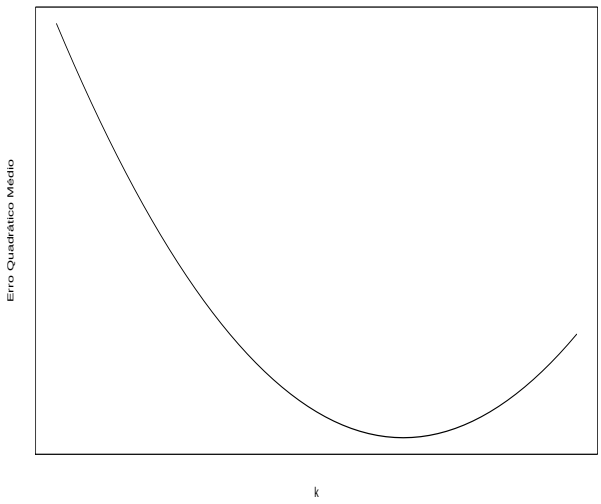

Figura 1.26: Comportamento do erro quadrático médio s_k^2 com o número k de coeficientes.

Portanto, deve-se escolher submodelos com C_k pequenos tais que $C_k \cong k$. Para um mesmo k, submodelos com $C_k < k$ têm uma SQRes menor, enquanto submodelos com $C_k > k$ têm uma SQRes maior.

Na Figura 1.27 são ilustrados 3 submodelos hipóteticos, A, B e C. O submodelo A é o pior submodelo, tem C_k alto e viés alto. O submodelo B tem um C_k menor e viés pequeno. Já o submodelo C tem um viés um pouco maior do que o submodelo B, porém um C_k bem menor, assim poderia ser o submodelo escolhido.

Critério Press

Finalmente, tem-se o critério Press que consiste em escolher o submodelo com o menor valor para a estatística

$$\text{Press}_k = \sum_{i=1}^{n} \{y_i - \widehat{y}_{(i)}\}^2,$$

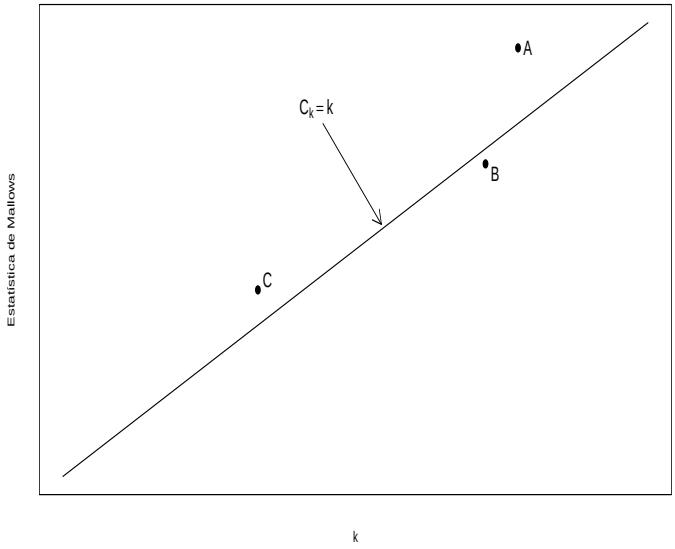

Figura 1.27: Descrição da reta $C_k = k$ e da estatística de Mallows para três submodelos hipotéticos A, B e C.

em que $\widehat{y}_{(i)}$ denota o valor predito para y_i do ajuste do submodelo com k coeficientes sem a i-ésima observação. Desde que $\widehat{y}_{(i)} = \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_{(i)}$, usando a expressão para $\widehat{\boldsymbol{\beta}}_{(i)}$ descrita na Seção 1.7.5 obtém-se

$$\begin{aligned} y_i - \widehat{y}_{(i)} &= y_i - \mathbf{x}_i^\top \left\{ \widehat{\boldsymbol{\beta}} - \frac{r_i}{(1-h_{ii})} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i \right\} \\ &= (y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}) + \frac{r_i}{(1-h_{ii})} \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_i \\ &= r_i + \frac{r_i h_{ii}}{(1-h_{ii})} \\ &= \frac{r_i}{(1-h_{ii})}. \end{aligned}$$

Logo, segue que

$$\text{Press}_k = \sum_{i=1}^{n} \left(\frac{r_i}{1-h_{ii}} \right)^2,$$

em que r_i e h_{ii} denotam, respectivamente, o i-ésimo resíduo ordinário e i-ésima medida de alavanca do submodelo com k coeficientes. Como a estatística Press_k cresce com o tamanho amostral n, uma proposta alternativa é considerar a estatística $\overline{\text{Press}}_k = \text{Press}_k/n$.

Assim, a fim de selecionar um submodelo usando os critérios: R_k^2 maior, s_K^2 menor, $\text{C}_k \cong k$ e pequeno e menor Press_k, deve-se ajustar todas as $T = 2^{(p-1)}$ regressões possíveis e selecionar um submodelo seguindo os 4 critérios descritos.

1.13.2 Métodos Sequenciais

Critérios de Akaike e de Schwartz

Seja $\text{L}(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança de um modelo de regressão com p coeficientes a serem estimados. O método de Akaike (1974) consiste em escolher um submodelo que maximize $\text{L}(\boldsymbol{\theta})$ minimizando o número de coeficientes. Isso é equivalente a minimizar a função penalizada abaixo

$$\text{AIC}_k = -2\text{L}(\widehat{\boldsymbol{\theta}}) + 2k,$$

em que $1 \leq k \leq p$ denota o número de coeficientes do submodelo. No caso de regressão linear múltipla mostra-se que

$$\text{AIC}_k = n \log\left(\frac{\text{SQRes}}{n}\right) + 2k$$

(vide Exercício 1.15). Similarmente ao método de Akaike o método de Schwartz (1978) consiste em maximizar $\text{L}(\boldsymbol{\theta})$ também minimizando o número de coeficientes da regressão, porém com uma penalização diferente. O método é equivalente a minimizar a função abaixo

$$\text{BIC}_k = -2\text{L}(\widehat{\boldsymbol{\theta}}) + k \log(n).$$

Para a regressão linear múltipla tem-se que $\text{BIC}_k = n \log\left(\frac{\text{SQRes}}{n}\right) + k \log(n)$.

Método LASSO

O método LASSO é utilizado para a seleção de variáveis explicativas (na forma padronizada) eliminando coeficientes da regressão cujas estimativas estejam próximas de zero. No contexto de mínimos quadrados o método é equivalente a minimizar a função abaixo

$$\mathrm{S}(\boldsymbol{\beta}, \lambda) = \sum_{i=1}^{n} (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 + \lambda \sum_{j=2}^{p} |\beta_j|,$$

em que $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ e $\lambda \geq 0$ é o parâmetro de penalização. Quando $\lambda = 0$ tem-se o método de mínimos quadrados e quando $\lambda \to \infty$ todos os coeficientes tendem a zero.

Critério *Forward*

Passo 1

Ajustar todas as regressões possíveis com apenas 1 variável explicativa. Isto é, ajustar as regressões

$$y_i = \beta_1 + \beta_j x_{ij} + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, para $i = 1, \ldots, n$ e $j = 2, \ldots, p$. Testar $\mathrm{H}_0 : \beta_j = 0$ contra $\mathrm{H}_1 : \beta_j \neq 0$ e obter a estatística

$$\mathrm{F}_j = \frac{\mathrm{SQReg}(x_j)}{s^2(x_j)} \overset{\mathrm{H}_0}{\sim} F_{1,(n-2)}.$$

Denote P_j o valor-P do teste. Seja $P_{\min} = \min\{P_2, \ldots, P_p\}$. Se $P_{\min} \leq P_E$ então a variável explicativa correspondente entra no modelo. Supor que X_2 entra no modelo.

Passo 2

Ajustar todas as regressões possíveis com apenas X_2 mais uma variável explicativa. Isto é, ajustar as regressões

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_j x_{ij} + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$ e $j = 3, \ldots, p$. Testar $\text{H}_0 : \beta_j = 0$ contra $\text{H}_1 : \beta_j \neq 0$ e obter a estatística

$$\text{F}_j = \frac{\text{SQReg}(x_j|x_2)}{s^2(x_2, x_j)} \overset{\text{H}_0}{\sim} F_{1,(n-3)}.$$

Denote P_j o valor-P do teste. Seja $P_{\min} = \min\{P_3, \ldots, P_p\}$. Se $P_{\min} \leq P_E$ então a variável explicativa correspondente entra no modelo. Supor que X_3 entra no modelo.

Passo 3

Ajustar todas as regressões possíveis com apenas X_2 e X_3 mais uma variável explicativa. Isto é, ajustar as regressões

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_j x_{ij} + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$ e $j = 4, \ldots, p$. Testar $\text{H}_0 : \beta_j = 0$ contra $\text{H}_1 : \beta_j \neq 0$ e obter a estatística

$$\text{F}_j = \frac{\text{SQReg}(x_j|x_2, x_3)}{s^2(x_2, x_3, x_j)} \overset{\text{H}_0}{\sim} F_{1,(n-4)}.$$

Denote P_j o valor-P do teste. Seja $P_{\min} = \min\{P_4, \ldots, P_p\}$. Se $P_{\min} \leq P_E$ então a variável explicativa correspondente entra no modelo. Se $P_{\min} > P_E$ parar o processo, nenhuma variável entra no modelo.

Critério *Backward*

Passo 1

Ajustar a regressão com todas as variáveis explicativas. Isto é, ajustar o seguinte modelo:

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$. Testar $\text{H}_0 : \beta_j = 0$ contra $\text{H}_1 : \beta_j \neq 0$ e obter a estatística

$$\text{F}_j = \frac{\text{SQReg}(x_j|\text{demais})}{s^2(x_2, \ldots, x_p)} \overset{\text{H}_0}{\sim} F_{1,(n-p)}.$$

Denote P_j o valor-P do teste, para $j = 2, \ldots, p$. Seja $P_{\max} = \max\{P_2, \ldots, P_p\}$. Se $P_{\max} \geq P_S$ então a variável explicativa correspondente sai do modelo. Supor que X_2 sai do modelo.

Passo 2

Ajustar a regressão sem a variável explicativa X_2. Isto é, ajustar o seguinte modelo:

$$y_i = \beta_1 + \beta_3 x_{i3} + \cdots + \beta_p x_{ip} + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$. Testar $\text{H}_0 : \beta_j = 0$ contra $\text{H}_1 : \beta_j \neq 0$ e obter a estatística

$$\text{F}_j = \frac{\text{SQReg}(x_j|\text{demais})}{s^2(x_3, \ldots, x_p)} \overset{\text{H}_0}{\sim} F_{1,(n-p-1)}.$$

Denote P_j o valor-P do teste, para $j = 3, \ldots, p$. Seja $P_{\max} = \max\{P_3, \ldots, P_p\}$. Se $P_{\max} < P_S$ o processo é terminado, nenhuma variável sai do modelo.

Critério *Stepwise*

O critério *stepwise* é uma combinação dos critérios *forward* e *backward*.

Passo 1

Ajustar todas as regressões com apenas uma variável explicativa, além do intercepto. Verificar se alguma variável explicativa entra no modelo. Supor que X_2 entrou no modelo.

Passo 2

Ajustar todas as regressões com X_2 mais uma variável explicativa, além do intercepto. Verificar se alguma variável explicativa entra no modelo. Supor que X_3 entrou no modelo. Verificar se X_2 sai do modelo dado que X_3 está no modelo.

Passo 3

O processo *stepwise* deve continuar até que não seja possível incluir nenhuma variável no modelo, nem retirar nenhuma variável do modelo.

Critérios de Parada

Não há um consenso na área de regressão a respeito de critérios de parada para os processos sequenciais. Alguns critérios mais utilizados:

(i) usar $F_E = F_S = 4$ que equivale aproximadamente a usar $P_E = P_S = 0,05$;

(ii) ser mais flexível na entrada do que na saída $P_E = 0,25$ e $P_S = 0,10$, ou com os mesmos critérios na entrada e na saída $P_E = P_S = 0,15$.

1.13.3 Estratégias para a Seleção de Modelos

Portanto, não há uma receita pronta para a seleção de modelos a partir de um conjunto de variáveis explicativas. Em Montgomery et al. (2021, Seção

10.3) há uma longa discussão a respeito de possíveis estratégias para seleção de modelos através dos critérios propostos nesta seção.

Segundo os autores, quando o número de variáveis explicativas é relativamente pequeno pode ser factível ajustar todas as regressões possíves e selecionar algumas candidatas segundo os critérios R_k^2 maior, s_K^2 menor, $C_k \cong k$ e pequeno e menor $\overline{\text{Press}}_k$. Para as regressões selecionadas sugere-se fazer uma análise de diagnóstico e levar em conta aspectos como a importância, custo e facilidade de interpretação das variáveis explicativas, bem como da capacidade de predição do modelo.

Os métodos sequenciais *forward*, *backward* e *stepwise* são recomendados quando há um número médio ou alto de variáveis explicativas, contudo exigem os níveis de significância de entrada e saída das variáveis explicativas. Já os métodos de Akaike e de Schwartz são mais recomendados quando há um grande número de variáveis explicativas no sentido de se fazer uma pré-seleção de variáveis sem a necessidade de estabelecer níveis de significância. Todos os métodos sequencias podem ser combinados com o ajuste de todas as regressões possíveis.

A seleção de modelos pode ficar mais complexa quando há interesse em selecionar variáveis explicativas que estejam relacionadas no sentido causa-efeito com a resposta, como ocorre por exemplo na área médica. Nesses casos, os algoritmos em geral são combinações de procedimentos sequenciais com procedimentos que procuram evitar a eliminação precoce de variáveis explicativas potenciais no sentido causa-efeito. Em Dunkler et al. (2014) há uma proposta de algoritmo híbrido que combina o procedimento de eliminação *backward* com procedimentos que levam em conta o efeito da eliminação de variáveis explicativas nos coeficientes das variáveuis mantidas no modelo.

1.14 Aplicações

1.14.1 Venda de Telhados

Considere novamente os dados descritos em Neter et al. (1996, p.449) referentes à venda no ano anterior de um tipo de telhado de madeira em $n = 26$ filiais de uma rede de lojas de construção civil, agora com as seguintes variáveis:

(i) Telhados: total de telhados vendidos (em mil metros quadrados),

(ii) Nclientes: número de clientes cadastrados na loja (em milhares),

(iii) Gastos: gastos pela loja com promoções do produto (em mil USD),

(iv) Marcas: número de marcas concorrentes do produto e

(v) Potencial: potencial da loja (quanto maior o valor maior o potencial).

O interesse é explicar o número médio de telhados vendidos dadas as demais variáveis. Na Tabela 1.12 tem-se as estimativas da correlação linear de Pearson entre as variáveis do exemplo vendas de telhados. Nota-se uma baixa correlação entre telhados e gastos, altas correlações entre telhados com número de clientes e marcas e uma correlação moderada com potencial da loja. Entre as variáveis explicativas nota-se correlações baixas, exceto uma correlação moderada entre número de clientes e potencial da loja. As correlações descritas na Tabela 1.12 estão coerentes com os diagramas de dispersão apresentados nas Figuras 1.28 e 1.29.

O primeiro critério a ser aplicado para selecionar um submodelo linear normal é com todas as regressões possíveis, cujos resultados das medidas resumo são apresentados na Tabela 1.13. Dois submodelos se destacam segundo os 4 critérios utilizados: 1 + Nclientes + Marcas e 1 + Gastos + Ncientes + Marcas. Levando-se em conta o número de variáveis explicativas

Tabela 1.12: Matriz de correlações lineares amostrais de Pearson entre as variáveis do exemplo vendas de telhados.

	Telhados	Gastos	Nclientes	Marcas	Potencial
Telhados	1,0	0,159	0,783	-0,833	0.407
Gastos		1,0	0,173	-0,038	-0,070
Nclientes			1,0	-0,324	0,468
Marcas				1,0	-0,202
Potencial					1,0

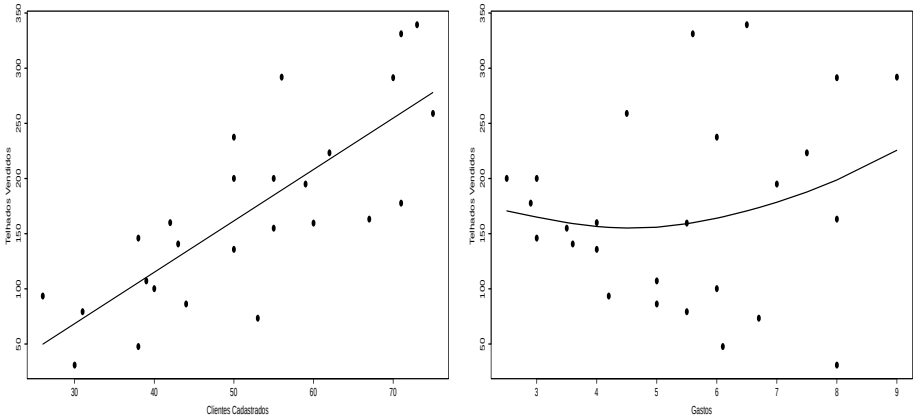

Figura 1.28: Diagramas de dispersão (com tendência) entre o total de telhados vendidos e o número de clientes cadastrados (esquerda) e gastos pela loja com promoções (direita).

o submodelo 1 + Nclientes + Marcas poderia ser escolhido, contudo deve-se fazer antes uma análise de diagnóstico com cada submodelo.

Os dois submodelos selecionados 1 + Nclientes + Marcas e 1 + Gastos + Ncientes + Marcas apresentaram excelentes ajustes, conforme pode ser observado pelas Tabelas 1.14 e 1.15 e pelos gráficos de resíduos descritos nas Figuras 1.30 e 1.31. Porém, a variável explicativa gastos aparece marginal-

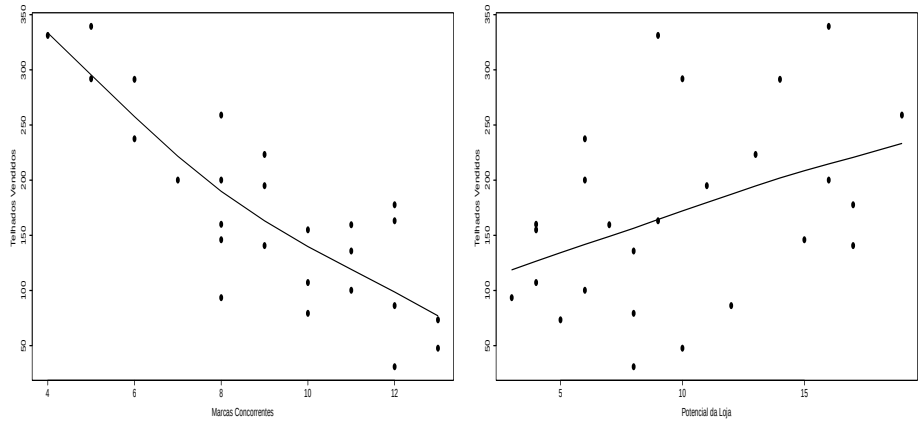

Figura 1.29: Diagramas de dispersão (com tendência) entre o total de telhados vendidos e o número de marcas concorrentes do produto (esquerda) e o potencial da loja (direita).

mente não significativa no 2º submodelo. Ambos os submodelos destacam os mesmos pontos potencialmente influentes pela distância de Cook com $k = 2$ (Figura 1.32). A eliminação da observação #21 deixa a variável explicativa gastos significativa ao nível de 5% no 2º submodelo. Portanto, essa observação está mascarando o efeito de gastos. Assim, deve-se escolher o submodelo 1 + Gastos + Nclientes + Marcas.

O segundo critério a ser aplicado é o método sequencial *stepwise* com $P_E = P_S = 0,15$. Na Tabela 1.16 tem-se um resumo dos 6 passos necessários para selecionar um submodelo. No 1º passo entra a variável marcas e no 2º passo entra a variável número de clientes. No 3º passo a variável marcas não sai do modelo. Já no 4º passo entra no modelo a variável gastos e no 5º passo nenhuma variável sai do modelo e finalmente no 6º passo a última variável potencial não entra no modelo. Assim, o submodelo selecionado pelo procedimento *stepwise* coincide com o submodelo selecionado pelo critério

Tabela 1.13: Medidas resumo dos 16 submodelos para explicar o número médio de telhados vendidos, em que T:Telhados, G:Gastos, N:Nclientes, M:Marcas, P:Potencial e k denota o número de parâmetros.

Submodelo[1]	$k-1$	k	R_k^2	s_k	C_k	$\overline{Press}_k$
1	0	1	0,00	84,6	1960,2	7434,5
1 + G	1	2	0,025	85,2	1912,1	7829,8
1 + N	1	2	0,613	53,7	746,2	3115,0
1 + M	1	2	0,694	47,8	585,4	2428,8
1 + P	1	2	0,166	78.8	1633,1	6522,2
1 + G + N	2	3	0,613	54,8	747,0	3508,8
1 + G + M	2	3	0,710	47,5	555,4	2543,8
1 + G + P	2	3	0,201	78,8	1564,9	6770,1
1 + N + M	2	3	0,988	9,8	4,5	113,6
1 + N + P	2	3	0,615	54,7	744,0	3330,4
1 + M + P	2	3	0,753	43,8	469,3	2166,2
1 + G + N + M	3	4	0,989	9,5	4,0	115,4
1 + G + N + P	3	4	0,616	55,9	743,9	3726,5
1 + G + P + M	3	4	0,775	42,6	428,4	2222,4
1 + N + P + M	3	4	0,988	10,0	6,4	120,8
1 + G + N + P + M	4	5	0,989	9,6	5,5	119,5

Tabela 1.14: Estimativas referentes ao submodelo 1 + N + M.

Efeito	Estimativa	E.Padrão	valor-t	valor-P
Constante	186,694	12,259	15,23	0,00
Nclientes	3,408	0,146	23,37	0,00
Marcas	-21,193	0,803	-26,40	0,00
s	9,803			
R^2	0,988			
$\overline{R}^2$	0,987			

Tabela 1.15: Estimativas referentes ao submodelo 1 + G + N + M.

Efeito	Estimativa	E.Padrão	valor-t	valor-P
Constante	179,844	12,621	14,25	0,00
Gastos	1,677	1,052	1,59	0,12
Nclientes	3,369	0,143	23,52	0,00
Marcas	-21,217	0,773	-27,30	0,00
s	9,491			
R^2	0,989			
$\overline{R}^2$	0,987			

Tabela 1.16: Resumo dos passos do procedimento *stepwise* com $P_E = P_S = 0,15$ e valores-P em cada passo para selecionar as variáveis explicativas do exemplo venda de telhados.

Passo	Gastos	Nclientes	Marcas	Potencial
Passo 1	0,4382	0,0000	0,0000	0,0389
Passo 2	0,2693	0,0000	-	0,0274
Passo 3	-	-	0,0000	-
Passo 4	0,1252	-	-	0,6968
Passo 5	-	0,0000	0,0000	-
Passo 6	-	-	-	0,4854

com todas as regrossões possíveis.

Finalmente, aplicando o critério de Akaike obtém-se como menor valor `AIC` $= 120,67$, que corresponde ao mesmo submodelo obtido com os dois procedimentos anteriores. Portanto, o submodelo selecionado contém as variáveis explicativas gastos, número de clientes e marcas, além da constante, cujas estimativas são apresentadas na Tabela 1.15. Interpretando as estimativas tem-se que a cada aumento de USD 1000 nos gastos da loja com promoções e de 100 clientes cadastrados, espera-se aumento de 1677 mil m^2

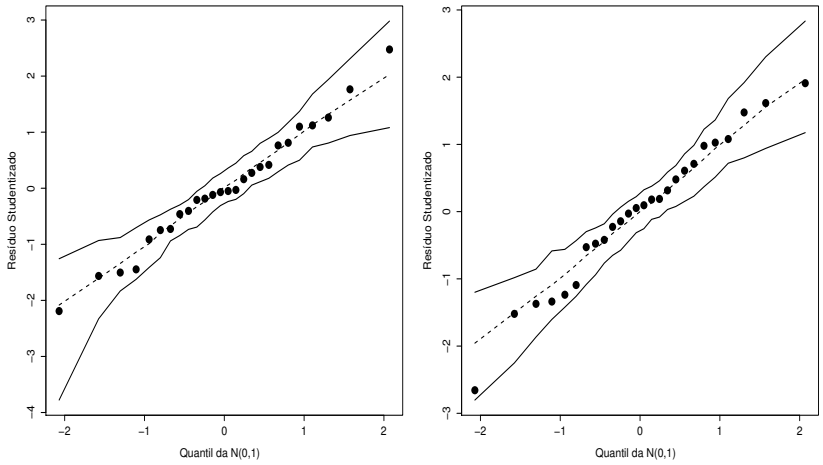

Figura 1.30: Gráficos normais de probabilidades referentes aos submodelos 1 + N + M (esquerda) e 1 + G + N + M (direita).

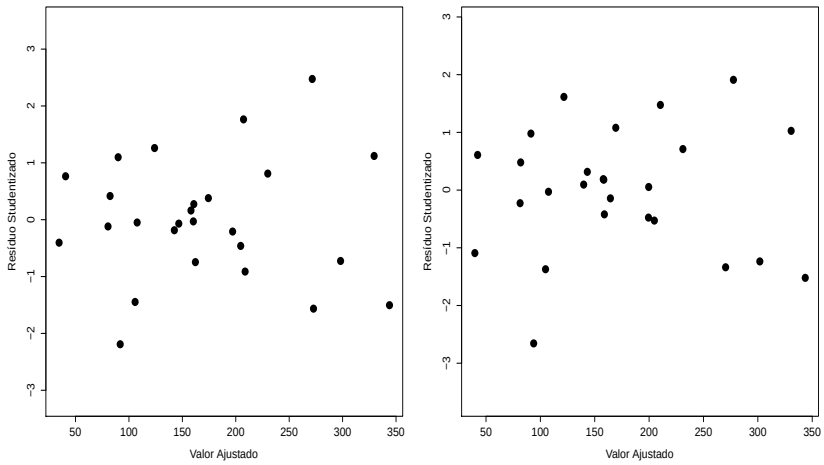

Figura 1.31: Gráficos do resíduo Studentizado contra o valor ajustado referentes aos submodelo 1 + N + M (esquerda) e 1 + G + N + M (direita).

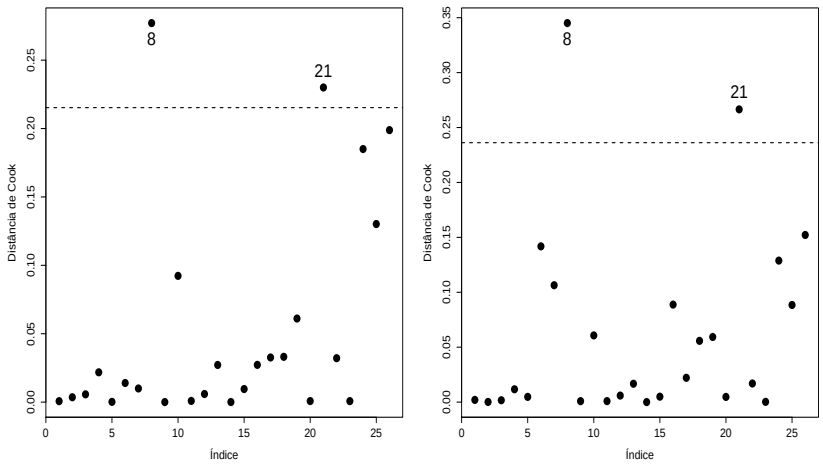

Figura 1.32: Gráficos da distância de Cook referentes aos submodelos 1 + N + M (esquerda) e 1 + G + N + M (direita).

e 337 mil m^2 de telhados vendidos, respectivamente. Por outro lado, um aumento de 10 marcas concorrentes leva a uma redução média de 212 mil m^2 de telhados vendidos.

1.14.2 Salário de Executivos

Considere os dados de uma pesquisa realizada por uma revista de negócios sobre o salário anual de executivos (em mil USD) descrita em Foster et al. (1998, pp. 180-188), em que uma amostra aleatória de 220 executivos (145 homens e 75 mulheres) foi coletada. Além do salário anual foram consideradas as seguintes variáveis explicativas:

(i) Gênero (1: masculino; 0: feminino),

(ii) Posição: posição na empresa (varia de 1 a 9), quanto maior o valor mais alta a posição e

(iii) Experiência: anos de experiência no cargo ou tempo no cargo.

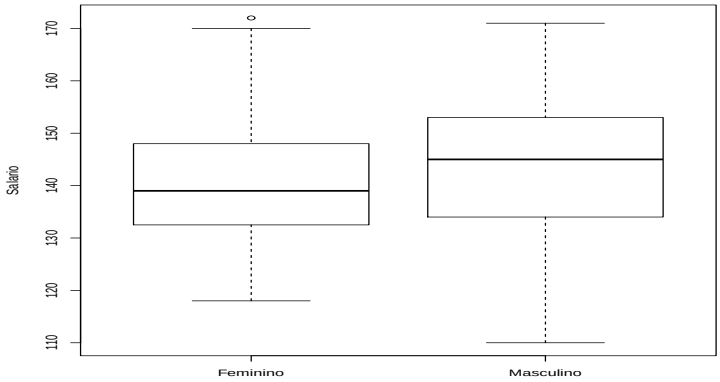

Figura 1.33: Boxplot robusto do salário anual segundo o gênero.

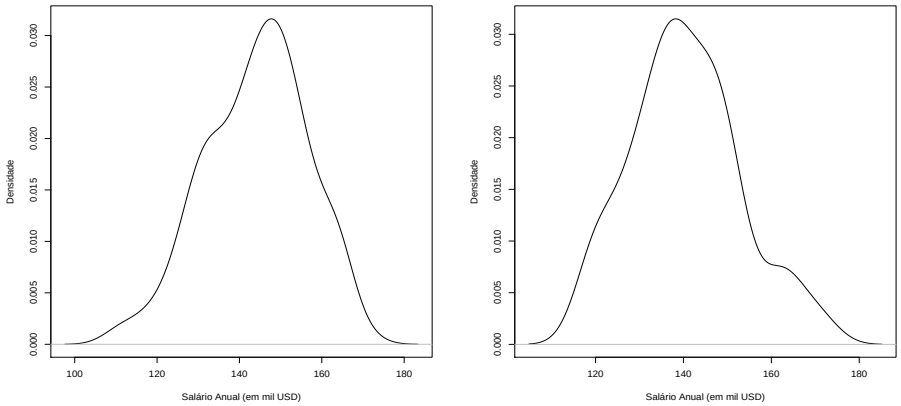

Figura 1.34: Densidade do salário anual dos executivos (esquerda) e das executivas (direita).

Os dados estão descritos no arquivo **salarios.txt**. O objetivo principal do estudo é explicar o salário médio anual segundo as três variáveis explicativas.

Tabela 1.17: Descrição dos salários médios anuais com os respectivos erros padrão e do teste-t de igualdede de médias.

Gênero	Amostra	Média	E.Padrão
Masculino	145	144,11	1,03
Feminino	75	140,47	1,43
	Diferença	Teste-t	valor-P
Estimativa	3,64	2,06	0,04
E.Padrão	1,77		

As Figuras 1.33 e 1.34 descrevem, respectivamente, os bloxplots robustos do salário anual segundo o gênero e as respectivas densidades empíricas. Nota-se uma ligeira superioridade dos salários anuais dos executivos. Isso é confirmado pela Tabela 1.17 onde são descritas as médias salariais com os respectivos erros padrão e o test-t para comparação de médias. A hipótese de igualdade de médias entre os dois grupos é rejeitada ao nível de significância de 5%. Há, portanto, indícios que os executivos em média ganham mais do que as executivas.

Com relação à posição na empresa e experiência no cargo, nota-se pela Figura 1.35 que os executivos ecupam em geral posições mais altas e têm mais experiência do que as executivas. Os diagramas de dispersão entre o salário anual e a posição para ambos os gêneros (Figura 1.36) descrevem tendências crescentes, enquanto os diagramas de dispersão entre salário e experiência indicam também tendências crescentes (Figura 1.37), porém com menor intensidade.

Essas análises descritivas sugerem, em princípio, o seguinte modelo linear:

$$y_i = \beta_1 + \beta_2 \text{gênero}_i + \beta_3 \text{experiência}_i + \beta_4 \text{posição}_i + \epsilon_i, \qquad (1.6)$$

em que y_i denota o salário do i-ésimo executivo da amostra com $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$,

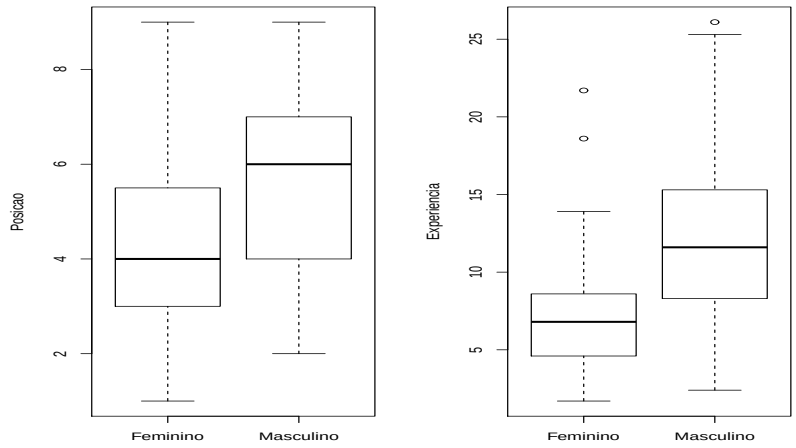

Figura 1.35: Boxplots robustos da posição e da experiência segundo o gênero.

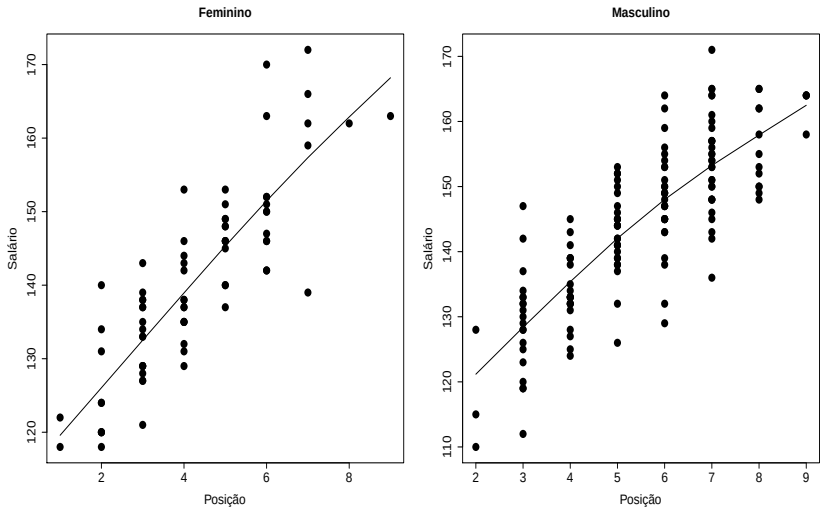

Figura 1.36: Diagrama de dispersão (com tendência) entre salário e posição segundo o gênero.

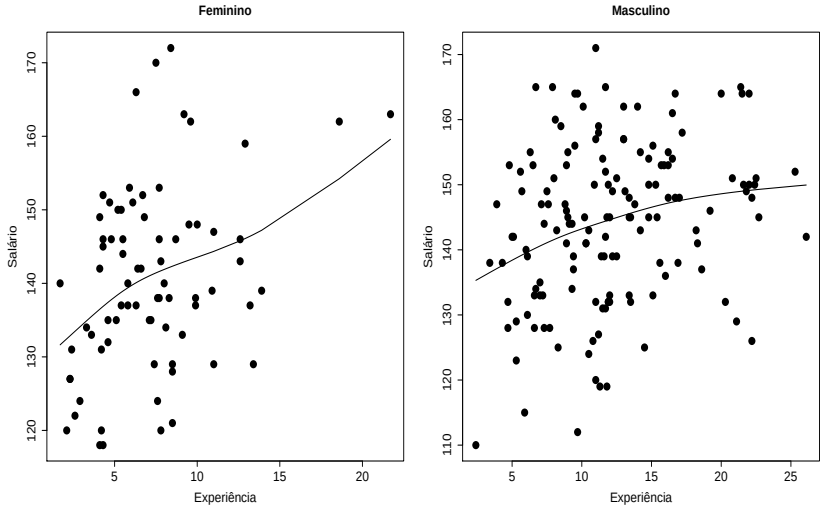

Figura 1.37: Diagrama de dispersão (com tendência) entre salário e experiência segundo o gênero.

para $i = 1, \ldots, 220$.

As estimativas referentes ao modelo (1.6) estão descritas na Tabela 1.18 e pode-se notar que todos os efeitos são marginalmente significativos. Em particular, nota-se que à medida que aumenta a posição na empresa espera-se maior salário, fixados os demais efeitos. A experiência, segundo o modelo ajustado, à medida que aumenta tende a reduzir o salário médio e as executivas, quando comparadas com os executivos nos mesmos níveis de posição e experiência, têm um salário esperado maior. Esses resultados parecem contradizer parte da análise descritiva, contudo são interpretações diferentes. A análise descritiva faz comparações marginais, enquanto a análise de regressão leva em conta todas as variáveis conjuntamente. Segundo as análises de resíduos (omitidas aqui) o modelo está bem ajustado, porém Foster et al.(1998) sugerem a inclusão de interações para agregar mais interpretações.

A Tabela 1.19 apresenta os valores da estatística F com os respectivos

Tabela 1.18: Estimativas dos parâmetros referentes ao modelo de regressão linear múltipla (1.6) ajustado aos dados sobre salário de executivos.

Efeito	Estimativa	E.Padrão	valor-t	valor-P
Constante	115,262	1,491	82,25	0,00
Experiência	-0,472	0,113	-4,17	0,00
GêneroM	-2,201	1,080	-2,04	0,04
Posição	6,710	0,313	21,46	0,00
s	6,77			
R^2	0,71			
$\overline{R}^2$	0,71			

Tabela 1.19: Teste F para a inclusão de interação no modelo (1.6).

Interação	valor-F	valor-P
gênero*experiência	1,615	0,20
gênero*posicão	0,001	0,97
experiência*posição	7,594	0,00

valores-P para a inclusão de cada interação no modelo (1.6). Nota-se que apenas a interação entre experiência e posição será incluída no modelo. Assim, o seguinte modelo será considerado:

$$\begin{aligned} y_i &= \beta_1 + \beta_2 \text{gênero}_i + \beta_3 \text{experiência}_i + \beta_4 \text{posição}_i + \\ &\quad + \gamma \text{experiência}_i * \text{posição}_i + \epsilon_i, \end{aligned} \tag{1.7}$$

em que y_i denota o salário do i-ésimo executivo da amostra com $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, 220$. Na Tabela 1.20 são apresentadas as estimativas do ajuste do modelo (1.7) aos dados sobre salário de executivos. Nota-se confirmação da inclusão da interação entre experiência e posição, contudo o efeito princi-

pal de experiência ficou não significativo. Não houve variações importantes nos coeficientes de determinação, indicando que a qualidade do ajuste permanece a mesma. Confirma-se pela estimativa do coeficiente de gênero que as executivas ganham em média mais do que os executivos, fixando-se os níveis de posição e experiência.

Tabela 1.20: Estimativas dos parâmetros referentes ao modelo de regressão linear múltipla (1.7) ajustado aos dados sobre salário de executivos.

Efeito	Estimativa	E.Padrão	valor-t	valor-P
Constante	108,042	2,961	36,48	0,00
Experiência	0,336	0,314	1,07	0,28
GêneroM	-2,811	1,087	-2,59	0,01
Posição	8,096	0,590	13,73	0,00
Exper*Posição	-0,135	0,049	-2,76	0,00
s	6,67			
R^2	0,72			
$\overline{R}^2$	0,72			

Pela Figura 1.38 não há indícios de afastamentos da normalidade e da constância de variância dos erros, bem como ausência de observações aberrantes. Contudo, pelo gráfico da distância de Cook com $k = 4$ (Figura 1.39) três observações são destacadas como possivelmente influentes. Apenas as observações #4 e #30 causam variações desproporcionais, respectivamente, de -14% e 11% na estimativa do coeficiente de gênero, embora não ocorram mudanças inferencias. A observação #4 é de uma executiva com salário anual de USD 139 mil (média USD 140,5 mil), posição 7 (média 4,3) e 13,9 anos de experiência (média 7,3 anos), enquanto a observação #30 é de um executivo com salário anual de USD 110 mil (média USD 144,1 mil), posição 2 (média 5,3) e 2,4 anos de experiência (média 12,2 anos).

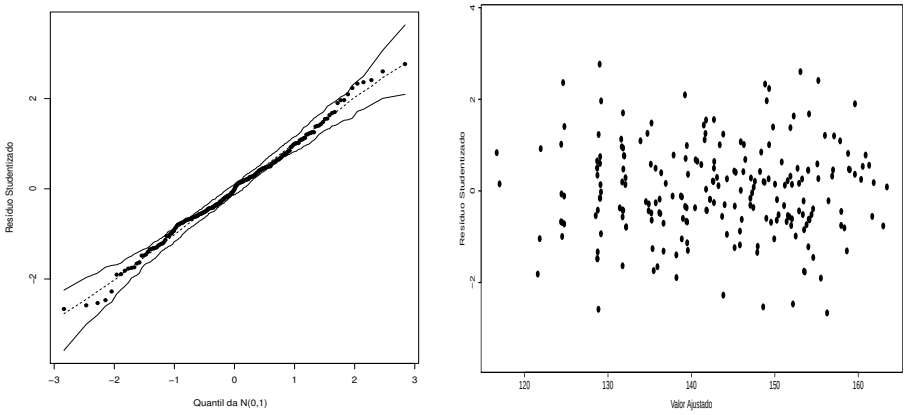

Figura 1.38: Análise de resíduos referente ao modelo (1.7) ajustado aos dados sobre salário de executivos.

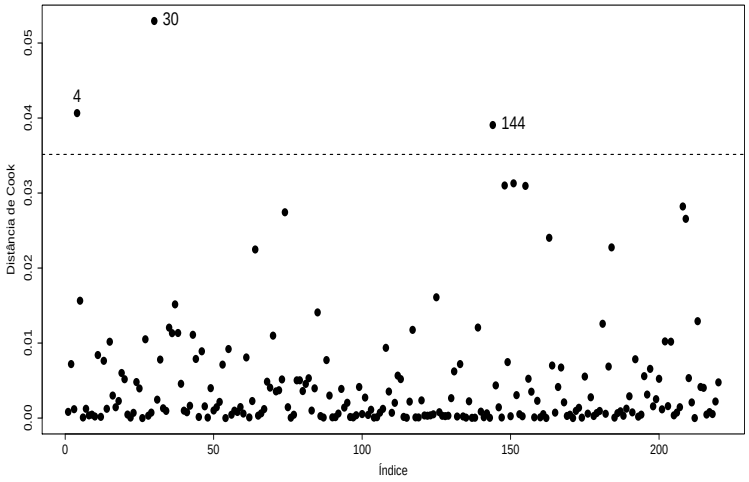

Figura 1.39: Distância de Cook contra a ordem das observações referente ao modelo (1.6) ajustado aos dados sobre salário de executivos.

O modelo ajustado fica então dado por

$$\begin{aligned}\widehat{y}(\mathbf{x}) &= 108,042 + 0,336\text{experiência} - 2,811\text{gênero} + \\ &\quad +8,096\text{posição} - 0,135\text{posição} * \text{experiência},\end{aligned}$$

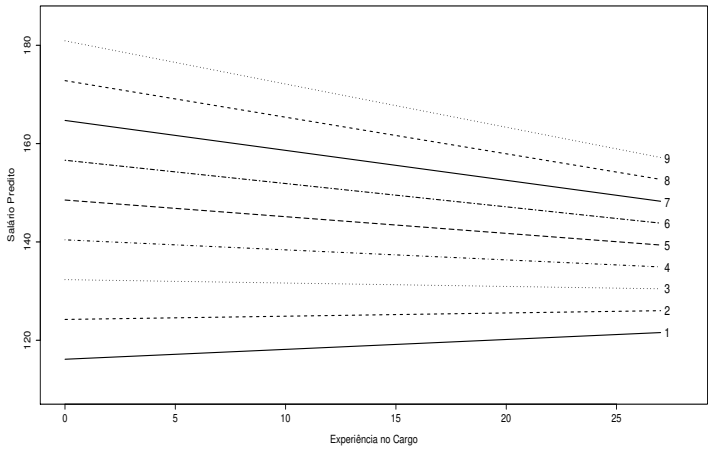

Figura 1.40: Salário médio estimado das executivas segundo a experiência e a posição.

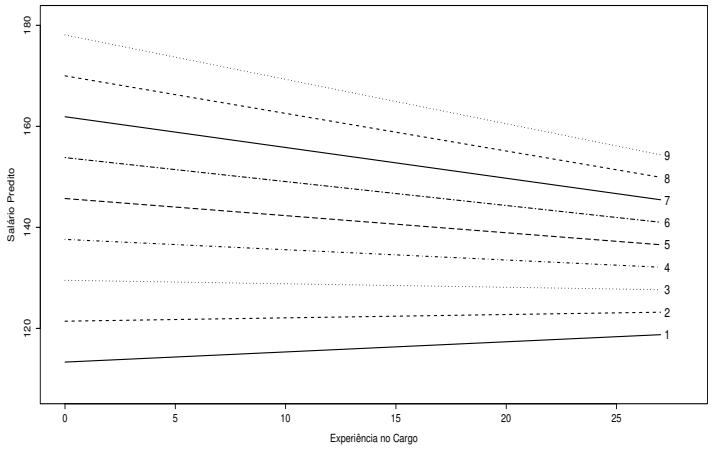

Figura 1.41: Salário médio estimado dos executivos segundo a experiência e a posição.

em que $\mathbf{x} = (1, \text{experiência}, \text{gênero}, \text{posição})^\top$.

Finalmente, nas Figuras 1.40 e 1.41 tem-se os salários preditos para exe-

cutivas e executivos, conforme variam a experiência e a posição. Nota-se que o salário predito para as executivas é sempre maior do que o salário predito para os executivos, fixados os níveis de experiência e posição. Para ambos os grupos o salário tende a crescer com o aumento do tempo no cargo nas posições iniciais 1 e 2. Contudo, nas demais posições o salário tende a decrescer com o aumento do tempo no cargo. Fixando-se a experiência o salário aumenta à medida que aumenta a posição. Todavia, a diferença salarial entre duas posições quaisquer tende a diminuir à medida que aumenta a experiência. Portanto, uma conclusão que pode-se extrair da interação entre posição e experiência é que não vale a pena do ponto de vista salarial ficar muito tempo no mesmo cargo.

1.15 Regressão por Partes

Quando a relação entre a variável resposta e alguma variável explicativa contínua é não linear, pode-se pensar em ajustar um polinômio a fim de obter um ajuste adequado, ou aplicar algum tipo de transformação na variável explicativa de modo que a relação entre as duas variáveis fique aproximadamente linear. Nesse segundo caso, muda-se a escala da variável explicativa dificultando a interpretação do coeficiente correspondente da regressão, contudo implicando num modelo mais simples. No caso polinomial, à medida que o grau do polinômio aumenta tem-se um modelo mais complexo com possibilidade de multicolinearidade. Uma forma de amenizar a complexidade desses polinômios, sem comprometer a aplicação do método de mínimos quadrados, é através da regressão por partes. Nesse procedimento, o domínio da variável explicativa é dividido em partes através de nós (pontos de mudança) escolhidos pelo analista, sendo ajustada uma regressão polinomial de grau cúbico em cada uma das partes que são segmentadas formando um único ajuste.

Esse procedimento é intermediário entre a regressão tradicional paramétrica e a regressão não paramétrica ou aditiva, em que métodos mais sofisticados são utilizados.

Como motivação, considere os dados do experimento em que a queda de tensão da bateria (em voltagem) de um motor de míssil guiado é observada ao longo do tempo (em segundos), em 41 instantes (Montgomery et al., 2012, Seção 7.2.2). Esses dados são descritos no arquivo **bateria.txt** e na Figura 1.42. Nota-se um comportamente não linear, aumento da tensão da voltagem até aproximadamente 12 segundos seguido de uma queda até aproximadamente 20 segundos.

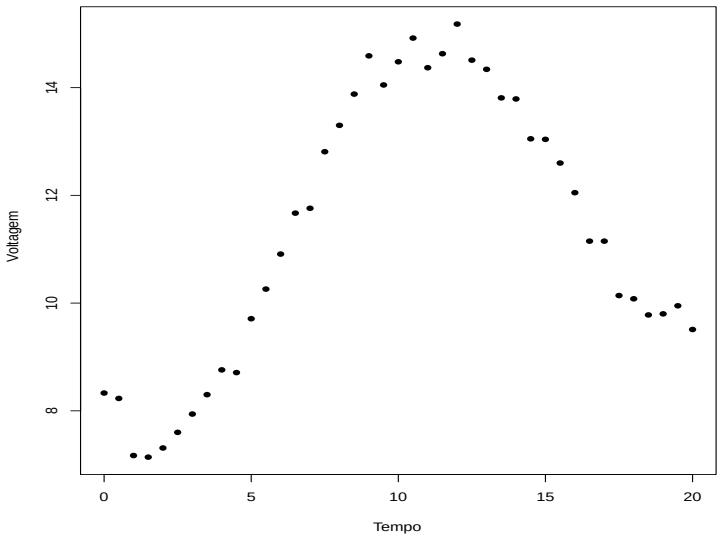

Figura 1.42: Diagrama de dispersão entre a queda da tensão da bateria (em voltagem) e o tempo (em segundos).

Pode-se propor para ajustar os dados o seguinte modelo:

$$y_i = f(x_i) + \epsilon_i,$$

em que y_i denota a tensão da voltagem no i-ésimo instante, ϵ_i's são erros aleatórios, $i = 1, \ldots, 41$, e $f(x)$ uma função suave do tempo. Como mencionado anteriormente, pode-se dividir o domínio da variável explicativa X em partes separadas por nós, sendo em cada parte ajustada uma curva de regressão. Depois junta-se as curvas.

Para um único ponto de mudança t, define-se o seguinte tipo de função:

$$(x-t)_+^r = \begin{cases} (x-t)^r & \text{se} \quad x > t \\ 0 & \text{se} \quad x \leq t, \end{cases}$$

para $r = 0, 1, 2, \ldots$. Como ilustração de um exemplo com um único ponto de mudança t, supor o ajuste de duas retas com inclinações diferentes através do modelo

$$y = f(x) + \epsilon,$$

em que $f(x) = \beta_0 + \beta_1 x + \gamma(x-t)_+$. Logo, para $x \leq t$ tem-se $f_1(x) = \beta_0 + \beta_1 x$ e para $x > t$ tem-se $f_2(x) = (\beta_0 - \gamma t) + (\beta_1 + \gamma)x$. Note que quando $x = t$ tem-se $f_1(x) = f_2(x)$, portanto há continuidade das duas retas. Assim, um modelo de regressão linear seria dado por

$$y_i = \beta_0 + \beta_1 x_i + \gamma(x_i - t)_+ + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$. Supondo $x_1 < x_2 < \cdots < x_n$ e que $x_s \leq t < x_{s+1}$, a matriz modelo fica dada por

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & x_s & 0 \\ 1 & x_{s+1} & (x_{s+1} - t) \\ \vdots & \vdots & \vdots \\ 1 & x_n & (x_n - t) \end{bmatrix}.$$

Uma proposta mais flexível, para um único ponto de mudança t, é consi-

derar a seguinte função cúbica:

$$\begin{aligned} f(x) &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \gamma_0 (x-t)_+^0 + \gamma_1 (x-t)_+^1 + \\ &\quad + \gamma_2 (x-t)_+^2 + \gamma_3 (x-t)_+^3. \end{aligned}$$

Contudo, é necessário impor condições de continuidade para $f(x)$, $f'(x)$ e $f''(x)$ em $x = t$, que implica nas restrições $\gamma_0 = 0$, $\gamma_1 = 0$ e $\gamma_2 = 0$. Assim, tem-se uma função cúbica mais simples

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \gamma_3 (x-t)_+^3.$$

O modelo correspondente de regressão linear fica dado por

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \gamma_3 (x_i - t)_+^3 + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$. Supondo $x_1 < x_2 < \cdots < x_n$ e que $x_s \leq t < x_{s+1}$ a matriz modelo fica dada por

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_3^2 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_s & x_s^2 & x_s^3 & 0 \\ 1 & x_{s+1} & x_{s+1}^2 & x_{s+1}^3 & (x_{s+1} - t)^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & (x_n - t)^3 \end{bmatrix}.$$

Generalizando, para h pontos de mudança $t_1 < t_2 < \cdots < t_h$ a função cúbica fica dado por

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{\ell=1}^{h} \gamma_\ell (x - t_\ell)_+^3.$$

Assim, uma regressão linear parcial aditiva em que k variáveis explicativas contínuas são ajustadas através de funções por partes pode ser expressa na forma

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + f_1(u_1) + \cdots + f_k(u_k) + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$. Nesse tipo de modelo há dois componentes, o primeiro referente a variáveis explicativas discretas ou contínuas cujos coeficientes são intepretáveis e o segundo formado por um conjunto de funções aditivas cujos coeficientes não são diretamente interpretáveis, contudo procuram captar da melhor maneira os efeitos não lineares de variáveis explicativas contínuas. Em muitas situações práticas $U_1, \ldots, U_k$ são variáveis de controle, tais como tempo e temperatura, havendo interesse principal na interpretação dos coeficientes do componente linear.

Voltando ao experimento sobre a queda de tensão da bateria de um motor de míssil, considere os pontos de mudança $t_1 = 6,5$ e $t_2 = 13$ (vide Figura 1.43) propostos por Montgomery et al. (2012, Seção 7.2.2).

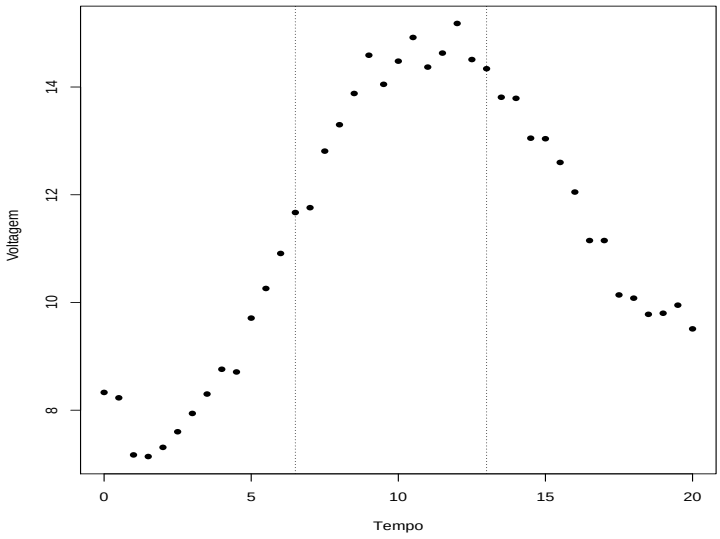

Figura 1.43: Diagrama de dispersão entre a queda da tensão da bateria (em voltagem) e o tempo (em segundos) com os pontos e mudança.

Tem-se portanto a seguinte regressão linear por partes:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \gamma_1 (x_i - 6,5)_+^3 + \gamma_2 (x_i - 13)_+^3 + \epsilon_i, \quad (1.8)$$

para $i = 1, \ldots, 41$ e cuja matriz modelo fica dada por

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_r & x_r^2 & x_r^3 & 0 & 0 \\ 1 & x_{r+1} & x_{r+1}^2 & x_{r+1}^3 & (x_{r+1} - t_1)^3 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_s & x_s^2 & x_s^3 & (x_s - t_1)^3 & 0 \\ 1 & x_{s+1} & x_{s+1}^2 & x_{s+1}^3 & (x_{s+1} - t_1)^3 & (x_{s+1} - t_2)^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & (x_n - t_1)^3 & (x_n - t_2)^3 \end{bmatrix},$$

em que $r = 14$ e $s = 27$. Supondo erros independentes e homocedásticos as estimativas de mínimos quadrados são apresentadas na Tabela 1.21. Nota-de que todos os coeficientes são altamente significativos com coeficiente de determinação bastante alto. Os gráficos de resíduos da Figuras 1.44 indicam para um ajuste adequado. Mesmo o gráfico do resíduo Studentizado contra o tempo (omitido nas análises) não indica erros correlacionados. Na Figura 1.45 tem-se a curva ajustada aos dados.

1.16 Métodos Robustos

Quando aparecem observações suspeitas de serem atípicas (alavanca, aberrante ou influente) num ajuste de regressão, deve-se através de algum procedimento de análise confirmatória verificar se de fato essas observações são mesmo atípicas. O procedimento mais utilizado é avaliar o impacto dessas observações nos coeficientes estimados da regressão através, por exemplo, de

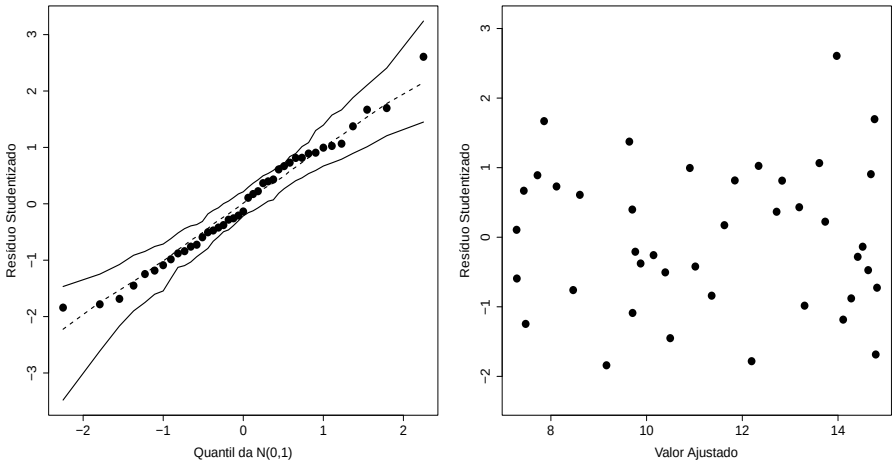

Figura 1.44: Análise de resíduos referente ao ajuste da regressão por partes (1.8) aos dados sobre a queda de tensão da bateria de um motor de míssil.

Tabela 1.21: Estimativas dos parâmetros referentes ao modelo de regressão por partes (1.8) ajustado aos dados sobre a queda de tensão da bateria de um motor de míssil.

Parâmetro	Estimativa	E.Padrão	valor-t	valor-P
β_0	8,4657	0,2005	42,22	0,00
β_1	-1,4531	0,1816	-8,00	0,00
β_2	0,4899	0,0430	11,39	0,00
β_3	-0,0294	0,0028	-10,35	0,00
γ_1	0,0247	0,0040	6,12	0,00
γ_2	0,0271	0,0036	7,58	0,00
s	0,268			
R^2	0,990			
$\overline{R}^2$	0,989			

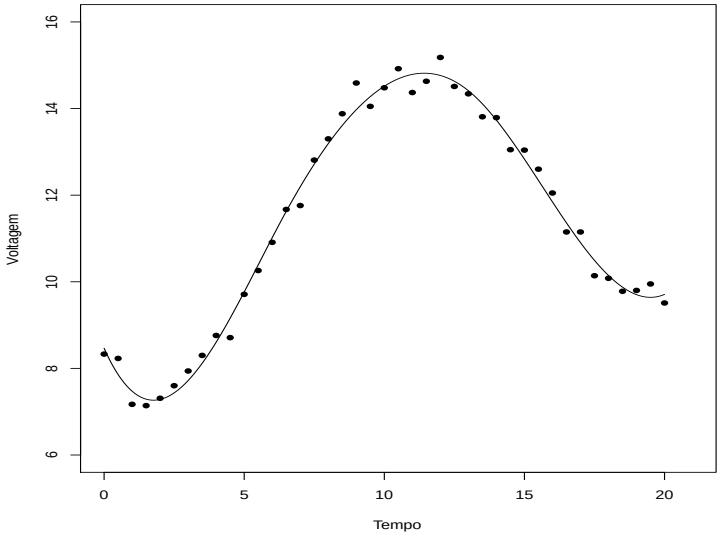

Figura 1.45: Curva ajustada pela regressão por partes (1.8) aos dados sobre a queda de tensão da bateria de um motor de míssil.

comparações com observações não destacadas como atípicas. Se for confirmado que as observações suspeitas de serem atípicas apresentam variações desproporcionais nos coeficientes estimados da regressão ou causam mudanças inferencias, deve-se inicialmente tentar amenizar ou mesmo eliminar esses impactos sem mudar o procedimento de estimação. Contudo, quando essas medidas tornam-se inócuas a aplicação de métodos de estimação robusta (ou resistente) pode ser uma opção a ser considerada. Neste tópico será apresentado apenas um tipo de estimador robusto, conhecido como estimador-M na classe de regressão linear múltipla. Este tipo de estimador é resistente a observações aberrantes sendo obtido através de procedimentos iterativos em que observações mais discrepantes recebem pesos menores com relação às demais observações. Eventualmnte esse tipo de estimador pode também

funcionar para outros tipos de observações atípicas.

1.16.1 Estimadores-M

Considere o modelo de regressão linear

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i,$$

em que ϵ_i, para $i = 1, \ldots, n$, são variáveis aleatórias independentes de média zero e variância σ^2. Note que está sendo relaxada a suposição de erros normais. Os estimadores-M são obtidos através da minimização de funções do tipo

$$S_\rho(\boldsymbol{\beta}) = \sum_{i=1}^{n} \rho(\epsilon_i), \tag{1.9}$$

em que $\epsilon_i = y_i - \mathbf{x}_i^\top \boldsymbol{\beta}$, para $i = 1, \ldots, n$ e $\rho(\epsilon)$ é uma função diferenciável em $\boldsymbol{\beta}$. Dependendo da escolha da função $\rho(\epsilon)$ e da distribuição dos erros, a minimização de (1.9) pode levar ao estimador de máxima verossimilhança. Por exemplo, se $\rho(\epsilon) = \frac{\epsilon^2}{2}$ (erros normais), tem-se em (1.9) o estimador de máxima verossimilhança (mínimos quadrados). Esse estimador é conhecido como estimador L_2. Quando $\rho(\epsilon) = \frac{|\epsilon|}{2}$ a minimização de (1.9) leva ao estimador de máxima verossimilhança da distribuição exponencial dupla ou distribuição de Laplace. O estimador obtido nesse caso é conhecido como estimador L_1.

Um dos estimadores mais conhecidos em métodos robustos é o estimador de Huber que é uma mistura entre os estimadores L_1 e L_2, sendo definido por

$$\rho(\epsilon) = \begin{cases} \frac{1}{2}\epsilon^2 & \text{para} \quad |\epsilon| \leq c \\ c\{|\epsilon| - \frac{c}{2}\} & \text{para} \quad |\epsilon| > c, \end{cases}$$

em que $c > 0$ é uma constante apropriada. Quando $c \to \infty$ tem-se o estimador L_2 e quando $c \to 0$ tem-se o estimador L_1. Outros estimadores robustos, tais como estimadores de Ramsay, de Andrews ou de Hampel são descritos em Montgomery et al. (2021, Cap. 15).

1.16.2 Estimação

Um problema com a minimização de (1.9) é que a solução pode não ser invariante com mudanças de escala dos regressores. Ou seja, se os regressores forem multiplicados por constantes a solução pode não continuar sendo a mesma. Assim, uma solução proposta é considerar no lugar de (1.9) a seguinte função objetivo:

$$S_\rho(\boldsymbol{\beta}) = \sum_{i=1}^{n} \rho(z_i), \tag{1.10}$$

em que $z_i = \frac{\epsilon_i}{s}$, com s sendo uma estimativa robusta de escala de modo que a solução em (1.10) seja invariante com mudanças de escala nos regressores. Uma escolha bastante conhecida para s é o desvio absoluto da mediana (vide Montogomery et al., 2021, Cap. 15) definido por

$$s = \text{mediana}|\epsilon_i - \text{mediana}(\epsilon_i)|/0,6745,$$

para $i = 1, \ldots, n$. A constante $0,6745$ faz com que s seja um estimador não tendencioso de σ se os erros são assumidos normais.

O estimador tipo M é obtido minimizando (1.10) cujas equações de estimação são dadas por

$$\mathbf{U}_\beta = \frac{\partial S_\rho(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}.$$

Definindo $\psi(z) = \rho'(z) = d\rho(z)/dz$, então para cada componente β_j tem-se o seguinte:

$$\begin{aligned} \mathrm{U}_{\beta_j} &= \frac{\partial S_\rho(\boldsymbol{\beta})}{\partial \beta_j} \\ &= \sum_{i=1}^{n} \frac{d\rho(z_i)}{dz_i} \frac{\partial z_i}{\partial \beta_j} \\ &= -\sum_{i=1}^{n} x_{ij} \psi(z_i)/s \end{aligned}$$

$$\propto \quad -\sum_{i=1}^{n} x_{ij}\omega_i(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}),$$

para $j = 1, \ldots, p$, em que $\omega_i > 0$ é um peso correspondente à i-ésima observação definido por

$$\omega_i = \begin{cases} \psi\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{s}\right) / \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{s} & \text{se} \quad y_i \neq \mathbf{x}_i^\top \boldsymbol{\beta} \\ 1 & \text{se} \quad y_i = \mathbf{x}_i^\top \boldsymbol{\beta}, \end{cases}$$

para $i = 1, \ldots, n$. Esses pesos serão estimados para cada observação no processo de estimação.

Em forma matricial as equações de estimação ficam dadas por

$$\mathbf{U}_\beta = \mathbf{X}^\top \mathbf{W}\{\mathbf{y} - \mathbf{X}^\top \widehat{\boldsymbol{\beta}}\} = \mathbf{0},$$

em que $\mathbf{W} = \text{diag}\{\omega_1, \ldots, \omega_n\}$. Essas equações são resolvidas através do processo iterativo de mínimos quadrados reponderados

$$\boldsymbol{\beta}^{(m+1)} = \{\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X}\}^{-1} \mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{y},$$

para $m = 0, 1, 2, \ldots$. Valor inicial $\boldsymbol{\beta}^{(0)}$ pode ser a estimativa da regressão L_2. Para n grande tem-se que $\widehat{\boldsymbol{\beta}} \sim \text{N}_p(\boldsymbol{\beta}, \text{Var}(\widehat{\boldsymbol{\beta}}))$, em que

$$\text{Var}(\widehat{\boldsymbol{\beta}}) = \text{fator}_c\{\sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}\},$$

com

$$\text{fator}_c = \frac{\text{E}\{\psi^2(\epsilon/\sigma)\}}{[\text{E}\{\psi'(\epsilon/\sigma)\}]}.$$

Uma estimativa para a matriz de variância-covariância de $\widehat{\boldsymbol{\beta}}$ descrita em Montgomery et al. (2021, Cap. 15) é dada por

$$\widehat{\text{Var}}(\widehat{\boldsymbol{\beta}}) = \frac{n\hat{s}^2}{n-p} \frac{\sum_{i=1}^{n} \psi^2\{(y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}})/\hat{s}\}}{[\sum_{i=1}^{n} \psi'\{(y_i - \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}})/\hat{s}\}]^2} (\mathbf{X}^\top \mathbf{X})^{-1}, \tag{1.11}$$

em que $\hat{s}$ é a estimativa robusta de escala. As estimativas assintóticas das variâncias e covariâncias de $\widehat{\boldsymbol{\beta}}$ devem ser extraídas de (1.11).

1.16.3 Função de Influência

A função $\psi(z) = \rho'(z)$, também conhecida como função de influência, desenpenha um papel importante em estimação robusta, uma vez que avalia o comportamento de $\rho'(z)$ à medida que $|z|$ aumenta. Assim, espera-se para os estimadores robustos que $\psi(z)$ fique limitada para valores altos de $|z|$. Por exemplo, para o estimador L_1 a função de influência fica dada por

$$\psi(z) = \rho'(z) = \frac{d}{dz}(|z|) = \text{sinal}(z),$$

sendo portanto uma função limitada em $[-1, 1]$ (Figura 1.46). Para a regressão L_2 tem-se que

$$\psi(z) = \rho'(z) = \frac{d}{dz}(\frac{1}{2}z^2) = z.$$

Ou seja, $\psi(z)$ é uma reta passando pela origem, logo é ilimitada (Figura 1.47).

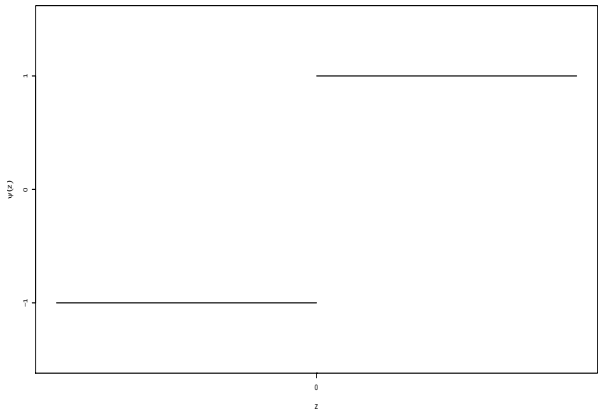

Figura 1.46: Função de influência para o estimador L_1.

Para o estimador de Huber a função de influência fica expressa na forma

$$\psi(z) = \begin{cases} z & \text{para} \quad |z| \leq c \\ c * \text{sinal}(z) & \text{para} \quad |z| > c. \end{cases}$$

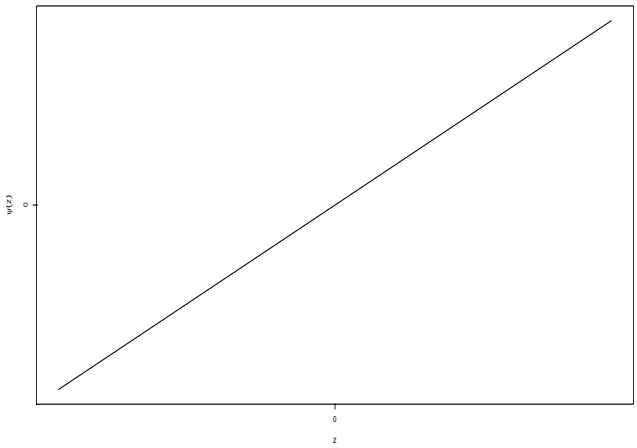

Figura 1.47: Função de influência para o estimador L_2.

Portanto, $\psi(z)$ é uma função limitada em $[-c, c]$ (Figura 1.48).

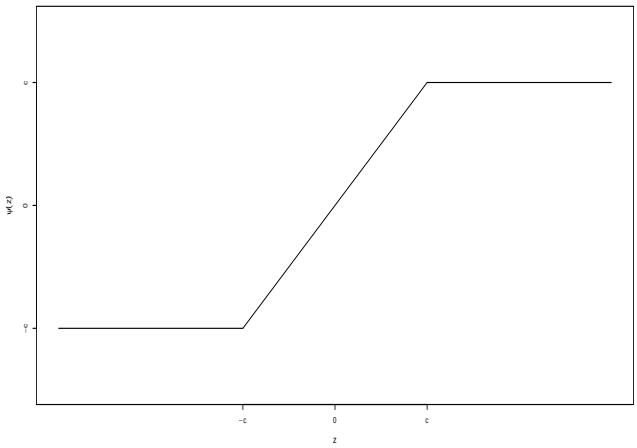

Figura 1.48: Função de influência para o estimador de Huber.

1.16.4 Pesos

Os pesos $\omega_i's$, que são estimados através do processo iterativo de mínimos qudrados reponderados, indicam a importância de cada observação no processo de estimação. Esses pesos agora são estimados ao invés de serem pré determinados como no caso da regressão linear ponderada (Seção 1.10). Por exemplo, na regressão L_2 os pesos ficam dados por

$$\omega_i = \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})/s}{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})/s} = 1, \forall i.$$

Logo, todas as observações recebem o mesmo peso. Na regressão L_1 os pesos assumem a forma

$$\begin{aligned} \omega_i &= \frac{\text{sinal}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})/s}{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})/s} \\ &= 1/|y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|, \end{aligned}$$

supondo $y_i \neq \mathbf{x}_i^\top \boldsymbol{\beta}$, para $i = 1, \ldots, n$. Portanto, o peso de cada observação é o inverso do valor absoluto do resíduo ordinário. Finalmente, na regressão de Huber tem-se que

$$\omega_i = \begin{cases} 1 & \text{se} \quad |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|/s \leq c \\ \frac{cs}{|y_i - \mathbf{x}_i^\top \beta|} & \text{se} \quad |y_i - \mathbf{x}_i^\top \boldsymbol{\beta}|/s > c, \end{cases}$$

para $i = 1, \ldots, n$. Nesse caso os pesos são uma mistura entre os pesos das regressões L_1 e L_2. Portanto, tem-se que os estimadores L_1 e de Huber são resistentes a observações aberrantes.

1.16.5 Aplicação

Como ilustração neste tópico considere o exemplo descrito em Montgomery et al. (2021, Cap.2) em que uma engarrafadora de refrigerantes está analisando o serviço de abastecimento das máquinas de refrigerantes atendidas

pela empresa. O serviço de abastecimento inclui o estoque das garrafas nas máquinas e pequenas manutenções feitas pelo próprio motorista do veículo com os carregamentos. O engenheiro industrial responsável pela logística da distribuição dos refrigerantes acredita que o tempo gasto (em minutos) pelo motorista para o abastecimento das máquinas pode estar relacionado com a distância percorrida pelo motorista do veículo até as máquinas (em pés) e pelo número de caixas de produtos estocados. Uma amostra aleatória de 25 abastecimentos foi considerada para análise. Os dados estão descritos no arquivo **delivery.txt**.

Na Figura 1.49 tem-se os diagramas de dispersão entre o tempo gasto pelo motorista e o número de caixas estocadas e a distância percorrida pelo motorista, respectivamente. Nota-se tendências aproximadamente lineares, sugerindo o seguinte modelo:

$$y_i = \beta_1 + \beta_2 \text{ncaixas}_i + \beta_3 \text{distância}_i + \epsilon_i, \tag{1.12}$$

para $i = 1, \ldots, 25$, em que y_i denota o tempo gasto pelo i-ésimo motorista com $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$. Na Tabela 1.22 tem-se as estimativas do ajuste e nota-se que todos os efeitos são altamente significtivos.

Na Figura 1.50 tem-se os gráficos de diagnóstico com a observação #9 sendo destacada como aberrante e influente. Refere-se ao abastecimento com os maiores valores para a resposta e para as variáveis explicativas. A fim de reduzir a influência dessa observação nas estimativas dos parâmetros o método de Huber é aplicado com $c = 1,345$ cujas estimativas são apresentadas na Tabela 1.23. Todos os efeitos são altamente significativos.

Nota-se pela Tabela 1.24 que a observação #9 recebe o menor peso através do processo de estimação, porém outras observações também têm o peso alterado com relação ao procedimento de mínimos quadrados. Na Figura 1.51 tem-se o gráfico do resíduo Studentizado da regressão L_2 contra os pesos

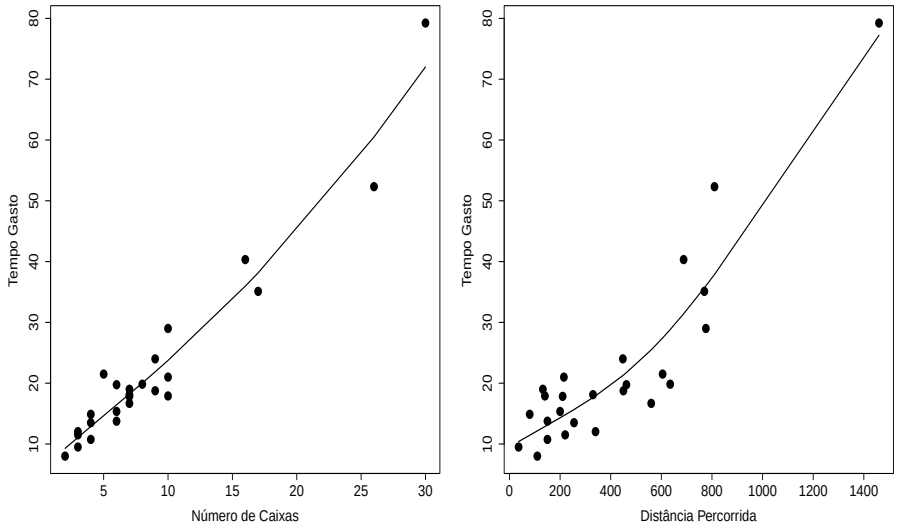

Figura 1.49: Diagramas de dispersão (com tendência) entre o tempo gasto pelo motorista e o número de caixas estocadas (esquerdo) e a distância percorrida pelo motorista (direito).

Tabela 1.22: Estimativas dos parâmetros referentes ao modelo (1.12) ajustado pelo método de mínimos quadrados aos dados sobre abastecimento de refrigerantes.

Efeito	Estimativa	Erro padrão	valor-t	valor-P
Constante	2,341	1,097	2,13	0,044
Ncaixas	1,616	0,171	9,47	0,001
Distância	0,014	0,004	3,89	0,000
s	3,259			
R^2	0,96			
$\overline{R}^2$	0,96			

estimados pelo método de Huber, e pode ser observado que as observações com resíduos altos em geral recebem pesos menores, confirmando a resistência

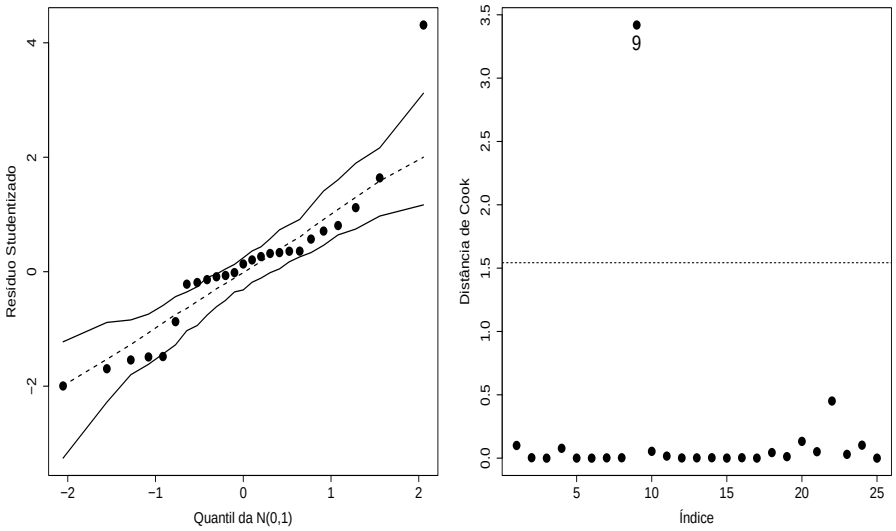

Figura 1.50: Gráfico normal de probabilidades e distância de Cook (k=2) referentes ao ajuste do modelo (1.12) aos dados sobre abastecimento de refrigerantes.

Tabela 1.23: Estimativas dos parâmetros referentes ao modelo (1.12) ajustado pelo método de Huber aos dados sobre abastecimento de refrigerantes.

Efeito	Estimativa	Erro padrão	valor-z	valor-P
Constante	3,469	0,841	4,12	0,000
Ncaixas	1,465	0,131	11,19	0,000
Distância	0,015	0,003	5,27	0,000
s	1,536			

do procedimento de estimação com relação a observações aberrantes.

Finalmente, tem-se na Tabela 1.25 a comparação entre estimativas e nota-se que as maiores correções pelo método de Huber com relação ao método de mínimos quadrados ocorrem na estimativa do intercepto e do coeficiente

Tabela 1.24: Pesos estimados das observações do exemplo sobre abastedimento de refrigerantes através do processo iterativo pelo método de Huber.

Obs	Peso	Obs	Peso	Obs	Peso
1	0,395	2	1	3	1
4	0,472	5	1	6	1
7	1	8	1	9	0,197
10	1	11	0,614	12	1
13	1	14	1	15	1
16	1	17	1	18	0,618
19	1	20	0,456	21	0,912
22	1	23	0,461	24	0,446
25	1				

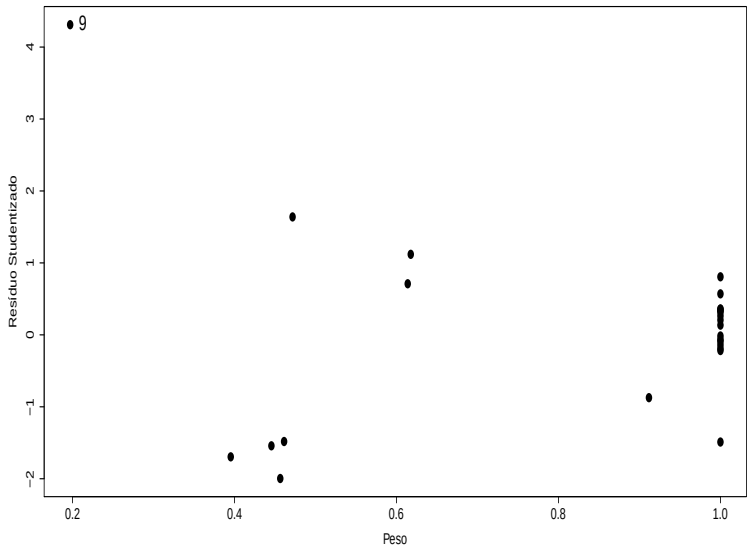

Figura 1.51: Gráfico entre o resíduo Studentizado do ajuste de mínimos quadrados do modelo (1.12) aos dados sobre abastecimento de refrigerantes e os pesos estimados pelo método de Huber.

do número de caixas. Esta última estimativa muito similar à estimativa de mínimos quadrados quando a observação #9 é excluída.

Tabela 1.25: Comparação das estimativas dos parâmetros pelos métodos de mínimos quadrados e Huber referentes ao exemplo sobre abastecimento de refrigerantes.

Efeito	L_2	L_2(-#9)	Huber
Constante	2,341	4,447	3,469
Ncaixas	1,616	1,498	1,465
Distância	0,014	0,010	0,015

1.17 Métodos de Reamostragem

Os métodos tradicionais de inferência estatística para parâmetros populacionais nem sempre podem ser aplicados com segurança, uma vez que dependem do conhecimento das propriedades dos estimadores e do tamanho amostral. O caso mais trivial de inferência para um único parâmetro populacional θ depende da existência de algum estimador T com propriedades ótimas. Na maioria dos casos essas propriedades são conhecidas apenas assintoticamente, sendo em geral complexo ou mesmo intratável a inferência exata. Assim, métodos alternativos de inferência para tamanhos amostrais fixos tornam-se mais atrativos, dentre os quais destacam-se os procedimentos de reamostragem.

O método jackknife (Quenouille, 1956) é o procedimento mais simples e talvez mais antigo de reamostragem, que consiste em construir um estimador para θ através de n estimadores obtidos da amostra, retirando uma observação de cada vez. Esses estimadores são agrupados para gerar o estimador jackknife. Mais recentemente, (Efron, 1979) propôs uma extensão

do estimador jackknife em que R amostras são retiradas com reposição da amostra gerando R estimadores do parâmetro θ, os quais são agrupados para gerar o estimador denominado *bootstrap*. Em ambos os casos aplica-se o TCL para os procedimentos inferênciais.

Esses métodos podem ser estendidos para modelagem de regressão, contudo requerem que o modelo postulado para ajustar os dados não esteja especificado de forma incorreta. Essa suposição pode ser facilmente verificada através de análise de resíduos. Por exemplo, pelo método *bootstrap*, o procedimento consiste em reamostrar com reposição R vezes dos dados e para cada amostra ajustar o modelo postulado. Tem-se portanto R estimativas para cada um dos coeficientes da regressão e supondo R suficientemente grande, pode-se obter estimativas intervalares para os coeficientes da regressão aplicando o TCL.

1.17.1 Estimador *jackknife*

Supor que o parâmetro populacional θ está sendo estimado por um estimador T e que foi observada a amostra $S = \{x_1, \ldots, x_n\}$, em que x_i denota o i-ésimo valor observado da variável aleatória X. A ideia é obter n estimativas para a estatística T considerando $n-1$ observações de cada vez, em que a i-ésima observação não é considerada. Assim, para a i-ésima amostra jackknife tem-se o espaço amostral $S_i = \{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\}$ sendo $T_{(i)} = T(S_i)$, $i = 1, \ldots, n$. O estimador jackknife para θ é definido por

$$T_{\text{jack}} = \frac{1}{n} \sum_{i=1}^{n} T_{(i)}.$$

Contudo, fazendo correção de viés utiliza-se o estimador corrigido

$$T^*_{\text{jack}} = nT - (n-1)T_{\text{jack}}.$$

Uma forma alternativa de obter o estimador jackknife corrigido é considerando os pseudo-valores

$$T^*_{(i)} = nT - (n-1)T_{(i)},$$

sendo

$$T^*_{\text{jack}} = \frac{1}{n}\sum_{i=1}^{n} T^*_{(i)}.$$

A variância estimada para T^*_{jack} é obtida como sendo a média da variância estimada de $T^*_{(i)}$, ou seja $\widehat{\text{Var}}(T^*_{\text{jack}}) = s^2_{\text{jack}}/n$, em que

$$\begin{aligned} s^2_{\text{jack}} &= \frac{1}{n-1}\sum_{i=1}^{n}(T^*_{(i)} - T^*_{\text{jack}})^2 \\ &= (n-1)\sum_{i=1}^{n}(T_{(i)} - T_{\text{jack}})^2. \end{aligned}$$

Assim, obtém-se

$$\widehat{\text{Var}}(T^*_{\text{jack}}) = \frac{n-1}{n}\sum_{i=1}^{n}(T_{(i)} - T_{\text{jack}})^2.$$

Para n suficientemente grande segue pelo TCL que uma estimativa intervalar de coeficiente $(1-\alpha)$ para θ fica dada por

$$[T^*_{\text{jack}} \pm z_{(1-\alpha/2)}\widehat{\text{EP}}(T^*_{\text{jack}})],$$

em que $\widehat{\text{EP}}(T^*_{\text{jack}}) = \sqrt{\widehat{\text{Var}}(T^*_{\text{jack}})}$ e $z_{(1-\alpha/2)}$ denota o quantil $(1-\alpha/2)$ da distribuição N(0,1).

1.17.2 *Bootstrap* Não Paramétrico

O procedimento *bootstrap* não paramétrico (vide, por exemplo, Fox e Weisberg, 2019) permite o estudo da distribuição amostral de um estimador T do parâmetro populacional θ sem fazer suposições para a distribuição

amostral de T. A ideia é retirar R amostras com reposição da amostra $\mathrm{S} = \{x_1, \ldots, x_n\}$, sendo a b-ésima amostra *bootstrap* denotada por $\mathrm{S}_b^* = \{x_{b1}^*, \ldots, x_{bn}^*\}$ e a estatística T será calculada para cada amostra *bootstrap*, cujos valores serão dados por $T_b^* = T(\mathrm{S}_b^*)$, para $b = 1, \ldots, R$. Assim, pode-se calcular algumas quantidades tais como a média amostral e a variância amostral da estatística *bootstrap* T^* dadas, respectivamente, por

$$\begin{aligned} \overline{T}^* &= \frac{\sum_{b=1}^{R} T_b^*}{R} \text{ e} \\ \widehat{\mathrm{Var}}(T^*) &= \frac{1}{R-1}\sum_{b=1}^{R}(T_b^* - \overline{T}^*)^2. \end{aligned}$$

A estimativa *bootstrap* do viés de T é dada por $\widehat{B}^* = \overline{T}^* - T$. Ou seja, $\widehat{B}^*$ é uma estimatia de $\mathrm{E}(T) - \theta$. Tem-se que $\widehat{\mathrm{EP}}(T^*) = \sqrt{\widehat{\mathrm{Var}}(T^*)}$ é também considerada uma estimativa do erro padrão da estatística T. Assim, supondo R grande usando o TCL e corrigindo o viés da estatística T, uma estimativa intervalar de coeficiente $(1-\alpha)$ para θ é dada por

$$[T - \widehat{B}^* \pm \mathrm{z}_{(1-\alpha/2)}\widehat{\mathrm{EP}}(T^*)],$$

que é conhecida como estimativa intervalar normal *bootstrap*.

Alternativamente, tem-se a estimativa intervalar percentílica *booststrap* dada por

$$T_{(\mathrm{inf})}^* < \theta < T_{(\mathrm{sup})}^*,$$

em que $T_{(1)}^* \leq T_{(2)}^* \leq \ldots \leq T_{(R)}^*$ denotam os percentis *bootstrap* da estatística T^*, inf = [(R+1)α/2], sup = [(R+1)(1 − α/2)] e [·] indica o inteiro mais próximo. Por exemplo, supondo R=999 e $\alpha = 0,05$ tem-se que inf=25 e sup=975.

Finalmente, tem-se o método BC_a (viés-corrigido e acelerado) *bootstrap* que é recomendado para situações em que o estimador *bootstrap* T^* é relativamente viesado com distribuição amostral assimétrica. Nesses casos calcula-se

inicialmene um tipo de correção para o viés definida por

$$z_0 = \Phi^{-1}\left(\frac{\#(T_b^* \leq T)}{R}\right),$$

em que $\Phi(\cdot)$ denota a fda da distribuição $\mathrm{N}(0,1)$. Em seguida, tem-se o fator de aceleração que leva em conta a assimetria da estatística T^*. Esse fator é estimado pelo método jackknife que envolve sistematicamente a utilização da estatística T sem uma observação, sendo definido por

$$a = \frac{\sum_{i=1}^{n}(\overline{T} - T_{(i)})^3}{6\{\sum_{i=1}^{n}(\overline{T} - T_{(i)})^2\}^{\frac{3}{2}}},$$

em que $T_{(i)}$ denota a estatística T sem a i-ésima observação e $\overline{T} = \sum_{i=1}^{n} T_{(i)}/n$. Calcula-se então os novos fatores

$$\begin{aligned} a_1 &= \Phi\left(z_0 + \frac{z_0 + z_{\alpha/2}}{1 - a(z_0 + z_{\alpha/2})}\right) \quad \text{e} \\ a_2 &= \Phi\left(z_0 + \frac{z_0 + z_{(1-\alpha/2)}}{1 - a(z_0 + z_{(1-\alpha/2)})}\right). \end{aligned}$$

A estimativa intervalar percentílica *booststrap* corrigida fica dada por

$$T^*_{(\inf^*)} < \theta < T^*_{(\sup^*)},$$

em que $\inf^* = [Ra_1]$ e $\sup^* = [Ra_2]$. Mostra-se facilmente que quando $z_0 = 0$ e $a = 0$ as estimativas intervalares *bootstrap* percentílica e percentílica corrigida coincidem.

1.17.3 Extensão para Regressão

No contexto de regressão o espaço amostral é formado pelos valores da variável resposta Y e pelos valores das variáveis explicativas $(X_1, \ldots, X_p)$, sendo denotado por $\mathrm{S} = \{z_1, \ldots, z_n\}$, em que $z_i = (y_{i1}, x_{i1}, \ldots, x_{ip})$ para

$i = 1, \ldots, n$. No caso de aplicação do método *bootstrap*, amostra-se R vezes com reposição de S sendo a b-ésima amostra denotada por $\mathrm{S}_b^* = \{z_{b1}^*, \ldots, z_{bn}^*\}$, para $b = 1, \ldots, R$. Calcula-se então as estimativas *bootstrap* $\beta_1^*, \ldots, \beta_p^*$ dos coeficientes da regressão e os respectivos erros padrão amostral. As estimativas intervalares *bootstrap* para os coeficientes da regressão são obtidas similarmente ao caso de um único parâmetro populacional descrito na seção anterior.

Há várias bibliotecas na plataforma R com procedimentos de reamostragem. Por exemplo, a biblioteca `boot` (Canty et al., 2024) é bastante flexível para a análise de uma grande variedade de dados. No contexto de regressão, a função `Boot` da biblioteca `car` (Fox e Weisberg, 2024), também disponível no R, permite uma aplicação direta de *bootstrap* em várias classes de regressão, incluindo a regressão linear e modelos lineares generalizados.

1.17.4 Aplicação

Como ilustração de aplicação do procedimento *bootstrap* considere o submodelo selecionado na Seção 1.14, em que as variáveis explicativas Gastos, Nclientes e Marcas foram selecionadas para explicar o total de telhados vendidos. Na Tabela 1.26 são apresentadas as estimativas intervalares de 95% para os coeficientes da regressão linear das três variáveis explicativas, usando o método de mínimos quadrados e os três métodos *bootstrap* os quais serão denotados por normal, percentílico e BC_a.

Nota-se uma forte concordância entre as estimativas intervalares *bootstrap* com as estimativas intervalares de mínimos quadrados. Essa concordância é reforçada pela Figura 1.52, extraída da biblioteca `car` da plataforma R, em que nota-se excelente aproximação entre as distribuições empíricas das estimativas de mínimos quadrados dos coefiecentes da regressão e a distri-

Tabela 1.26: Estimativas intervalares de 95% para os coeficientes da regressão linear das variáveis explicativas Gastos, Nclientes e Marcas usando o método de mínimos quadradose e os procedimentos *bootstrap* normal, percentílico e BC_a.

Efeito	M.Quadrados	Normal	Percentílico	BC_a
Gastos	[-0,38;3,74]	[-0,57;3,91]	[-0,54;3,87]	[-0,53;3,89]
Nclientes	[3,09;3,65]	[3,10;3,68]	[3,04;3,61]	[3,07;3,64]
Marcas	[-22,73;-19,70]	[-23,12;-19,37]	[-23,04;-19,34]	[-23,28;-19,54]

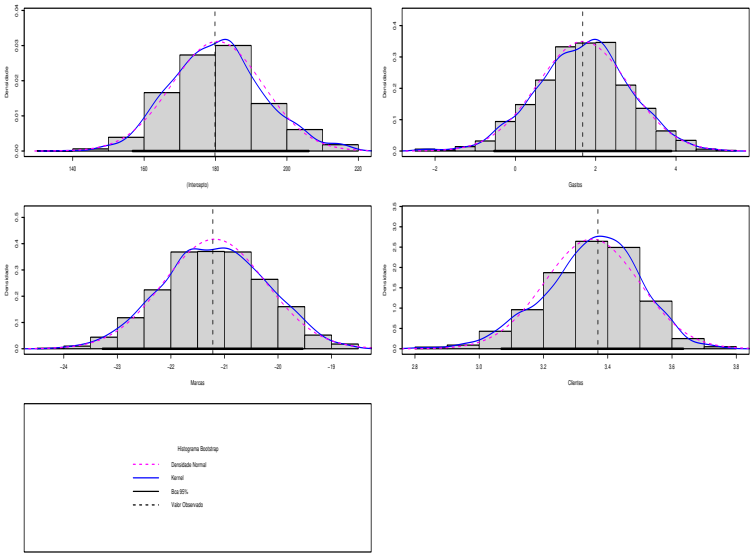

Figura 1.52: Distribuição empírica dos coeficientes estimados pelo método de mínimos quadrados através do procedimento *bootstrap* e estimativas intervalares BC_a de 95% em torno das estimativas de mínimos quadrados

.

buição normal. As estimativas intervalares BC_a (em negrito) aparecem em geral simétricas em torno das estimativas de mínimos quadrados, indicando

que não há necessidade de correções importantes de assimetria e viés das estimativas de mínimos quadrados. A única estimativa intervalar BC_a que não cobre o valor zero refere-se ao coeficiente da variável explicativa Gastos que foi mantida no modelo uma vez que seu efeito foi mascarado por uma observação.

1.18 Regressão Não Linear

Os modelos de regressão não linear podem ser expressos na seguinte forma:

$$y = f(\boldsymbol{\theta}; \mathbf{x}) + \epsilon, \tag{1.13}$$

em que y denota o valor observado da variável resposta, $f(\boldsymbol{\theta}; \mathbf{x})$ é uma função não linear nos parâmetros $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$, $\mathbf{x}$ contém valores de variáveis explicativas e ϵ é um erro aditivo. Recupera-se o modelo linear quando $f(\boldsymbol{\theta}; \mathbf{x}) = \mathbf{x}^\top \boldsymbol{\theta}$. Diferentemente dos modelos lineares, os modelos não lineares descritos em (1.13) apresentam algumas características particulares:

- A função $f(\boldsymbol{\theta}; \mathbf{x})$ é conhecida e em geral desenvolvida através de suposições teóricas, por exemplo equações diferenciais.
- Os parâmetros têm alguma interpretação, por exemplo física, biológica ou econométrica. Logo, a aproximação dessas funções por outras funções mais simples pode levar à perda da interpretação paramétrica.
- Essas funções podem ter formas equivalentes obtidas através de reparametrizações de $\boldsymbol{\theta}$. Essas reparametrizações são utilizadas para reduzir o viés dos estimadores de $\boldsymbol{\theta}$.
- A estimação de $\boldsymbol{\theta}$ é obtida através de procedimentos iterativos.
- As propriedades dos estimadores de $\boldsymbol{\theta}$ são em geral assintóticas.

Vários modelos não lineares são descritos em Ratkowsky (1983) e Seber e Wild (1989). Alguns exemplos são descritos a seguir.

1.18.1 Modelo de von Bertalanffy

Este modelo, que é uma curva de crescimento, tem sido aplicado na área de Ecologia para explicar o comprimento esperado de uma espécie de peixe dada sua idade. Uma das formas mais utilizadas do modelo é a seguinte:

$$y = \theta_1[1 - \exp\{-\theta_2(x - \theta_3)\}] + \epsilon,$$

em que y denota o comprimento do peixe, x denota a respectiva idade, enquanto $\theta_1 > 0$ representa o comprimento máximo esperado para a espécie (assíntota), $\theta_2 > 0$ denota a taxa média de crescimento e θ_3 é um valor nominal em que o comprimento esperado da espécie é zero. Tem-se na Figura 53 a descrição de um exemplo da curva de von Bertalanffy.

As curvas de crescimento apresentam formas equivalentes obtidas através de reparametrizações, que podem ser aplicados dependendo da área de interesse ou mesmo para reduzir o viés da estimativa de máxima verossimilhança de $\boldsymbol{\theta}$. As funções abaixo, extraídas do livro de Fox e Weisberg (2019), são formas equivalentes de curvas de crescimento que recebem nomes diferentes dependendo da área:

1. $f_1(\boldsymbol{\theta}; x) = \theta_1 - \theta_3\theta_2^x$
2. $f_2(\boldsymbol{\theta}; x) = \theta_1 - \theta_3\exp(-\theta_2 x)$
3. $f_3(\boldsymbol{\theta}; x) = \theta_1 + (\theta_3 - \theta_1)\theta_2^x$
4. $f_4(\boldsymbol{\theta}; x) = \theta_1 + (\theta_3 - \theta_1)\exp(-\theta_2 x)$
5. $f_5(\boldsymbol{\theta}; x) = \theta_1 - \exp\{-(\theta_3 + \theta_2 x)\}$

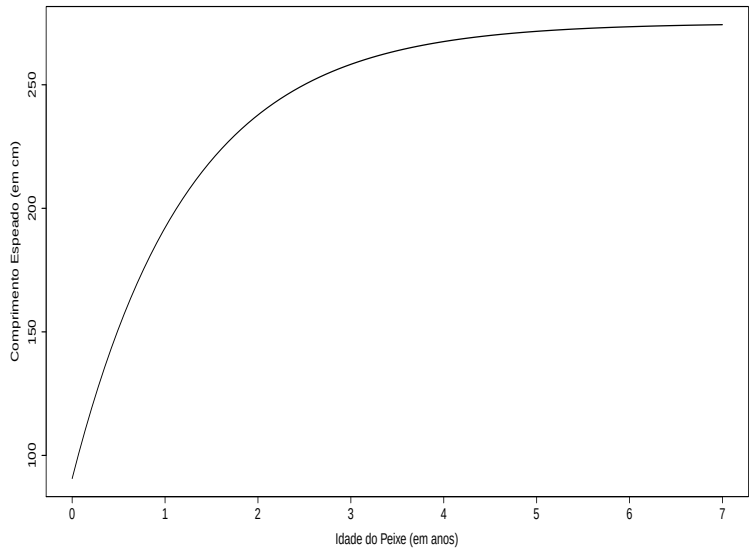

Figura 1.53: Curva de von Bertalanffy para $\theta_1 = 275$, $\theta_2 = 0.5$ e $\theta_3 = -0.5$.

6. $f_6(\boldsymbol{\theta}; x) = \theta_1 + \theta_3\{1 - \exp(-\theta_2 x)\}$,

em que θ_1 denota a assíntota (valor esperado quando $x \to \infty$) e θ_2 denota a taxa média de crescimento em todos os modelos. Nos modelos 1,2 e 6 tem-se $\theta_3 = \theta_1 - \mu$, em que μ denota o valor esperado $\mathrm{E}(Y|x)$ quando $x = 0$, enquanto nos modelos 3 e 4 tem-se $\theta_3 = \mu$ e no modelo 5 $\theta_3 = \log(\theta_1 - \mu)$.

Se qualquer um desses 6 modelos for ajustado ao mesmo conjunto de dados, a curva ajustada será a mesma (invariância dos valores preditos), contudo as estimativas dos parâmetros, respectivos erros padrão e vieses deverão ser diferemtes. Assim, pode-se optar pelo modelo cujas estimativas tenham os menores vieses. Nesse tipo de modelo há dois tipos de não linearidade, paramétrica e intrínseca. A principal diferenca é que a não linearidade paramétrica pode sempre ser reduzida com reparametrizações, enquanto a não linearidade intrínseca é invariante com reparametrizações.

Uma sugestão de valores inciais para o modelo de von Bertalanffy é considerar $\theta_1^{(0)} \cong y_{\max}$, com os parâmetros θ_2 e θ_3 sendo definidos através da relação

$$\log(1 - y/\theta_1^{(0)}) \cong \gamma + \eta x,$$

em que $\gamma = \theta_2\theta_3$ e $\eta = -\theta_2$. Os valores iniciais $\theta_2^{(0)}$ e $\theta_3^{(0)}$ podem ser obtidos do ajuste de mínimos quadrados de $z = \log\{1 - y/\theta_1^{(0)}\}$ contra $\gamma + \eta x$.

1.18.2 Modelo de Crescimento Logístico

Esse modelo sigmoidal é frequentemente aplicado para estudar o crescimento populacional. Sua forma mais conhecida é dada por

$$y = \frac{\theta_1}{1 + \exp\{-(\theta_2 + \theta_3 x)\}} + \epsilon,$$

em que y denota o tamanho da população num dado ano x. O parâmetro $\theta_1 > 0$ representa o tamanho máximo esperado para a população (assíntota), θ_3 controla o crescimento da curva no intervalo $(0, \theta_1)$. Pode-se mostrar que a curva é simétrica em $x = -\theta_2/\theta_3$. Ou seja, $\mathrm{E}(Y|x = -\theta_2/\theta_3) = \frac{\theta_1}{2}$ que corresponde ao ponto médio entre as duas assíntotas. Um exemplo da curva logística é descrito na Figura 1.54.

Para valores iniciais a sugestão é considerar $\theta_1^{(0)} \cong y_{\max}$ com os parâmetros θ_2 e θ_3 sendo definidos tais que

$$\log\left(\frac{y/\theta_1^{(0)}}{1 - y/\theta_1^{(0)}}\right) \cong \theta_2 + \theta_3 x.$$

Logo, os valores iniciais $\theta_2^{(0)}$ e $\theta_3^{(0)}$ podem ser obtidos do ajuste de mínimos quadrados de $z = \log\{(y/\theta_1^{(0)})/(1 - y/\theta_1^{(0)})\}$ contra $\theta_2 + \theta_3 x$.

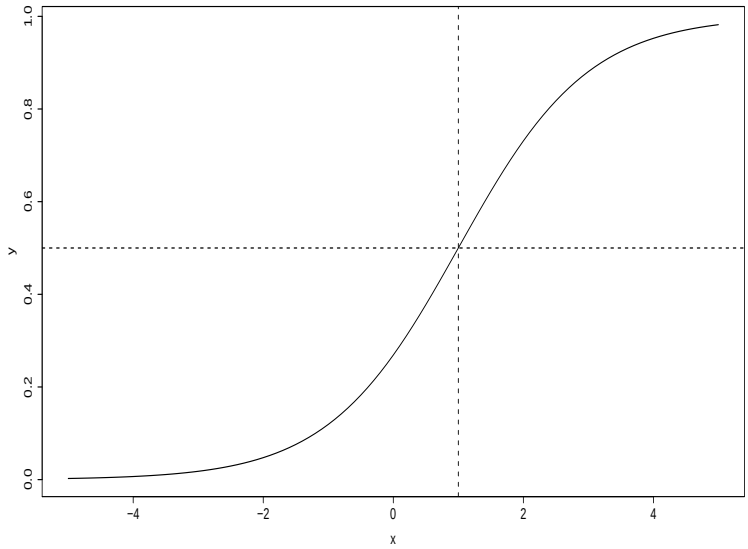

Figura 1.54: Curva Logística para $\theta_1 = 1$, $\theta_2 = -1$ e $\theta_3 = 1$.

1.18.3 Modelo de Mistura de Duas Drogas

O modelo de Finney para mistura de drogas tem sido aplicado na área de Farmacologia para avaliar a interação de duas drogas A e B de mesmo tipo, porém com princípios ativos diferentes. Assume a seguinte forma:

$$y = \alpha + \delta \log(x_1 + \rho x_2 + \kappa\sqrt{\rho x_1 x_2}) + \epsilon,$$

em que y denota o valor observado da resposta, x_1 e x_2 representam, respectivamente, as doses das drogas A e B, δ é a relação comum log(dose) e resposta, ρ é a potência da droga B em relação à droga A e κ denota a interação entre as duas drogas, sendo interpretado da seguinte maneira: $\kappa = 0$ efeitos aditivos, $\kappa > 0$ sinergismo e $\kappa < 0$ antagonismo.

1.18.4 Modelo de Michaelis-Menten

O modelo de Michaelis-Menten é muito aplicado em cinética química para relacionar a velocidade inicial de uma reação enzimática (Y) (contagem/min) dada a concentração de um substrato (X) (em ppm), sendo expresso na forma

$$y = \frac{\theta_1 x}{x + \theta_2} + \epsilon,$$

em que θ_1 denota a velocidade máxima obtida (assíntota) e θ_2 é conhecido como a constante de Michaelis. A curva de Michaelis-Menten é ilustrada na Figura 1.55 para um caso particular.

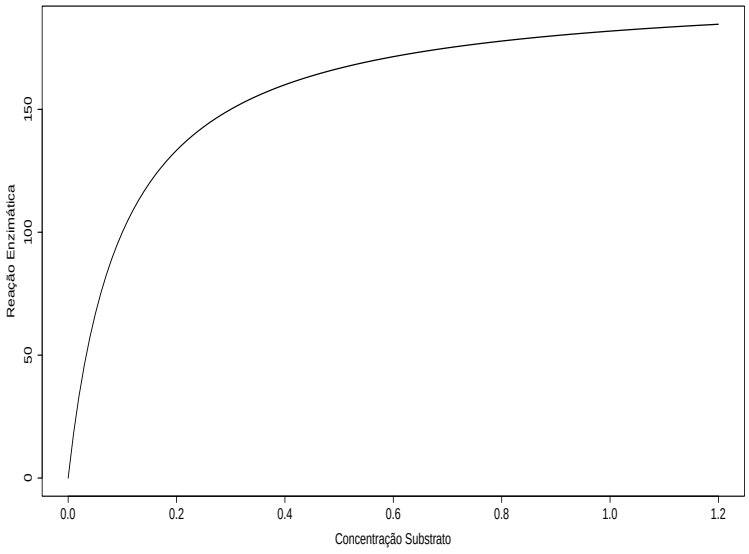

Figura 1.55: Curva de Michaelis-Menten para $\theta_1 = 200$ e $\theta_2 = 0,10$.

Para valores iniciais para o modelo de Michaelis-Menten utiliza-se a aproximação

$$\frac{1}{y} \cong \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1}\frac{1}{x}.$$

Assim, os valores iniciais $\theta_1^{(0)}$ e $\theta_2^{(0)}$ podem ser obtidos do ajuste de mínimos quadrados de y^{-1} contra $\beta_1 + \beta_2 x^{-1}$, em que $\beta_1 = 1/\theta_1$ e $\beta_2 = \theta_2/\theta_1$.

1.18.5 Estimação

Considere agora o modelo de regressão não linear

$$y_i = f(\boldsymbol{\theta}; \mathbf{x}_i) + \epsilon_i, \tag{1.14}$$

em que $y_1, \ldots, y_n$ denotam os valores observados da variável resposta, $f(\boldsymbol{\theta}; \mathbf{x}_i)$ é uma função não linear nos parâmetros $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$, $\mathbf{x}_i$ contém valores de variáveis explicativas e $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$. Similarmente à regressão linear, a estimação de $\boldsymbol{\theta}$ em (1.14) é obtida minimizando a seguinte função objetivo:

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \{y_i - f(\boldsymbol{\theta}; \mathbf{x}_i)\}^2 = \{\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})\}^\top \{\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})\},$$

em que $\mathbf{y} = (y_1, \ldots, y_n)^\top$ e $\mathbf{f}(\boldsymbol{\theta}) = \{f(\boldsymbol{\theta}; \mathbf{x}_1), \ldots, f(\boldsymbol{\theta}; \mathbf{x}_n)\}^\top$.

A derivada parcial de $S(\boldsymbol{\theta})$ com relação a $\boldsymbol{\theta}$ fica dada por

$$\frac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -2\mathbf{J}(\boldsymbol{\theta})^\top \{\mathbf{y} - \mathbf{f}(\boldsymbol{\theta})\},$$

em que $\mathbf{J}(\boldsymbol{\theta})$ é a matriz Jacobiana de dimensão $n \times p$ da transformação de $\mathbf{f}(\boldsymbol{\theta})$ com relação a $\boldsymbol{\theta}$, sendo denotada por

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial f(\theta; \mathbf{x}_1)}{\partial \theta_1} & \cdots & \frac{\partial f(\theta; \mathbf{x}_1)}{\partial \theta_p} \\ \vdots & \cdots & \vdots \\ \frac{\partial f(\theta; \mathbf{x}_n)}{\partial \theta_1} & \cdots & \frac{\partial f(\theta; \mathbf{x}_n)}{\partial \theta_p} \end{bmatrix}.$$

Portanto, como $\mathbf{J}(\boldsymbol{\theta})$ depende de $\boldsymbol{\theta}$ o estimador de mínimos quadrados (máxima verossimilhança) deve ser obtido iterativamente. O processo iterativo de

Newton-Raphson fica dado por

$$\begin{aligned}\boldsymbol{\theta}^{(m+1)} &= \boldsymbol{\theta}^{(m)} + \{\mathbf{J}(\boldsymbol{\theta}^{(m)})^\top \mathbf{J}(\boldsymbol{\theta}^{(m)})\}^{-1}\mathbf{J}(\boldsymbol{\theta}^{(m)})^\top\{\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}^{(m)})\} \\ &= \{\mathbf{J}(\boldsymbol{\theta}^{(m)})^\top \mathbf{J}(\boldsymbol{\theta}^{(m)})\}^{-1}\mathbf{J}(\boldsymbol{\theta}^{(m)})^\top \mathbf{z}(\boldsymbol{\theta}^{(m)}), \qquad (1.15)\end{aligned}$$

para $m = 0, 1, 2, \ldots$ e $\mathbf{z}(\boldsymbol{\theta}) = \mathbf{y} - \{\mathbf{f}(\boldsymbol{\theta}) - \mathbf{J}(\boldsymbol{\theta})\boldsymbol{\theta}\}$ é uma pseudo resposta ou variável dependente modificada. Ou seja, $\widehat{\boldsymbol{\theta}}$ é obtido através de um processo iterativo de mínimos quadrados, contudo valores iniciais $\boldsymbol{\theta}^{(0)}$ são necessário para iniciar o processo iterativo.

1.18.6 Inferência

Mostra-se para n grande que $\widehat{\boldsymbol{\theta}}$ segue aproximadamente distribuição normal p-variada de média $\boldsymbol{\theta}$ e matriz de variância-covariância dada por

$$\mathrm{Var}(\widehat{\boldsymbol{\theta}}) = \sigma^2\{\mathbf{J}(\boldsymbol{\theta})^\top \mathbf{J}(\boldsymbol{\theta})\}^{-1},$$

sendo o estimador para σ^2 definido por

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n}\{y_i - f(\widehat{\boldsymbol{\theta}}; \mathbf{x}_i)\}^2}{n-p}.$$

Se o interesse é testar as hipóteses $\mathrm{H}_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{0}$ contra $\mathrm{H}_1 : \mathbf{R}\boldsymbol{\theta} \neq \mathbf{0}$, em que $\mathbf{R}$ é uma matriz de dimensão $r \times p$ e posto linha completo $r \leq p$, tem-se sob H_0 e para n grande que

$$\mathrm{F} = \frac{(\mathbf{R}\widehat{\boldsymbol{\theta}})^\top[\mathbf{R}\{\mathbf{J}(\widehat{\boldsymbol{\theta}})^\top \mathbf{J}(\widehat{\boldsymbol{\theta}})\}^{-1}\mathbf{R}^\top]^{-1}(\mathbf{R}\widehat{\boldsymbol{\theta}})}{r\widehat{\sigma}^2} \sim \mathrm{F}_{r,(n-p)}.$$

Logo, para um nível de significância $0 < \alpha < 1$, rejeita-se H_0 se $\mathrm{F} > \mathrm{F}_{(1-\alpha),(p-1),(n-p)}$, em que $\mathrm{F}_{(1-\alpha),(p-1),(n-p)}$ denota o quantil $(1-\alpha)$ de uma distribuição F com $(p-1)$ e $(n-p)$ graus de liberdade.

1.18.7 Métodos de Diagnóstico

Na convergência do processo iterativo (1.15) tem-se que

$$\widehat{\boldsymbol{\theta}} = \{\mathbf{J}(\widehat{\boldsymbol{\theta}})^\top \mathbf{J}(\widehat{\boldsymbol{\theta}})\}^{-1} \mathbf{J}(\widehat{\boldsymbol{\theta}})^\top \mathbf{z}(\widehat{\boldsymbol{\theta}}).$$

Portanto, similarmente à regressão linear, pode-se escrever $\mathbf{J}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\theta}} = \mathbf{H}(\widehat{\boldsymbol{\theta}})\mathbf{z}(\widehat{\boldsymbol{\theta}})$, em que

$$\mathbf{H}(\widehat{\boldsymbol{\theta}}) = \mathbf{J}(\widehat{\boldsymbol{\theta}})\{\mathbf{J}(\widehat{\boldsymbol{\theta}})^\top \mathbf{J}(\widehat{\boldsymbol{\theta}})\}^{-1} \mathbf{J}(\widehat{\boldsymbol{\theta}})^\top.$$

Ou seja, $\mathbf{H}(\widehat{\boldsymbol{\theta}})$ é um projetor linear da pseudo resposta $\mathbf{z}(\widehat{\boldsymbol{\theta}})$ no plano explicado pelas colunas da matriz $\mathbf{J}(\widehat{\boldsymbol{\theta}})$, conhecido como plano tangente à superfície $\mathbf{f}(\boldsymbol{\theta})$ em $\widehat{\boldsymbol{\theta}}$. Os elementos da diagonal principal da matriz $\widehat{\mathbf{H}}$, $\widehat{h}_{11}, \ldots, \widehat{h}_{nn}$, podem ser considerados como medidas de ponto de alavanca. O resíduo padronizado

$$t_i = \frac{\{y_i - f(\widehat{\boldsymbol{\theta}}; \mathbf{x}_i)\}}{\widehat{\sigma}\sqrt{1 - \widehat{h}_{ii}}}$$

seria uma extensão natural do resíduo Studentizado da regressão linear para a regressão não linear, contudo esse resíduo não tem distribuição conhecida sendo necessário no gráfico normal de probabilidades a inclusão de bandas empíricas de confiança. Para detectar observações influentes, uma aproximação da distância de Cook para a regressão não linear é dada por $\mathrm{D}_i = t_i^2 \widehat{h}_{ii}/p(1 - \widehat{h}_{ii})$, para $i = 1, \ldots, n$.

1.18.8 Aplicação

Como ilustração considere o arquivo **lakemary** da biblioteca `alr4` do R, em que são descritos o comprimento em mm e a idade em anos de uma amostra de $n = 78$ peixes de uma espécie de água doce. Ajustar aos dados o modelo não linear de von Bertalanffy

$$y_i = \theta_1[1 - \exp\{-\theta_2(x_i - \theta_3)\}] + \epsilon_i,$$

em que y_i e x_i denotam, respectivamente, o comprimento (em mm) e a idade (anos) do i-ésimo peixe, enquanto $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, 78$.

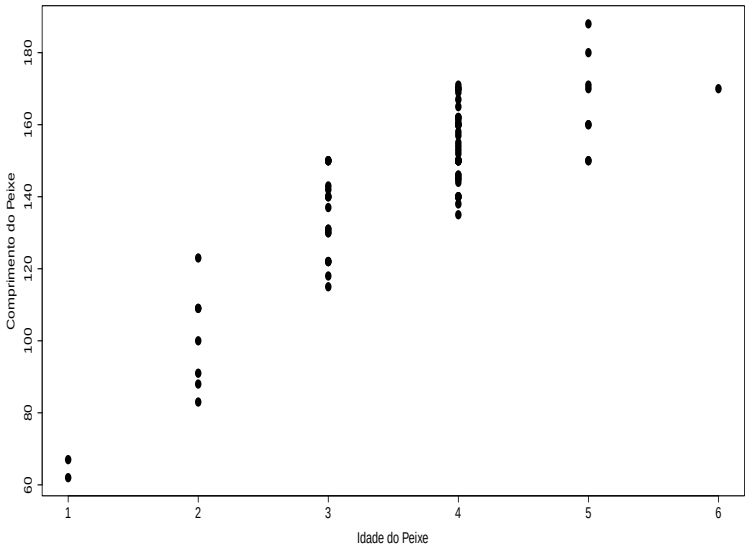

Figura 1.56: Diagrama de dispersão entre o comprimento e a idade do peixe do arquivo **lakemary** da biblioteca `alr4`.

Os dados estão descritos na Figura 1.56 e o ajuste da curva de von Bertalanffy é apresentado na Figura 1.57. Nota-se pelas estimativas que o parâmetro θ_3 não é significativo, sugerindo que os dados podem ser ajustados com um modelo mais simples envolvendo apenas os parâmetros θ_1 e θ_2. Para ilustrar, uma estimativa intervalar de 95% para o comprimento máximo esperado para a espécie fica (em mm) dada por $[191,809 \pm 1,96 \times 13,079]$. Análise de resíduos descrita na Figura 1.58 sugere adequação da suposição de normalidade e homocedasticidade dos erros. As variações dos resíduos dentro da banda de confiança é muito provavelmene devido ao fato de termos para uma mesma idade do peixe várias réplicas.

Tabela 1.27: Estimativas dos parâmetros referentes ao modelo de von Bertalanffy ajustado aos dados do arquivo **lakemary** da biblioteca `alr4`.

Parâmetro	Estimativa	Erro padrão	valor-z	valor-P
θ_1	191,809	13,079	14,74	0,000
θ_2	0,406	4,593	9,47	0,000
θ_3	0,081	0,240	0,34	0,737
s	10,960			

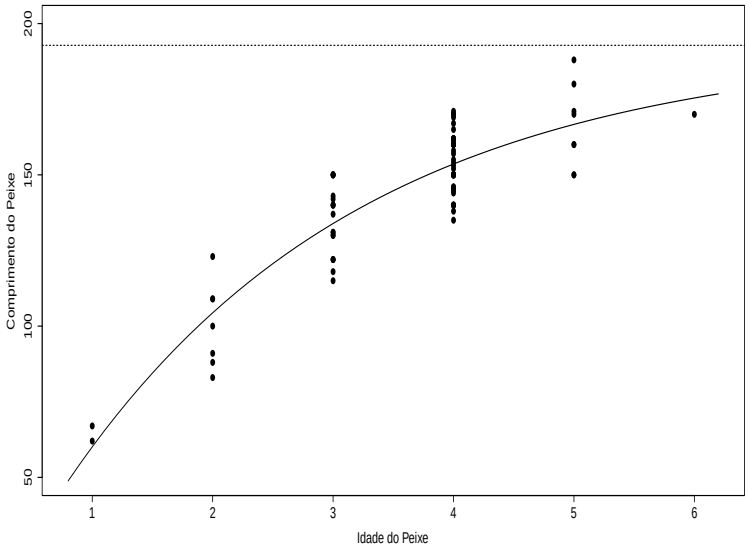

Figura 1.57: Curva ajustada de von Bertalanffy aos dados do arquivo **lakemary** da biblioteca `alr4`.

1.19 Erros Autoregressivos AR(1)

Em algumas situações práticas em que a regressão linear é aplicada pode haver suspeita de correlação temporal nas observações. Isso ocorre em particular quando as unidades experimentais são coletadas de forma temporal,

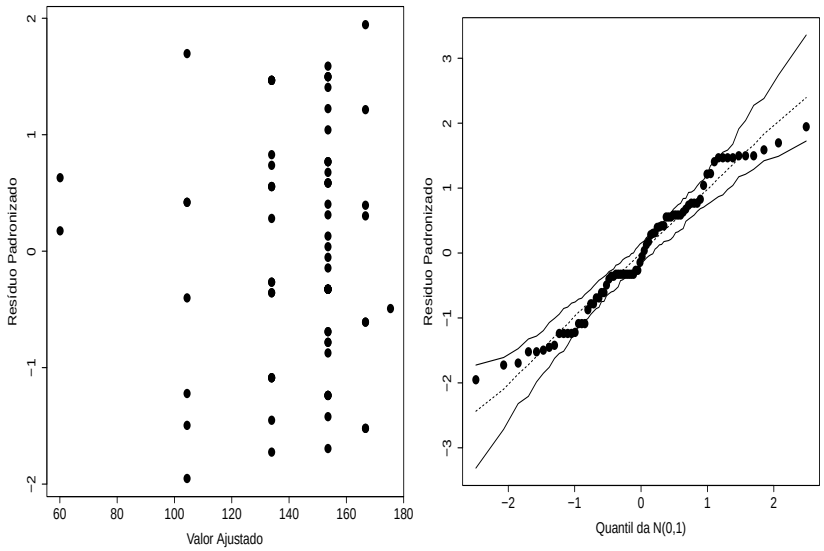

Figura 1.58: Análise de resíduos do ajuste do modelo de von Bertalanffy aos dados do arquivo **lakemary** da biblioteca `alr4`.

por exemplo, diariamente, semanalmente, mensalmente ou anualmente. O gráfico temporal do resíduo Studentizado pode revelar a necessidade de inclusão de alguma estrutura nos errros para acomodar a correlação temporal. Testes mais formais, como por exemplo o teste clássico de Durbin-Watson pode ser aplicado para avaliar se há correlação autoregressiva de ordem 1. Nesta seção será discutido o caso mais simples em que há suspeita de autocorrelação AR(1) nos erros. Embora procedimentos de máxima verossimilhança possam ser aplicados de uma forma geral, como ilustração de solução mais simples para o caso AR(1) será discutido o método de Cochrane-Orcutt, que procura reduzir o modelo linear normal com erros autoregressivos a um modelo com erros independentes e igualmente distribuídos.

Assim, considere o seguinte modelo de regressão linear:

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \epsilon_t, \tag{1.16}$$

em que $\epsilon_t = \phi\epsilon_{t-1} + e_t$ com $|\phi| < 1$ e $e_t \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $t = 1, \ldots, T$. Substituindo $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}$ na expressão $\epsilon_t = \phi\epsilon_{t-1} + e_t$ obtém-se

$$\begin{aligned} \epsilon_t &= \phi^2\{\phi\epsilon_{t-3} + e_{t-2}\} + \phi e_{t-1} + e_t \\ &= \phi^3\epsilon_{t-3} + \phi^2 e_{t-2} + \phi e_{t-1} + e_t. \end{aligned}$$

E assim sucessivamente segue que

$$\epsilon_t = \sum_{j=1}^{\infty} \phi^j e_{t-j}.$$

E dessa relação obtém-se os resultados

$$\text{E}(\epsilon_t) = 0, \quad \text{Var}(\epsilon_t) = \sigma^2\left(\frac{1}{1-\phi^2}\right) \quad \text{e} \quad \text{Cov}(\epsilon_t, \epsilon_{t\pm j}) = \phi^j\sigma^2\left(\frac{1}{1-\phi^2}\right),$$

para $j = 0, 1, 2, \ldots$. Logo, a autocorrelação entre os erros ϵ_t e ϵ_{t+1} fica dada por

$$\begin{aligned} \rho_t &= \frac{\text{Cov}(\epsilon_t, \epsilon_{t+1})}{\sqrt{\text{Var}(\epsilon_t)}\sqrt{\text{Var}(\epsilon_{t+1})}} \\ &= \frac{\phi\sigma^2\left(\frac{1}{1-\phi^2}\right)}{\sqrt{\sigma^2\left(\frac{1}{1-\phi^2}\right)}\sqrt{\sigma^2\left(\frac{1}{1-\phi^2}\right)}} \\ &= \phi. \end{aligned}$$

Pode-se mostrar de forma similar que a autocorrelação entre os erros ϵ_t e ϵ_{t+k} fica dada por $\rho_k = \phi^k$. Em particular, quando ϕ é positivo a magnitude da autocorrelação entre dois erros decresce à medida que a distância temporal entre os erros aumenta.

1.19.1 Teste de Durbin-Watson

Em geral, na prática, tem-se autocorrelação positiva entre os erros e um teste bastante conhecido para avaliar a necessidade de inclusão de uma estrutura de correlação AR(1) é o teste de Durbin-Watson (DW). Mais especificamente o teste de DW considera as hipóteses $\mathrm{H}_0 : \phi = 0$ contra $\mathrm{H}_1 : \phi > 0$, sendo a estatística do teste definida por

$$d = \frac{\sum_{t=2}^{T}(r_t - r_{t-1})^2}{\sum_{t=1}^{T} r_t^2},$$

em que $r_t = y_t - \widehat{y}_t$ é o resíduo ordinário da regressão de mínimos quadrados com erros independentes e igualmente distribuídos. Há tabelas disponíveis para avaliar o teste de DW que levam em conta o tamanho amostral, o nível de significância do teste e o número de variáveis explicativas no modelo (vide, por exemplo, Tabela A.6 de Montgomery et al.(2021)). Nessas tabelas são apresentados valores críticos d_U e d_L para a estatística do teste com o seguinte critério de decisão:

- Se $d < d_L$ rejeitar H_0
- Se $d > d_U$ não tejeitar H_0
- Se $d_L \leq d \leq d_U$ inconclusivo.

Há também bibliotecas que calculam diretamente o teste de DW com o respectivo valor-P, como por exemplo a biblioteca `lmtest` do R.

1.19.2 Método de Cochrane-Orcutt

Do modelo de regressão linear com erros AR(1) descrito em (1.16) segue que

$$y_{t-1} = \mathbf{x}_{t-1}^{\top}\boldsymbol{\beta} + \epsilon_{t-1}.$$

Logo, obtém-se $\epsilon_{t-1} = y_{t-1} - \mathbf{x}_{t-1}^\top \boldsymbol{\beta}$ e portanto pode-se escrever

$$\begin{aligned} y_t &= \mathbf{x}_t^\top + \phi\epsilon_{t-1} + e_t \\ &= \mathbf{x}_t^\top \boldsymbol{\beta} + \phi y_{t-1} - \phi \mathbf{x}_{t-1}^\top \boldsymbol{\beta} + e_t \\ y_t - \phi y_{t-1} &= \{x_t - \phi x_{t-1}\}^\top \boldsymbol{\beta} + e_t. \end{aligned}$$

Implicando para ϕ fixo na seguinte regressão linear:

$$u_t = \mathbf{z}_t^\top \boldsymbol{\beta} + e_t,$$

em que

$$u_t = y_t - \phi y_{t-1} \text{ e } \mathbf{z}_t = \mathbf{x}_t - \phi \mathbf{x}_{t-1}$$

com $e_t \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $t = 1, \ldots, T$. Portanto, para ϕ fixo, pode-se estimar $\boldsymbol{\beta}$ através do procedimento de mínimos quadrados

$$\widehat{\boldsymbol{\beta}} = (\mathbf{U}^\top \mathbf{U})^{-1} \mathbf{U}^\top \mathbf{z}, \tag{1.17}$$

em que $\mathbf{U}$ é uma matriz $n \times p$ de linhas $\mathbf{u}_t^\top$ e $\mathbf{z} = (z_1, \ldots, z_T)^\top$. Porém, na prática ϕ não é fixo, podendo ser estimado através de um estimador de momentos dado por $\widehat{\phi} = \sum_{t=2}^{T} r_t r_{t+1} / \sum_{t=1}^{T} r_t^2$ com $r_t = y_t - \widehat{y}_t$.

Assim, um procedimento iterativo para obter uma estimativa de mínimos quadrados para $\boldsymbol{\beta}$ fica dado por

1. Fornecer uma estimativa para ϕ.
2. Obter $\widehat{\boldsymbol{\beta}}$ de (1.17).
3. Aplicar o teste de DW.
4. Se H_0 não for for rejeitada, parar. Caso contrário, atualizar a estimativa para ϕ e repetir (1)-(3). Parar quando o teste for rejeitado e não

for mais possível mudar a estimativa de $\boldsymbol{\beta}$. Nesse último caso provavelmente uma estrutura de erros de ordem maior deve ser considerada em (1.16).

O processo iterativo acima pode ser aplicado através, por exemplo, da biblioteca `orcutt` do R.

1.20 Estimação por Máxima Verossimilhança

Como visto anteriormente o modelo de regressão linear múltipla assume que

- $Y_i|\mathbf{x}_i \overset{\text{ind}}{\sim} \text{N}(\mu_i, \sigma^2)$
- $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$,

em que $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ e $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, para $i = 1, \ldots, n$. Denotando $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$, em que $\phi = \sigma^2$, a função densidade de probabilidade de $Y_i|\mathbf{x}_i$ fica expressa na forma

$$f(y_i; \mathbf{x}_i, \boldsymbol{\theta}) = \left(\frac{1}{\sqrt{2\pi\phi}}\right) \exp\left\{-\frac{1}{2\phi}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2\right\},$$

para $i = 1, \ldots, n$. Assim, o logaritmo da função de verossimilhança fica dado por

$$\begin{aligned}
\text{L}(\boldsymbol{\theta}) &= \log[\Pi_{i=1}^n \{f(y_i; \mathbf{x}_i, \boldsymbol{\theta})\}] \\
&= n \log\left(\frac{1}{\sqrt{2\pi\phi}}\right) - \frac{1}{2\phi}\sum_{i=1}^n (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2 \\
&= -\frac{n}{2}\log(2\pi\phi) - \frac{1}{2\phi}\sum_{i=1}^n (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2.
\end{aligned}$$

Para obter as estimativas de máxima verossimilhança de $\boldsymbol{\beta}$ e ϕ é preciso derivar a função escore

$$\mathbf{U}_\theta = \frac{\partial \text{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \text{U}_\beta \\ \text{U}_\phi \end{pmatrix} = \begin{pmatrix} \frac{\partial \text{L}(\theta)}{\partial \beta} \\ \frac{\partial \text{L}(\theta)}{\partial \phi} \end{pmatrix}.$$

As estimativas de máxima verossimilhança são obtidas resolvendo-se as equações $\mathbf{U}_\beta = \mathbf{0}$ e $\mathrm{U}_\phi = 0$.

A derivada parcial de $\mathrm{L}(\boldsymbol{\theta})$ com relação a β_j fica dada por

$$\frac{\partial \mathrm{L}(\boldsymbol{\theta})}{\partial \beta_j} = \frac{1}{\phi}\sum_{i=1}^{n} x_{ij}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}),$$

para $j = 1, \ldots, p$. Em forma matricial obtém-se

$$\mathbf{U}_\beta = \frac{\partial \mathrm{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \frac{1}{\phi}\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

em que $\mathbf{y} = (y_1, \ldots, y_n)^\top$ e $\mathbf{X}$ é a matriz modelo. A estimativa de máxima verossimilhança $\widehat{\boldsymbol{\beta}}$ é obtida tal que

$$\mathbf{U}_\beta = \mathbf{0} \Rightarrow \mathbf{X}^\top(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{0}.$$

Então, se $\mathbf{X}$ é uma matriz de posto coluna completo tem-se solução única

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y},$$

que coincide com a estimativa de mínimos quadrados. Por outro lado, a derivada parcial de $\mathrm{L}(\boldsymbol{\theta})$ com relação a ϕ fica dada por

$$\mathrm{U}_\phi = \frac{\partial \mathrm{L}(\boldsymbol{\theta})}{\partial \phi} = -\frac{n}{2\phi} + \frac{1}{2\phi^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2,$$

e fazendo $\mathrm{U}_\phi = 0$ obtém-se

$$\widehat{\phi} = \widehat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n},$$

em que $\widehat{y}_i = \mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}$. Portanto, tem-se que $\widehat{\sigma}^2 = \frac{(n-p)}{n}s^2$ e $\mathrm{E}(\widehat{\sigma}^2) = \frac{(n-p)}{n}\sigma^2$. Logo, $\widehat{\sigma}^2$ é um estimador tendencioso de σ^2.

A matriz de informação de Fisher para $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$ é definada como sendo o valor esperado da curvatura de $\mathrm{L}(\boldsymbol{\theta})$

$$\mathbf{K}_{\theta\theta} = \mathrm{E}\left(-\frac{\partial^2 \mathrm{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right) = \begin{bmatrix} \mathbf{K}_{\beta\beta} & \mathbf{K}_{\beta\phi} \\ \mathbf{K}_{\phi\beta} & \mathrm{K}_{\phi\phi} \end{bmatrix},$$

em que $\mathbf{K}_{\beta\beta}$ e $\mathbf{K}_{\phi\beta}$ são submatrizes de informação de Fisher, respectivamente, de $\boldsymbol{\beta}$ e de $\boldsymbol{\beta}$ e ϕ simultaneamente, enquanto $\mathrm{K}_{\phi\phi}$ é a informação de Fisher de ϕ.

As submatrizes $\mathbf{K}_{\beta\beta}$ e $\mathbf{K}_{\phi\beta}$ ficam dadas por

$$\begin{aligned}
\mathbf{K}_{\beta\beta} &= \mathrm{E}\left(-\frac{\partial^2 \mathrm{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right) \\
&= \frac{1}{\phi}(\mathbf{X}^\top\mathbf{X}) \text{ e} \\
\mathbf{K}_{\beta\phi} &= \mathrm{E}\left(-\frac{\partial^2 \mathrm{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\phi}\right) \\
&= \frac{1}{\phi}\mathrm{E}\{\mathbf{X}^\top(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})|\mathbf{X}\} \\
&= \mathbf{X}^\top\mathrm{E}\{(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})|\mathbf{X}\} = \mathbf{0}.
\end{aligned}$$

Assim, os parâmetros $\boldsymbol{\beta}$ e ϕ são ortogonais. Ainda tem-se que

$$\begin{aligned}
\mathrm{K}_{\phi\phi} &= \mathrm{E}\left(-\frac{\partial^2 \mathrm{L}(\boldsymbol{\theta})}{\partial\phi^2}\right) \\
&= -\frac{n}{2\phi^2} + \frac{1}{\phi^3}\sum_{i=1}^{n}\mathrm{E}\{(Y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2\} \\
&= -\frac{n}{2\phi^2} + \frac{n}{\phi^2} = \frac{n}{2\phi^2}.
\end{aligned}$$

Logo, a matriz de informação de Fisher para $\boldsymbol{\theta}$ assume a forma bloco diagonal

$$\mathbf{K}_{\theta\theta} = \begin{bmatrix} \mathbf{K}_{\beta\beta} & \mathbf{0} \\ \mathbf{0} & \mathrm{K}_{\phi\phi} \end{bmatrix},$$

e pelas propriedades de estimação por máxima verosimilhança, tem-se para n grande que $\widehat{\boldsymbol{\beta}} \sim \mathrm{N}_p(\boldsymbol{\beta}, \mathbf{K}_{\beta\beta}^{-1})$ e $\widehat{\sigma}^2 \sim \mathrm{N}(\sigma^2, \mathrm{K}_{\phi\phi}^{-1})$. Além disso, $\widehat{\boldsymbol{\beta}}$ e $\widehat{\sigma}^2$ são independentes. No caso de $\widehat{\boldsymbol{\beta}}$ o resultado vale para todo n. Similarmente, segue que $(n-p)s^2/\sigma^2 \sim \chi^2_{(n-p)}$.

Exercícios

1. Seja T um estimador do parâmetro θ e supor a existência dos dois primeiros momentos de T. Mostre que

$$\mathrm{E}\{(\mathrm{T}-\theta)^2\} = \mathrm{E}[\{T-\mathrm{E}(T)\}^2] + \{\mathrm{E}(\mathrm{T})-\theta\}^2.$$

   Ou seja, EQM(T) = Var(T) + {Viés(T)}^{2}.

2. Com base numa amostra independente de $n = 3$ de uma variável aleatória X de média μ_X e variância σ_X^2 foram propostos para μ_X os seguintes estimadores:

$$\mathrm{T}_1 = \frac{1}{5}(X_1 + 3X_2 + X_3), \quad \mathrm{T}_2 = \frac{1}{2}(X_1 + 2X_3),$$

$$\mathrm{T}_3 = \frac{1}{4}(2X_1 + X_2 + X_3) \text{ e } \mathrm{T}_4 = \frac{1}{3}(X_1 + X_2 + X_3).$$

   Obtenha o erro quadrático médio, a variância e o viés de cada estimador. Entre os não tendenciosos qual escolher? Justifique.

3. Considere a seguinte regressão linear simples:

$$y_i = \beta_1 + \beta_2 x_i + \epsilon_i,$$

   em que $\epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, $i = 1, \ldots, n$. Mostre que: (i) $\mathrm{Cov}(\bar{Y}, \hat{\beta}_2) = 0$, (ii) $\sum_{i=1}^n r_i \hat{y}_i = 0$, (iii) $\sum_{i=1}^n r_i x_i = 0$, (iv) $\sum_{i=1}^n r_i = 0$ e (v) $\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i$, em que $r_i = y_i - \hat{y}_i$.

4. Supor que foi ajustado através de mínimos quadrados o modelo de regressão $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_2$, porém o modelo verdadeiro é dado por

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon,$$

   em que $\epsilon \sim \mathrm{N}(0, \sigma^2)$. Mostre que o estimador $\hat{\beta}_2$ obtido no primeiro ajuste é tendencioso. Expresse o viés de $\hat{\beta}_2$.

5. Considere inicialmente o modelosde regressão $y_i = \beta x_i + \epsilon_i$, em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, n$. Obter $\hat{\beta}$ e $Var(\hat{\beta})$. Supor agora que o modelo correto seja $y_i = \alpha + \beta x_i + \epsilon_i$. Mostre que $\hat{\beta}$ do primeiro modelo é um estimador tendencioso e que o erro quadrático médio de $\hat{\beta}$ pode ser expresso na forma

$$\text{E}QM(\hat{\beta}) = \frac{1}{\sum_{i=1}^{n} x_i^2} \left\{ \sigma^2 + \frac{(n\alpha\bar{x})^2}{\sum_{i=1}^{n} x_i^2} \right\}.$$

6. São apresentados na tabela abaixo o consumo (galão/milha)(Y) e a cilindrada (polegadas^3) (X) de uma amostra de $n = 32$ automóveis de marcas diferentes (Montgomery et al., 2021, Tabela B3).

y	x	y	x	y	x	y	x
18,90	350,0	17,00	350,0	20,00	250,0	18,25	351,0
20,07	225,0	11,20	440,0	22,12	231,0	21,47	262,0
34,70	89,7	30,40	96,9	16,50	350,0	36,50	85,3
21,50	171,0	19,70	258,0	20,30	140,0	17,80	302,0
14,39	500,0	14,89	440,0	17,80	350,0	16,41	318,0
23,54	231,0	21,47	360,0	16,59	400,0	31,90	96,9
29,40	140,0	13,27	460,0	23,90	133,6	19,73	318,0
13,90	351,0	13,27	351,0	13,77	360,0	16,50	350.0

Responda às seguintes questões: (i) construir o diagrama de dispersão entre o consumo e a cilindrada dos automóveis, comente; (ii) obter a correlação linear amostral de Pearson; (iii) ajustar o modelo de regressão linear simples de mínimos quadrados, obtendo as estimativas $\hat{\beta}_1$ e $\hat{\beta}_2$ e os respectivos erros padrão; (iv) traçar a reta de regressão no diagrama de dispersão; (v) interpretar a estimativa $\hat{\beta}_2$; (vi) obter as estimativas intervalares de 95% para β_1 e β_2 e (vii) obter a estimativa intervalar de 97% para o consumo de um automóvel com cilindrada de $x = 300$ polegadas^3. Resultados úteis: $\bar{y} = 20,2231$, $\bar{x} = 284,7312$,

$\sum y_i^2 = 14324,74$. $\sum x_i^2 = 3019001$ e $\sum x_i y_i = 164118,10$. Este exercício deve ser feito manualmente. O diagrama de dispersão pode ser feito no R.

7. Suponha o modelo de regressão dado em (1.4). Mostre que $\hat{\gamma} \sim \mathrm{N}(\gamma, \sigma^2/(1-h_{ii}))$. Mostre também que, sob a hipótese $\mathrm{H}_1 : \gamma \neq 0$, a estatística F tem uma distribuição $\mathrm{F}_{1,(n-p-1)}(\lambda)$, em que $\lambda = \frac{1}{2}\frac{\gamma^2(1-h_{ii})}{\sigma^2}$ é o parâmetro de não centralidade (Cook e Weisberg, 1982). Comente sobre o poder desse teste para $0 \leq h_{ii} < 1$. Use o resultado: se $\mathbf{Y} \sim \mathrm{N}_n(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ então $\mathbf{y}^\top\mathbf{y}/\sigma^2 \sim \chi^2_n(\lambda)$, em que $\lambda = \frac{1}{2}\boldsymbol{\mu}^\top\boldsymbol{\mu}/\sigma^2$.

8. Supor o modelo de regressão linear múltipla $y_i = \mathbf{x}_i^\top\boldsymbol{\beta} + \epsilon_i$, em que $\epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, para $i = 1, \ldots, n$. Mostre que uma estimativa intervalar de menor comprimento para σ^2 pode ser expressa na forma

$$\left[\frac{(n-p)s^2}{a}; \frac{(n-p)s^2}{b}\right],$$

em que a e b são constantes tais que $a^2 g_{(n-p)}(a) = b^2 g_{(n-p)}(b)$ com $g_{(n-p)}(t)$ denotando a função densidade de probabilidade da distribuição $\chi^2_{(n-p)}$. Sugestão: minimizar (derivando em b) o comprimento do intervalo $\ell(b) = (n-p)s^2[1/b - 1/a]$ e derivar em ambos os lados (em b) a equação $\int_a^b g_{(n-p)}(t)dt = (1-\alpha)$ com $a = a(b)$.

9. Considere agora o modelo de regressão linear múltipla $y_i = \mathbf{x}_i^\top\boldsymbol{\beta} + \epsilon_i$, em que $\epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, $\mathbf{x}_i = (1, x_{i2}, \ldots, x_{ip})^\top$ e $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, para $i = 1, \ldots, n$. Mostre que a estatística F para testar $\mathrm{H}_0 : \beta_2 = \cdots = \beta_p$ contra $\mathrm{H}_1 : \beta_j \neq 0$, para pelo menos algum $j = 2, \ldots, p$, pode ser expressa na forma

$$\mathrm{F} = \frac{R^2(n-p)}{(p-1)(1-R^2)}.$$

10. Considere um experimento planejado em que os efeitos da rapidez (X_1, em dois níveis, $X_1 = 0$ e $X_1 = 1$), da pressão (X_2, em dois níveis, $X_1 = -1$ e $X_1 = 1$) e da distância (X_3, em 3 níveis, $X_3 = -1$, $X_3 = 0$ e $X_3 = 1$) foram observados em uma máquina com relação à qualidade da impressão colorida (Y) em rótulos de embalagem. A fim de comparar os níveis de cada efeito foi considerado o seguinte modelo de regressão linear:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, $i = 1, \ldots, 12$. A matriz modelo $\mathbf{X}$ (desenho do experimento) e o vator de respostas $\mathbf{y}$ ficam, respectivamente, dados por:

$$\mathbf{X} = \begin{bmatrix} 0 & -1 & -1 \\ 1 & -1 & -1 \\ 0 & 1 & -1 \\ 1 & 1 & -1 \\ 0 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

e $\mathbf{y} = (y_1, \ldots, y_{12})^\top$. Responda às seguintes questões:

- Obter os estimadores de mínimos quadrados $\hat{\beta}_1$, $\hat{\beta}_2$ e $\hat{\beta}_3$, suas variâncias e covariâncias.
- Qual a propriedade que é obtida com esse experimento? Comente.

11. Suponha duas populações normais com médias μ_1 e μ_2, mesma variância, e que amostras independentes de tamanhos n_1 e n_2 foram, respectiva-

mente, obtidas das duas populações. Para o modelo com parte sistemática $\mu_1 = \alpha + \beta$ e $\mu_2 = \alpha - \beta$, mostre que a estatística F para testar $H_0 : \beta = 0$ contra $H_1 : \beta \neq 0$ pode ser expressa na forma simplificada

$$F = \frac{(n-2)\frac{n_1 n_2}{n}(\bar{y}_1 - \bar{y}_2)^2}{\sum(y_i - \bar{y})^2 - \frac{n_1 n_2}{n}(\bar{y}_1 - \bar{y}_2)^2},$$

em que $\bar{y}, \bar{y}_1, \bar{y}_2$ são as respectivas médias amostrais.

12. No arquivo **reg3.txt** são descritas as seguintes variáveis referentes a 50 estados norte-americanos: (i) `estado` (nome do estado), (ii) `pop` (população estimada em julho de 1975), (iii) `percap` (renda percapita em 1974 em USD), (iv) `analf` (proporção de analfabetos em 1970), (v) `expvida` (expectativa de vida em anos 1969-70), (vi) `crime` (taxa de criminalidade por 100000 habitantes 1976), (vii) `estud` (porcentagem de estudantes que concluem o segundo grau 1970), (viii) `ndias` (número de dias do ano com temperatura abaixo de zero grau Celsus na cidade mais importante do estado) e (ix) `area` (área do estado em milhas quadradas). Para ler o arquivo no R use o comando

    `reg3 = read.table(‘‘reg3.txt’’, header=TRUE).`

    O objetivo do estudo é tentar explicar a `expvida` média usando um modelo de regressão normal linear dadas as variáveis explicativas `percap`, `analf`, `crime`, `estud`, `ndias` e `dens`, em que `dens=pop/area`.

    Inicialmente faça uma análise descritiva dos dados, por exemplo com boxplot e a densidade da variável resposta e com diagramas de dispersão com as respectivas tendências entre a variável resposta e as variáveis explicativas. Comente essa parte descritiva. Posteriormente, ajuste o modelo de regressão normal linear com todas as variáveis explicativas e através do método `stepwise` (com PE=PS=0,15) faça uma

seleção de variáveis. Uma vez selecionado o modelo faça uma análise de diagnóstico e apresente as interpretações dos coeficientes estimados do modelo final.

13. No arquivo **octana.txt** (Wood, 1973) são descritos dados referentes à produção de gasolina numa determinada refinaria segundo três variáveis observadas durante o processo e uma quarta variável que é uma combinação das três primeiras. A resposta é o número de octanas do produto produzido. A octanagem é a propriedade que determina o limite máximo que a gasolina, junto com o ar, pode ser comprimida na câmara de combustão do veículo sem queimar antes de receber a centilha vinda das velas. As melhores gasolinas têm uma octanagem alta. Em grandes refinarias, o aumento de um octana na produção de gasolina pode representar um aumento de alguns milhões de dolares no custo final da produção. Assim, torna-se importante o controle dessa variável durante o processo de produção. Para ler o arquivo no R use os comandos

    ```
    octana = read.table(‘‘octana.txt’’, header=TRUE).
    ```

    Faça inicialmente uma análise descritiva dos dados. Por exemplo, boxplots robustos para cada variável e diagrama de dispersão de cada variável explicativa e a variável resposta octanas. Selecione um submodelo através dos métodos de maior R_k^2, menor s_k e menor C_k. Faça uma análise de resíduos com o modelo selecionado, bem como destaque as observações atípicas através dos gráficos de alavanca, distância de Cook e DFFITS. Avalie o impacto de cada observação destacada. Interprete os coeficientes do modelo estimado. Apenas de forma ilustrativa ajustar o modelo final no GAMLSS.

14. No arquivo **capm.txt** estão os seguintes dados (Ruppert, 2004, Cap.7):

Tbill (taxa de retorno livre de risco), retorno Microsoft, SP500 (retorno do mercado), retorno GE e retorno FORD de janeiro de 2002 a abril de 2003. Todos os retornos são diários e estão em porcentagem. Construir inicialmente os diagramas de dispersão (com tendência) entre o excesso de retorno ($y_{rt} - r_{ft}$) de cada uma das empresas Microsoft, GE e FORD e o excesso de retorno do mercado ($r_{mt} - r_{ft}$), em que y_{rt} denota o retorno da ação da empresa, r_{mt} é o retorno do mercado e r_{ft} indica a taxa livre de risco durante o t-ésimo período. Posteriormente, ajustar o seguinte modelo de regressão linear simples para cada ação:

$$y_t = \alpha + \beta x_t + \epsilon_t,$$

em que $y_t = y_{rt} - r_{ft}$, $x_t = r_{mt} - r_{ft}$ e $\epsilon_t \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$. No modelo acima, o parâmetro β é denominado risco sistemático com a seguinte interpretação: se $\beta = 1$ o excesso de retorno é equivalente ao mercado (volatilidade similar ao mercado), se $\beta > 1$ o excesso de retorno é maior do que o excesso de retorno do mercado (ação mais volátil do que o mercado), e se $\beta < 1$ o excesso de retorno é menor do que o excesso de retorno do mercado (ação menos volátil do que o mercado). O intercepto é incluído para controlar eventuais precificações incorretas, porém em geral $\alpha = 0$ não é rejeitado.

Para ler o arquivo no R use os comandos

```
capm = read.table("capm.txt", header=TRUE).
```

Para deixar o arquivo disponível use o comando

```
attach(capm).
```

Por exemplo, para ajustar o excesso de retorno da Microsoft use os comandos

```
ymsf = rmsf - tbill

xmerc = sp500 - tbill

ajuste.msf = lm(ymsf ~ xmerc)

summary(ajuste.msf).
```

Verifique se os modelos estão bem ajustados através de análise de resíduos. Para cada ação encontre uma estimativa intervalar de 95% para o risco sistemático e classifique o excesso de retorno em relação ao mercado. Finalmente, construa para cada ação a banda de confiança de 95% para prever o excesso de retorno num determinado dia, dado o excesso de retorno do mercado.

15. Suponha o modelo de comparação de médias

$$y_{ij} = \mu_i + \epsilon_{ij},$$

em que $\epsilon_{ij} \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, $i = 1, \ldots, k$ e $j = 1, \ldots, n_i$. Mostre que $\hat{\mu}_i = \bar{y}_i$ e $\text{Var}(r_{ij}) = \sigma^2(1 - 1/n_i)$, em que $r_{ij} = y_{ij} - \bar{y}_i$.

16. Considere o modelo de regressão linear múltipla

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, $i = 1, \ldots, n$. Mostre que o critério de Akaike é equivalente a minimizar a quantidade

$$\text{AIC} = n \log \left\{ \frac{\text{SQRes}}{n} \right\} + 2p,$$

com $\text{SQRes} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$.

17. Considere o arquivo **BigMac2003** da biblioteca `alr4` do R, em que são descritas as seguintes variáveis de 69 cidades de diversos países:

- `BigMac`: minutos de trabalho para comprar um Big Mac
- `Bread`: minutos de trabalho para comprar 1kg de pão
- `Rice`: minutos de trabalho para comprar 1kg de arroz
- `FoodIndex`: índice de preços de alimentos
- `Bus`: valor da passagem de ônibus (em USD)
- `Apt`: valor do aluguel (em USD) de um apartamento padrão de 3 dormitórios
- `TeachGI`: salário bruto anual (em 1000 USD) de um professor de ensino fundamental
- `TeachNI`: salário líquido anual (em 1000 USD) de um professor de ensino fundamental
- `TaxRate`: imposto pago (em porcentagem) por um professor de ensino fundamental
- `TeachHours`: carga horária semanal (em horas) de um professor de ensino fundamental.

Para disponibilizar e visualizar um resumo dos dados use na sequência os seguintes comandos do R:

```
require(alr4)

require(MASS)

attach(BigMac2003)

summary(BigMac2003).
```

O objetivo principal do estudo é relacionar a variável BigMac com as demais variáveis explicativas. A fim de obter uma melhor aproximação para a normalidade considere log(BigMac) como variável resposta. Apresente os diagramas de dispersão (com tendência) entre

a variável resposta e cada uma das variáveis explicativas e comente. Padronize as variáveis explicativas. Por exemplo, para padronizar a variável explicativa Bread use o comando

`sBread = scale(Bread, center = TRUE, scale = TRUE).`

Através do procedimento `stepAIC` fazer uma seleção das variáveis explicativas. Para o modelo selecionado aplicar análises de resíduos e de sensibilidade. Comente. Classifique as variáveis explicativas segundo o impacto na explicação da média da variável resposta.

18. No arquivo **motorins** da biblioteca `faraway` do R são descritas informações relacionadas a 1797 grupos de apólices de seguro de automóvel no ano de 1977 na Suécia. Em particular, há interesse em saber se há diferenças significativas entre o seguro médio pago por sinistro em 7 regiões do país. Para ler o arquivo no R utilize os comandos

    `require(faraway)`

    `summary(motorins)`

    `attach(motorins).`

    Considere as variáveis `Zone` (região do país) e perd valor pago por sinistro (em coroas suecas). A fim de obter uma melhor aproximação para a normalidade considere como resposta a variável log(perd). Construir boxplots de log(perd) segundo a região. Comente. Aplique em seguida um ajuste de comparação de médias através do comando

    `fit1.motor = lm(log(perd) ~ Zone).`

    Construa a tabela ANOVA através do comando

    `fit2.motor = aov(log(perd) ~ Zone).`

Se for rejeitada a hipótese de homogeneidade de médias, aplique o método de Tukey para verificar quais contrastes são significativos através do comando

`TukeyHSD(fit2.motor)`

`plot(TukeyHSD(fit2.motor),las=2).`

Comente.

19. No arquivo **fuel2001** da biblioteca `alr4` do R, estão descritas as seguintes variáveis referentes aos 50 estados norte-americanos mais o Distrito de Columbia no ano de 2001:

    - `UF`: unidade da federação
    - `Drivers`: número de motoristas licenciados
    - `FuelC`: total de gasolina vendida (em mil galões)
    - `Income`, renda per capita em 2000 (em mil USD)
    - `Miles`, total de milhas em estradas federais
    - `MPC`, milhas per capita percorridas
    - `Pop`, população ≥ 16 anos
    - `Tax`, taxa da gasolina (em cents por galão).

    A fim de possibilitar uma comparação entre as UFs duas novas variáveis são consideradas `Fuel = 1000*FuelC/Pop` e `Dlic = 1000*Drivers/Pop`, além da variável `Miles` ser substituída por `log(Miles)`. Para ler o arquivo no R use os comandos

    `require(alr4)`

    `require(MASS)`

```
attach(fuel2001)

summary(fuel2001).
```

Considere como resposta a variável Fuel e como variáveis explicativas Dlic, log(Miles), Income e Tax. Faça inicialmente uma análise descritiva dos dados. Por exemplo, boxplot robusto para a variável resposta e diagramas de dispersão (com tendência) entre cada variável explicativa e a variável resposta. Comente. Aplique o procedimento `stepAIC` para selecionar as variáveis explicativas. Verifique se é possível incluir alguma interação. Com o modelo selecionado faça uma análise de diagnóstico: análise de resíduos, pontos de alavanca, distância de Cook e DFFITS. Avalie o impacto dos pontos destacados. Interprete os coeficientes estimados.

20. No arquivo **wine.txt** (Montgomery et al., 2021, Tabela B.11) são descritas características de uma amostra aleatória de 38 vinhos da marca "Pinot Noir". O objetivo do estudo é relacionar a qualidade do vinho com as seguintes variáveis explicativas: (i) `claridade`, (ii) `aroma`, (iii) `corpo`, (iv) `sabor`, (v) `aromac`, aroma do tonel de carvalho e (vi) `regiao` (1: região 1, 2: região 2 e 3: região 3). Para ler o arquivo no R use os comandos

    ```
    wine = read.table("wine.txt", header=TRUE).
    ```

    A variável região é categórica com três níveis. Assim é possível através do comando `factor` do R transformá-la em duas variáveis binárias: `regiao2` = 1 para região 2 e 0 caso contrário e `regiao3` = 1 para região 3 e 0 em caso contrário. A casela de referência será a região 1. Para acionar o procedimento use o comando

    ```
    regiao = factor(regiao).
    ```

Faça inicialmente uma análise descritiva dos dados com boxplot robusto para a variável resposta e diagramas de dispersão (com tendência) entre a variável resposta e variáveis explicativas. Calcule também as correlações lineares de Peasorn entre as variáveis (exceto região). Selecione incialmente um submodelo através dos métodos de maior R_k^2, menor s_k, menor C_k e menor $\overline{\text{Press}}_k$. Em seguida selecione outro submodelo através do procedimento `stepwise` usando PE=PS=0,15. Compare os submodelos escolhidos e para o submodelo selecionado aplicar análise de resíduos e sensibilidade. Interpretar os coeficientes estimados.

21. Considere o modelo linear simples

$$y_i = \beta_1 + \beta_2(x_i - \bar{x}) + \epsilon_i,$$

para $i = 1, \ldots, n$ com $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$. Encontrar os estimadores *ridge* $\hat{\beta}_{R1}$ e $\hat{\beta}_{R2}$ como também suas variâncias e covariância assintóticas $\text{Var}(\hat{\beta}_{R1})$, $\text{Var}(\hat{\beta}_{R2})$ e $\text{Cov}(\hat{\beta}_{R1}, \hat{\beta}_{R2})$. Expresse os estimadores *ridge* em função dos estimadores de mínimos quadrados e mostre que são estimadores tendenciosos.

22. Para avaliar a relação entre a energia necessária diária e a produção de carne, uma amostra aleatória de 64 ovelhas em fase de crescimento foi considerada, sendo observado para cada animal o consumo médio diário de energia (mcal) e o peso (em kg). Esses dados estão descritos no arquivo **sheep.txt** (vide Lindsey, 1997, Seção 9.4). Para ler o arquivo no R use os comandos

```
sheep = read.table("sheep.txt", header=TRUE).
```

Fazer inicialmente uma análise descritivia dos dados, boxplot robusto da variável resposta (peso) e diagrama de dispersão entre o peso do

animal e o consumo diário de energia (variável explicativa). Ajustar um modelo linear normal aos dados e verificar que há indícios de variância não constante dos erros. Ajustar um modelo normal ponderado com pesos apropriados. Fazer uma análise de diagnóstico e interpretar as estimativas.

23. Considere o modelo de regressão linear múltipla

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, $i = 1, \ldots, n$. Mostre que SQRes(k) $\geq$ SQRes, em que SQRes(k) $= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R)^\top(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R)$ e SQRes $= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ denotam, respectivamente, as somas de quadrados de resíduos da regressão *ridge* e da regressão de mínimos quadrados.

24. Supor o modelo linear ponderado $y_i = \alpha + \beta x_i + \epsilon_i$, em que $\epsilon_i \overset{\text{ind}}{\sim} \text{N}(0, \text{d}_i\sigma^2)$, $\text{d}_i > 0$, para $i = 1, \ldots, n$. Obter $\hat{\beta}$ e ASQ($\beta = 0$).

25. Considere os dados do arquivo **Rateprof** da biblioteca `alr4` do R, referentes a notas médias recebidas por 364 instrutores de uma universidade norte americana durante um período de 10 anos. O objetivo do estudo é relacionar o interesse do avaliador (`RaterInterest`) (escore de 1 a 5) com as seguintes avaliações feitas pelo avaliador:

    - `Quality`: qualidade das aulas do instrutor (escore de 1 a 5)
    - `Helpfulness`: prestatividade do instrutor (escore de 1 a 5)
    - `Clarity`: clareza das aulas do instrutor (escore de 1 a 5)
    - `Easiness`: facilidade que o instrutor tem com a matéria (escore de 1 a 5).

Inicialmente centralize as 5 variáveis através do comando

`cvariavel = variavel - mean(variavel).`

Fazer uma análise descritiva com os dados apresentando a matriz de correlações lineares de Pearson e os diagramas de dispersão (com tendência). Comente. Ajustar agora um modelo de regressão linear da variável resposta centralizada contra as demais variáveis explicativas centralizadas e passando pela origem. Use o comando

`fit1 = lm(cresposta ~ cv1 + cv2 + cv3 + cv4 -1).`

Verifique se há indícios de multicolineridade através do VIF. Tente contornar o problema através de componentes principais, considerando apenas o 1º componente. Qual a explicação desse componente? Expresse esse componente em função das 4 variáveis explicativas centralizadas. Fazer um ajuste da regressão linear da variável resposta centralizada contra esse componente e passando pela origem. Interprete o coeficiente estimado e apresente análises de diagnóstico.

26. Considere o arquivo **oldfaith** da biblioteca `alr4` do R, em que a duração (em segundos) da erupção da fonte termal "Old Faithful Geyser"no lançamento de água e o intervalo (em minutos) até a próxima erupção foram observados $n = 270$ vezes em 1980. O principal objetivo do estudo é fazer predições para o intervalo até a próxima erupção dado o tempo que durou a erupção anterior. Para disponibilizar e visualizar um resumo dos dados use na sequência os seguintes comandos do R:

`require(alr4)`

`attach(oldfaith)`

`summary(oldfaith).`

Faça a transformação `nDuration = Duration/100` e apresente os gráficos de densidade e boxplot para a variável resposta bem como o diagrama de dispersão (com tendência usando $df = 5$) entre `Interval` e `nDuration`. Tente identificar 1 ponto de mudança (por exemplo nDuration=2.2) e proponha uma regressão por partes. Faça uma análise de diagnóstico e apresente a banda de confiança de 95% para prever o intervalo até a próxima erupção dado o tempo de duração da última erupção.

27. Considere o arquivo **mcycle** da biblioteca `MASS` do R, em que um conjunto de medidas da aceleração da cabeça (em g, 9,80665 m/s^2) ao longo do tempo (em milissegundos) em um acidente simulado de motocicleta é utilizado para avaliar a resistência de capacetes. Um dos objetivos do estudo é estimar a relação funcional entre a aceleração média da cabeça em função do tempo. Isso possibilitaria, por exemplo, novos estudos de simulação. Para disponibilizar e visualizar um resumo dos dados use na sequência os seguintes comandos do R:

    ```
    require(MASS)

    attach(mcycle)

    summary(mcycle).
    ```

    Construir inicialmente o diagrama de dispersão entre a resposta (aceleação da cabeça) e o tempo. Por exemplo usando o comando

    ```
    plot(times, accel, xlab="Tempo", ylab="Aceleração da Cabeça",
    pch=16)
    ```

    Comente. Considerar quatro pontos de mudança e ajustar uma regressõa por partes com erros normais. Sugestão: $t_1^0 = 14, t_2^0 = 21, t_3^0 = 31, t_4^0 = 42$. Avaliar o ajuste no R com os gráficos de resíduos contra o

valor ajustado e o gráfico normal de probabilidades. Repetir as análises no GAMLSS. Comparar o ajuste da regressão por partes com os ajustes através de spline cúbico e P-splines pelo GAMLSS. Apresentar as três curvas num mesmo gráfico.

28. No arquivo **ginidh.txt** constam o índice de GINI de 2013 e o IDH de 2017 dos 26 estados brasileiros mais o distrito federal. Construir inicialmente o diagrama de dispersão entre GINI(X) e IDH(Y) e comente. Ajustar através de uma regressão linear simples o IDH contra o índice de GINI. Aplicar procedimentos de diagnóstico, análise de resíduos e distância de Cook e comente. Elimine a UF discrepante e reajuste o modelo. Tente agora acomodar a UF discrepante através do seguinte modelo:

$$y_i = \beta_1 + \beta_2 x_i + \gamma z_i + \epsilon_i,$$

em que z_i é uma variável explicativa com zeros e valor 1 na posição da UF discrepante. Refazer a análise de resíduos e a distância de Cook para esse modelo e comente. Finalmente, aplicar para o modelo inicial o procedimento de Huber para tentar acomodar a UF discrepante. Compare os 4 ajustes e comente.

29. Na tabela abaixo tem-se a quantidade de água (em mm) na raiz e o comprimento (em cm) de 15 tipos de feijoeiros. Propor valores inicias e ajustar um modelo de crescimento logístico aos dados. Obter as estimativas intervalares para os parâmetros e construir os gráficos de resíduos. Comente. Qual a quantidade de água necessária para o feijoeiro alcançar metade do comprimento?

Comprimento	1,3	1,3	1,9	3,4	5,3	7,1	10,6	16,0
Água	0,5	1,5	2,5	3,5	4,5	5,5	6,5	7,5
Comprimento	16,4	18,3	20,9	20,5	21,3	21,2	20,9	
Água	8,5	9,5	10,5	11,5	12,5	13,5	14,5	

30. A tabela abaixo descreve a evolução da população brasileira (em milhões) através dos censos realizados desde 1872. Apresentar inicialmente o gráfico de dispersão da evolução da população brasileira. Propor valores inicias e ajustar um modelo de crescimento logístico, apresentar as análises de resíduos e comentar. Encontre uma estimativa intervalar aproximada de 90% para o valor esperado do máximo a ser alcançado pela população brasileira. Estime o valor esperado para a população brasileria em 2030, apresentando uma estiamtiva intervalar aproximada de 95%.

Censo	População	Censo	População	Censo	População
1872	9.930478	1890	14.333915	1900	17.438434
1920	30.635605	1940	41.236315	1950	51.944397
1960	70.191370	1970	93.139037	1980	119.002706
1991	146.825475	2000	169.779170	2010	190.755799
2022	210.862.983				

Capítulo 2

Modelos Lineares Generalizados

2.1 Introdução

Durante muitos anos os modelos normais lineares foram utilizados na tentativa de descrever a maioria dos fenômenos aleatórios. Mesmo quando o fenômeno sob estudo não apresentava uma resposta para a qual fosse razoável a suposição de normalidade, algum tipo de transformação era sugerida a fim de alcançar a normalidade procurada. Provavelmente a transformação mais conhecida foi proposta por Box e Cox (1964), a qual transforma o valor observado y (positivo) em

$$z = \begin{cases} \frac{y^{\lambda}-1}{\lambda} & \text{se} \quad \lambda \neq 0 \\ \log y & \text{se} \quad \lambda = 0, \end{cases}$$

sendo λ uma constante desconhecida. O objetivo da transformação de Box e Cox, quando aplicada a um conjunto de valores observados, é produzir aproximadamente a normalidade, a constância de variância e também a linearidade $\mathrm{E}(Z) = \eta$, em que $\eta = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$. No entanto, isso raramente ocorre para um único valor de λ (Box e Draper, 1987).

Com o desenvolvimento computacional ocorrido na década de 70, alguns

modelos que exigiam a utilização de processos iterativos para a estimação dos parâmetros começaram a ser mais aplicados, como por exemplo o modelo normal não linear. Todavia, a proposta mais interessante e pode-se dizer inovadora no assunto foi apresentada por Nelder e Wedderburn (1972), que propuseram os modelos lineares generalizados (MLGs). A ideia básica consiste em abrir o leque de opções para a distribuição da variável resposta, permitindo que a mesma pertença á família exponencial linear de distribuições, bem como dar maior flexibilidade para a relação funcional entre a média da variável resposta e o preditor linear η. Assim, por exemplo, para dados de contagem, em vez de aplicar a transformação $\sqrt{y}$ no sentido de buscar a normalidade dos dados e constância de variância, pode-se supor que a distribuição de Y é Poisson e que a relação funcional entre a média de Y e o preditor linear é dada por $\log(\mu) = \eta$. Essa relação funcional é conveniente, uma vez que garante para quaisquer valores dos parâmetros do preditor linear um valor positivo para μ. Similarmente, para proporções, pode-se pensar na distribuição binomial para a resposta e numa relação funcional do tipo $\log\{\mu/(1-\mu)\}$, em que $0 < \mu < 1$ denota a proporção esperada de sucessos.

Nelder e Wedderburn propuseram também um processo iterativo para a estimação dos parâmetros e introduziram o conceito de desvio que tem sido largamente utilizado na avaliação da qualidade do ajuste dos MLGs, bem como no desenvolvimento de resíduos e medidas de diagnóstico. Inúmeros trabalhos relacionados com modelos lineares generalizados foram publicados desde 1972 bem como a implementação dos MLGs em alguns *softwares*. Neste texto as saídas e gráficas foram desenvolvidos no software `R` (`http://CRAN.R-project.org`).

Os modelos de quase-verossimilhança, que estendem a ideia dos MLGs para situações mais gerais incluindo dados correlacionados, foram propos-

tos por Wedderburn (1974). Os modelos de dispersão (Jørgensen, 1987) ampliam o leque de opções para a distribuição da variável resposta. Liang e Zeger (1986) estendem os modelos de quase-verossimilhança propondo as equações de estimação generalizadas (EEGs) que permitem o estudo de variáveis aleatórias correlacionadas não gaussianas. Os modelos não lineares de família exponencial (Cordeiro e Paula, 1989 e Wei, 1998) admitem preditor não linear nos parâmetros. Tem-se ainda os modelos aditivos generalizados (Hastie e Tibshirani, 1990; Green e Silverman, 1994; Wood, 2017) que supõem preditor linear formado também por funções aditivas e parciais aditivas e os modelos lineares generalizados mistos (Breslow e Clayton, 1993 e McCulloch e Searle, 2001) que admitem a inclusão de efeitos aleatórios gaussianos no preditor linear. Mais recentemente, Lee e Nelder (1996, 2001) estenderam o trabalho de Breslow e Clayton propondo modelos lineares generalizados hierárquicos em que o preditor linear pode ser formado por efeitos fixos e efeitos aleatórios não gaussianos. Muitos desses resultados são discutidos no livro de Lee et al. (2006). Extensões de MLGs para séries temporais, análise de dados de sobrevivência, modelos de espaço de estado e outros modelos multivariados são descritas, por exemplo, em Fahrmeir e Tutz (2001). Os modelos aditivos generalizados de localização, escala e forma propostos por Rigby e Stasinopoulos (2005) (vide também Stasinopoulos et al., 2017) contemplam as diversas extensões dos MLGs possibilitando a modelagem conjunta de todos os parâmetrso das distribuições, inclusive os quartis. Referências de texto no assunto são os livros de McCullagh e Nelder (1989) e Cordeiro et al. (2024).

Neste capítulo os modelos lineares generalizados são introduzidos juntamente com vários resultados relacionados com estimação, teste de hipóteses e métodos de diagnóstico. Algumas aplicações são apresentadas no final do

capítulo envolvendo as principais distribuições da família exponencial, como também vários exercícios teóricos e práticos.

2.2 Definição

Sejam $Y_1, \ldots, Y_n$ variáveis aleatórias independentes, cada uma com função densidade de probabilidade ou função de probabilidade na forma dada abaixo

$$f(y_i; \theta_i, \phi) = \exp[\phi\{y_i\theta_i - b(\theta_i)\} + c(y_i, \phi)], \tag{2.1}$$

denominada família exponencial linear. Pode-se mostrar sob condições usuais de regularidade que

$$\begin{aligned} \mathrm{E}\left\{\frac{\partial \log f(Y_i; \theta_i, \phi)}{\partial \theta_i}\right\} &= 0 \quad \text{e} \\ \mathrm{E}\left[\frac{\partial^2 \log f(Y_i; \theta_i, \phi)}{\partial \theta_i^2}\right] &= -\mathrm{E}\left[\left\{\frac{\partial \log f(Y_i; \theta_i, \phi)}{\partial \theta_i}\right\}^2\right], \end{aligned}$$

em que $\mathrm{E}(Y_i) = \mu_i = b'(\theta_i)$ e $\mathrm{Var}(Y_i) = \phi^{-1}V(\mu_i)$, sendo $V_i = V(\mu_i) = d\mu_i/d\theta_i$ é denominada função de variância e $\phi^{-1} > 0$ ($\phi > 0$) é o parâmetro de dispersão (precisão), $i = 1, \ldots, n$. A função de variância desempenha um papel importante na família exponencial, uma vez que a mesma caracteriza a distribuição. Isto é, dada a função de variância, tem-se uma classe de distribuições correspondentes, e vice-versa. Para ilustrar, a função de variância definida por $V(\mu) = \mu(1-\mu)$, $0 < \mu < 1$, caracteriza a classe de distribuições binomiais com probabilidades de sucesso μ e $1 - \mu$. Uma propriedade interessante envolvendo a distribuição de Y e a função de variância é a seguinte:

$$\sqrt{\phi}(Y - \mu) \ \to_d \ \mathrm{N}(0, V(\mu)), \quad \text{quando} \quad \phi \to \infty.$$

Ou seja, para ϕ grande Y segue distribuição aproximadamente normal de média μ e variância $\phi^{-1}V(\mu)$. Esse tipo de abordagem assintótica, diferente da usual em que n é grande, foi introduzida por Jørgensen (1987).

Os modelos lineares generalizados são definidos por (2.1) e pela parte sistemática

$$g(\mu_i) = \eta_i, \tag{2.2}$$

em que $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ é o preditor linear, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, $p < n$, é um vetor de parâmetros desconhecidos a serem estimados, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ representa os valores de variáveis explicativas e $g(\cdot)$ é uma função monótona e diferenciável, denominada função de ligação. Apresenta-se a seguir as distribuições mais conhecidas pertencentes à família exponencial linear.

2.2.1 Casos particulares

Normal

Seja Y uma variável aleatória com distribuição normal de média μ e variância σ^2, $Y \sim \mathrm{N}(\mu, \sigma^2)$. A função densidade de probabilidade de Y é expressa na forma

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} = \exp\left[\left\{\frac{1}{\sigma^2}(\mu y - \frac{\mu^2}{2}) - \frac{1}{2}\{\log(2\pi\sigma^2) + \frac{y^2}{\sigma^2}\right\}\right],$$

em que $-\infty < \mu, y < \infty$ e $\sigma^2 > 0$. Logo, para $\theta = \mu$, $b(\theta) = \theta^2/2$, $\phi = \sigma^{-2}$ e $c(y;\phi) = \frac{1}{2}\log(\phi/2\pi) - \frac{\phi y^2}{2}$ obtém-se (2.1). Verifica-se facilmente que a função de variância é dada por $V(\mu) = 1$.

Poisson

No caso de $Y \sim \mathrm{P}(\mu)$, a função de probabilidade fica dada por

$$e^{-\mu}\mu^y/y! = \exp\{y\log(\mu) - \mu - \log(y!)\},$$

em que $\mu > 0$ e $y = 0, 1, \ldots$. Fazendo $\log(\mu) = \theta$, $b(\theta) = e^\theta$, $\phi = 1$ e $c(y;\phi) = -\log(y!)$ obtém-se (2.1). Segue portanto que $V(\mu) = \mu$.

Binomial

Seja Y^* a proporção de sucessos em n ensaios independentes, cada um com probabilidade de ocorrência μ. Denota-se $nY^* \sim \mathrm{B}(n,\mu)$. A função de probabilidade de Y^* fica então expressa na forma

$$\binom{n}{ny^*}\mu^{ny^*}(1-\mu)^{n-ny^*} = \exp\left\{\log\binom{n}{ny^*} + ny^*\log\left(\frac{\mu}{1-\mu}\right) + n\log(1-\mu)\right\},$$

em que $0 < \mu, y^* < 1$. Obtém-se (2.1) fazendo $\phi = n$, $\theta = \log\{\mu/(1-\mu)\}$, $b(\theta) = \log(1+e^\theta)$ e $c(y^*;\phi) = \log\binom{\phi}{\phi y^*}$. A função de variância fica dada por $V(\mu) = \mu(1-\mu)$.

Gama

Seja Y uma variável aleatória com distribuição gama de média μ e coeficiente de variação $\phi^{-\frac{1}{2}}$, denota-se $Y \sim \mathrm{G}(\mu,\phi)$. A função densidade de probabilidade de Y é dada por

$$\begin{aligned}\frac{1}{\Gamma(\phi)}\left(\frac{\phi y}{\mu}\right)^{\phi}\exp\left(-\frac{\phi y}{\mu}\right)\mathrm{d(logy)} &= \exp[\phi\{(-y/\mu) - \log(\mu)\} - \log(\Gamma(\phi)) + \\ &\quad \phi\log(\phi y) - \log(y)],\end{aligned}$$

em que $y > 0$, $\phi > 0$, $\mu > 0$ e $\Gamma(\phi) = \int_0^\infty t^{\phi-1}e^{-t}dt$ é a função gama. Logo, fazendo $\theta = -1/\mu$, $b(\theta) = -\log(-\theta)$ e $c(y;\phi) = (\phi-1)\log(y) + \phi\log(\phi) - \log(\Gamma(\phi))$ obtém-se (2.1).

Para $0 < \phi < 1$ a densidade da gama tem uma *pole* na origem e decresce monotonicamente quando $y \to \infty$. A exponencial é um caso especial quando $\phi = 1$. Para $\phi > 1$ a função densidade assume zero na origem, tem um máximo em $y = \mu - \mu/\phi$ e depois decresce para $y \to \infty$. A χ^2_k é um outro caso especial quando $\phi = k/2$ e $\mu = k$. A distribuição normal é obtida fazendo $\phi \to \infty$. Isto é, quando ϕ é grande $Y \sim \mathrm{N}(\mu, \phi^{-1}V(\mu))$. Tem-se que $\phi = \mathrm{E}^2(Y)/\mathrm{Var}(Y)$ é o inverso do coeficiente de variação de Y ao quadrado,

ou seja, $\phi = 1/(\mathrm{CV}(Y))^2$, em que $\mathrm{CV}(Y) = \sqrt{\mathrm{Var(Y)}}/\mathrm{E}(Y)$. A função de variância da gama é dada por $V(\mu) = \mu^2$.

Normal inversa

Seja Y uma variável aleatória com distribuição normal inversa de média μ e parâmetro de precisão ϕ, denotada por $Y \sim \mathrm{NI}(\mu, \phi)$ e cuja função densidade de probabilidade é dada por

$$\sqrt{\frac{\phi}{2\pi y^3}}\exp\left\{-\frac{\phi(y-\mu)^2}{2\mu^2 y}\right\} = \exp\left[\phi\left\{-\frac{y}{2\mu^2}+\frac{1}{\mu}\right\} - \frac{1}{2}\left\{\log(2\pi y^3/\phi) + \frac{\phi}{y}\right\}\right],$$

em que $y > 0$, $\mu > 0$. Fazendo $\theta = -\frac{1}{2\mu^2}$, $b(\theta) = -(-2\theta)^{1/2}$ e $c(y;\phi) = \frac{1}{2}\log\{\phi/(2\pi y^3)\} - \frac{\phi}{2y}$ obtém-se (2.1). A normal inversa se aproxima da normal quando $\phi \to \infty$. Ou seja, para ϕ grande tem-se que $Y \sim \mathrm{N}(\mu, \phi^{-1}V(\mu))$. A função de variância fica aqui dada por $V(\mu) = \mu^3$.

Na Tabela 2.1 é descrito um resumo dessas distribuições.

Tabela 2.1
Principais distribuições pertencentes à família exponencial linear.

Distribuição	$b(\theta)$	θ	ϕ	$V(\mu)$
Normal	$\theta^2/2$	μ	σ^{-2}	1
Poisson	e^θ	$\log(\mu)$	1	μ
Binomial	$\log(1+e^\theta)$	$\log\{\mu/(1-\mu)\}$	n	$\mu(1-\mu)$
Gama	$-\log(-\theta)$	$-1/\mu$	$1/(\mathrm{CV}(Y))^2$	μ^2
N.Inversa	$-\sqrt{-2\theta}$	$-1/2\mu^2$	ϕ	μ^3

2.3 Ligações canônicas

Supondo ϕ conhecido, o logaritmo da função de verossimilhança de um MLG com respostas independentes pode ser expresso na forma

$$\mathrm{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \phi\{y_i\theta_i - b(\theta_i)\} + \sum_{i=1}^{n} c(y_i, \phi).$$

Um caso particular importante ocorre quando o parâmetro canônico (θ) coincide com o preditor linear, isto é, quando $\theta_i = \eta_i = \sum_{j=1}^{p} x_{ij}\beta_j$. Nesse caso, $\mathrm{L}(\boldsymbol{\beta})$ fica dado por

$$\mathrm{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \phi\{y_i \sum_{j=1}^{p} x_{ij}\beta_j - b(\sum_{j=1}^{p} x_{ij}\beta_j)\} + \sum_{i=1}^{n} c(y_i, \phi).$$

Definindo a estatística $S_j = \phi \sum_{i=1}^{n} Y_i x_{ij}$, $\mathrm{L}(\boldsymbol{\beta})$ fica então reexpresso na forma

$$\mathrm{L}(\boldsymbol{\beta}) = \sum_{j=1}^{p} s_j\beta_j - \phi \sum_{i=1}^{n} b(\sum_{j=1}^{p} x_{ij}\beta_j) + \sum_{i=1}^{n} c(y_i, \phi).$$

Logo, pelo teorema da fatorização a estatística $\mathbf{S} = (S_1, \ldots, S_p)^\top$ é suficiente minimal para o vetor $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$. As ligações que correspondem a tais estatísticas são chamadas de ligações canônicas e desempenham um papel importante na teoria dos MLGs. As ligações canônicas mais comuns são dadas abaixo.

Distribuição	Normal	Binomial	Poisson	Gama	N. Inversa
Ligação	$\mu = \eta$	$\log\left\{\frac{\mu}{1-\mu}\right\} = \eta$	$\log(\mu) = \eta$	$\mu^{-1} = \eta$	$\mu^{-2} = \eta$

Uma das vantagens de usar ligações canônicas é que as mesmas garantem a concavidade de $\mathrm{L}(\boldsymbol{\beta})$ e consequentemente muitos resultados assintóticos são obtidos mais facilmente. Por exemplo, a concavidade de $\mathrm{L}(\boldsymbol{\beta})$ garante a unicidade da estimativa de máxima verossimilhança de $\boldsymbol{\beta}$, quando essa existe. Para ligações não canônicas Wedderburn (1976) discute condições para a existência da concavidade de $\mathrm{L}(\boldsymbol{\beta})$.

2.3.1 Outras ligações

Ligação probito

Seja μ a proporção de sucessos de uma distribuição binomial. A ligação probito (Finney, 1971) é definida por

$$\Phi^{-1}(\mu) = \eta,$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada da normal padrão.

Ligação complemento log-log

A distribuição do valor extremo (logaritmo da exponencial) tem função densidade de probabilidade dada por

$$f(y) = \exp\{y - \exp(y)\},$$

em que $-\infty < y < \infty$. Logo, a função de distribuição acumulada fica dada por

$$F(y) = 1 - \exp\{-\exp(y)\}.$$

O modelo binomial com ligação complemento log-log é definido tal que

$$\mu = 1 - \exp\{-\exp(\eta)\},$$

ou, equivalentemente,

$$\log\{-\log(1-\mu)\} = \eta.$$

A ligação logito é definida de forma similar. A função densidade de probabilidade da distribuição logística é dada por

$$f(y) = \frac{\exp(y)}{\{1 + \exp(\mathrm{y})\}^2},$$

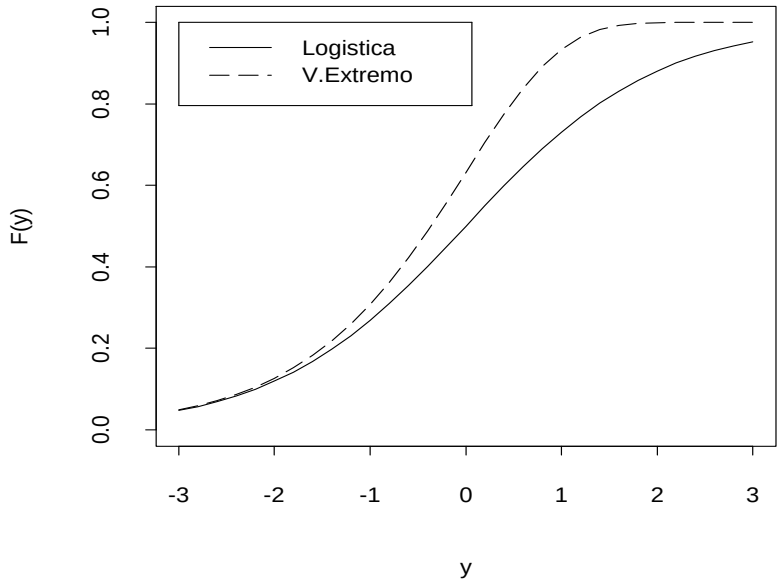

Figura 2.1: Função de distribuição acumulada das curvas logística e do valor extremo.

em que $-\infty < y < \infty$. Daí segue que a função de distribuição acumulada fica expressa na forma

$$F(y) = \frac{e^y}{(1+e^y)}.$$

O modelo logístico binomial é obtido substituindo $F(y)$ por μ e y por η na expressão acima. Como no caso binomial o parâmetro de interesse sempre é uma probabilidade, fica muito razoável que funções de distribuições acumuladas sejam utilizadas para gerarem novas ligações e consequentemente novos modelos. Na Figura 2.1 tem-se a $F(y)$ da distribuição logística e da distribuição do valor extremo para valores de y variando no intervalo $[-3\,,\,3]$. Note que a curva logística é simétrica em torno de $F(y) = 1/2$, enquanto que a curva do valor extremo apresenta comportamentos distintos para $F(y) \leq 1/2$ e $F(y) > 1/2$.

Ligação de Box-Cox

Uma classe importante de ligações, pelo menos para observações positivas, é a classe de ligações de Box-Cox definida por

$$\eta = (\mu^{\lambda} - 1)/\lambda,$$

para $\lambda \neq 0$ e $\eta = \log(\mu)$ para $\lambda \to 0$. A ideia agora é aplicar a transformação de Box-Cox, definida na Seção 2.1, na média da variável resposta ao invés de transfor a própria variável resposta. Tem-se na Figura 2.2 o comportamento de μ para alguns valores de λ e para η variando no intervalo [0 , 10].

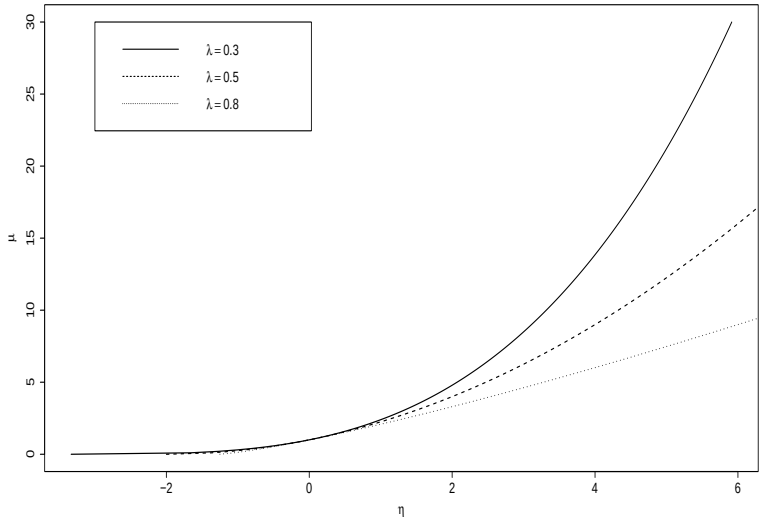

Figura 2.2: Ligação de Box-Cox para alguns valores de λ.

Ligação de Aranda-Ordaz

Uma outra transformação importante foi proposta por Aranda-Ordaz (1981) para dados binários. A transformação é dada por

$$\eta = \log\left\{\frac{(1-\mu)^{-\alpha} - 1}{\alpha}\right\},$$

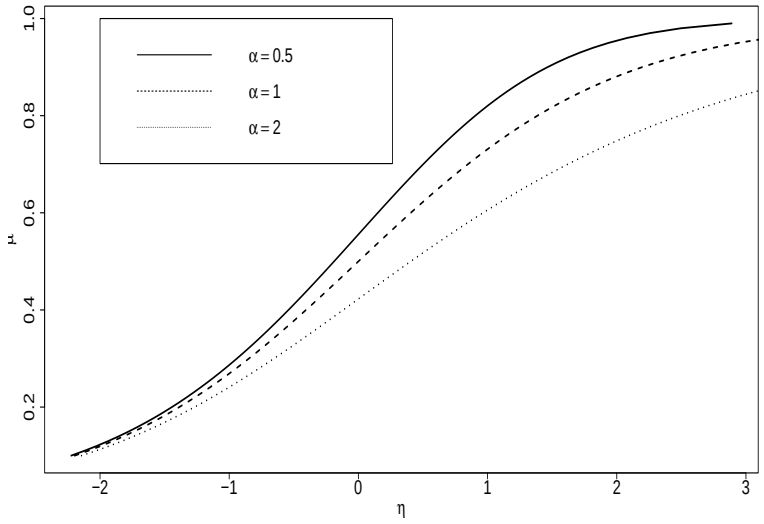

Figura 2.3: Ligação de Aranda-Ordaz para alguns valores de α.

em que $0 < \mu < 1$ e α é uma constante desconhecida. Quando $\alpha = 1$ tem-se a ligação logito $\eta = \log\{\mu/(1-\mu)\}$. Quando $\alpha \to 0$ tem-se $\{(1-\mu)^{-\alpha}-1\}/\alpha \to \log(1-\mu)^{-1}$ de modo que $\eta = \log\{-\log(1-\mu)\}$ e obtém-se portanto a ligação complemento log-log. Na Figura 2.3 tem-se o comportamento de μ para alguns valores de α. Em muitas situações práticas o interesse pode ser testar se o modelo logístico é apropriado, $H_0 : \alpha = 1$, contra a necessidade de uma transformação na ligação, $H_1 : \alpha \neq 1$.

Os MLGs são ajustados no aplicativo `R` através do comando `glm`. Para ilustrar uma aplicação, supor que o interesse é ajustar um modelo de Poisson com ligação canônica e que a variável resposta é denotada por `resp` com variáveis explicativas `cov1` e `cov2`. Pode-se mandar os resultados do ajuste para um arquivo (objeto no `R`), por exemplo com nome `fit.poisson`, através do comando

```
fit.poisson = glm( resp ~ cov1 + cov2, family=poisson).
```

Com o comando

```
summary(fit.poisson)
```

tem-se um resumo dos resultados do ajuste.

2.4 Função desvio

Sem perda de generalidade, supor que o logaritmo da função de verossimilhança seja agora definido por

$$\mathrm{L}(\boldsymbol{\mu};\mathbf{y}) = \sum_{i=1}^{n} \mathrm{L}(\mu_i; y_i),$$

em que $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$. Para o modelo saturado ($p = n$) a função $\mathrm{L}(\boldsymbol{\mu};\mathbf{y})$ é estimada por

$$\mathrm{L}(\mathbf{y};\mathbf{y}) = \sum_{i=1}^{n} \mathrm{L}(y_i; y_i).$$

Ou seja, a estimativa de máxima verossimilhança de μ_i fica nesse caso dada por $\tilde{\mu}_i = y_i$. Quando $p < n$, denota-se a estimativa de $\mathrm{L}(\boldsymbol{\mu};\mathbf{y})$ por $\mathrm{L}(\hat{\boldsymbol{\mu}};\mathbf{y})$. Aqui, a estimativa de máxima verossimilhança de μ_i será dada por $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$, em que $\hat{\eta}_i = \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}$.

A qualidade do ajuste de um MLG é avaliada através da função desvio

$$\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = \phi \mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}) = 2\{\mathrm{L}(\mathbf{y};\mathbf{y}) - \mathrm{L}(\hat{\boldsymbol{\mu}};\mathbf{y})\},$$

que é uma distância entre o logaritmo da função de verossimilhança do modelo saturado (com n parâmetros) e do modelo sob investigação (com p parâmetros) avaliado na estimativa de máxima verossimilhança $\hat{\boldsymbol{\beta}}$. Um valor pequeno para a função desvio indica que, para um número menor de parâmetros, tem-se um ajuste tão bom quanto o ajuste com o modelo saturado. Denotando por $\hat{\theta}_i = \theta_i(\hat{\mu}_i)$ e $\tilde{\theta}_i = \theta_i(\tilde{\mu}_i)$ as estimativas de máxima

verossimilhança de θ para os modelos com p parâmetros ($p < n$) e saturado ($p = n$), respectivamente, tem-se que a função $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})$ (não escalonada por ϕ) fica, alternativamente, dada por

$$\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\{y_i(\tilde{\theta}_i - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\tilde{\theta}_i))\}.$$

A seguir a função desvio é derivada para alguns casos particulares. O desvio no `R` sai com o nome *deviance* após o ajuste do modelo e o número de graus de liberdade correspondente é dado por $n - p$. É usual denotar $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} d^2(y_i; \hat{\mu}_i)$, em que $d^2(y_i; \hat{\mu}_i)$ será denominado componente do desvio não escalonado.

Normal

Aqui $\theta_i = \mu_i$, logo $\tilde{\theta}_i = y_i$ e $\hat{\theta}_i = \hat{\mu}_i$. O desvio fica portanto dado por

$$\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\{y_i(y_i - \hat{\mu}_i) + \hat{\mu}_i^2/2 - y_i^2/2\} = \sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2,$$

que coincide com a soma de quadrados de resíduos.

Poisson

Neste caso tem-se $\theta_i = \log(\mu_i)$, o que implica em $\tilde{\theta}_i = \log(y_i)$ para $y_i > 0$ e $\hat{\theta}_i = \log(\hat{\mu}_i)$. Assim,

$$\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\{y_i\log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)\}.$$

Se $y_i = 0$ o i-ésimo termo de $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})$ vale $2\hat{\mu}_i$. Resumindo, tem-se o seguinte resultado para o modelo de Poisson:

$$d^2(y_i; \hat{\mu}_i) = \begin{cases} 2\{y_i\log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)\} & \text{se } \ y_i > 0; \\ 2\hat{\mu}_i & \text{se } \ y_i = 0. \end{cases}$$

Binomial

No caso binomial em que $Y_i \sim \mathrm{B}(n_i, \mu_i)$, $i = 1, \ldots, k$, obtém-se $\tilde{\theta}_i = \log\{y_i/(n_i - y_i)\}$ e $\hat{\theta}_i = \log\{\hat{\mu}_i/(1-\hat{\mu}_i)\}$ para $0 < y_i < n_i$. Logo, o desvio assume a seguinte forma:

$$\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{k}[y_i \log(y_i/n_i\hat{\mu}_i) + (n_i - y_i)\log\{(1 - y_i/n_i)/(1-\hat{\mu}_i)\}].$$

Todavia, quando $y_i = 0$ ou $y_i = n_i$, o i-ésimo termo de $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})$ vale $-2n_i\log(1-\hat{\mu}_i)$ ou $-2n_i\log\hat{\mu}_i$, respectivamente. Portanto, os componentes do desvio no caso binomial assumem as seguintes formas:

$$d^2(y_i; \hat{\mu}_i) = \begin{cases} y_i\log(y_i/n_i\hat{\mu}_i) + (n_i - y_i)\log\{(1 - y_i/n_i)/(1-\hat{\mu}_i)\} & \text{se } 0 < y_i < n_i; \\ -2n_i\log(1-\hat{\mu}_i) & \text{se } y_i = 0; \\ -2n_i\log\hat{\mu}_i & \text{se } y_i = n_i. \end{cases}$$

Gama

No caso gama, $\tilde{\theta}_i = -1/y_i$ e $\hat{\theta}_i = -1/\hat{\mu}_i$. Assim, segue que o desvio (quando todos os valores são positivos) pode ser expresso na forma

$$\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\{-\log(y_i/\hat{\mu}_i) + (y_i - \hat{\mu}_i)/\hat{\mu}_i\}.$$

Se algum componente de y_i é igual a zero o desvio fica indeterminado. McCullagh e Nelder (1989) sugerem substituir $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})$ nesse caso por

$$\mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\phi C(\mathbf{y}) + 2\phi\sum_{i=1}^{n}\log(\hat{\mu}_i) + 2\phi\sum_{i=1}^{n}(y_i/\hat{\mu}_i),$$

em que $C(\mathbf{y})$ é uma função arbitrária, porém limitada. Pode-se, por exemplo, usar $C(\mathbf{y}) = \sum_{i=1}^{n} y_i/(1+y_i)$. Em geral os programas não aceitam ajustes da distribuição gama com $y = 0$.

Normal inversa

Para este caso $\tilde{\theta}_i = -1/2y_i^2$ e $\hat{\theta}_i = -1/2\hat{\mu}_i^2$. A função desvio fica então dada por

$$\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2/(y_i\hat{\mu}_i^2).$$

2.4.1 Medida R^2

Na regressão normal linear, como é bem conhecido, uma medida de qualidade do ajuste é dada pelo coeficiente de determinação (Seção 1.3), definido por

$$\mathrm{R}^2 = 1 - \frac{\mathrm{SQRes}}{\mathrm{SQT}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$

em que SQRes e SQT denotam, respectivamente, a soma de quadrados de resíduos e a soma de quadrados total, e $0 \leq \mathrm{R}^2 \leq 1$. Um refinamento dessa medida é obtido ajustando-se os graus de liberdade das formas quadráticos, obtendo-se o coeficiente de determinação ajustado

$$\overline{\mathrm{R}}^2 = 1 - \frac{\mathrm{SQRes}/(n-p)}{\mathrm{SQT}/(n-1)} = 1 - \frac{(n-1)\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{(n-p)\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

Mostra-se facilmente que $\overline{\mathrm{R}}^2 \leq \mathrm{R}^2$ e não necessariamente $\overline{\mathrm{R}}^2$ aumenta com o aumento do número de variáveis explicativas.

Uma extensão natural para os MLGs é dada por

$$\mathrm{R}^2 = 1 - \frac{\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})}{\mathrm{D}(\mathbf{y}; \bar{\mathbf{y}})},$$

em que $\mathrm{D}(\mathbf{y}; \bar{\mathbf{y}})$ denota o desvio do modelo apenas com o intercepto. Na prática o coeficiente de determinação para os MLGs (exceto para o caso normal) raramente é superior a $0,40$, sendo portanto esse valor utilizado como referência de excelente ajuste. Contudo, há várias outras propostas de pseudo R^2 em regressão. Por exemplo, a proposta de Cox e Snell em que

$\mathrm{R}^2 = 1 - \{\mathrm{L}(\bar{\mathbf{y}}; \mathbf{y})/\mathrm{L}(\hat{\boldsymbol{\mu}}; \mathbf{y})\}^{\frac{2}{n}}$. A ideia aqui é subtrair de 1 a n-ésima raiz de duas vezes a razão entre o menor valor e o valor ajustado do logaritmo da função de verossimilhança. Como essa quantidade em geral não alcança o valor 1, é proposta uma correção $\mathrm{R}^2 = [1-\{\mathrm{L}(\bar{\mathbf{y}}; \mathbf{y})/\mathrm{L}(\hat{\boldsymbol{\mu}}; \mathbf{y})\}^{\frac{2}{n}}]/[1-\mathrm{L}(\bar{\mathbf{y}}; \mathbf{y})^{\frac{2}{n}}]$. Esss duas quantidades podem ser obtidas na biblioteca `GAMLSS` do R (ver, por exemplo, Stasinopoulos et al., 2017) através dos comandos

```
require(gamlss)
ajuste = gamlss(resp ~ cov1 + cov2, family=PO)
Rsq(ajuste, type="both").
```

2.4.2 Resultados assintóticos

Embora seja usual comparar os valores observados da função desvio com os quantis da distribuição qui-quadrado com $n-p$ graus de liberdade, em geral $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})$ não segue assintoticamente uma χ^2_{n-p}. No caso binomial quando k é fixo e $n_i \to \infty$ para cada i, $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})$ segue sob a hipótese de que o modelo é verdadeiro uma χ^2_{k-p}. Isso não vale quando $n \to \infty$ e $n_i\mu_i(1-\mu_i)$ permanece limitado. Para o modelo de Poisson, quando $\mu_i \to \infty$ para todo i, segue que $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) \sim \chi^2_{n-p}$. No caso normal, como é conhecido para σ^2 fixo, $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) \sim \sigma^2\chi^2_{n-p}$. Lembre que $\mathrm{E}(\chi^2_r) = r$, assim um valor do desvio próximo de $n-p$ pode ser uma indicação de que o modelo está bem ajustado. Em geral, para os casos em que $\mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ depende do parâmetro de dispersão ϕ^{-1}, o seguinte resultado (Jørgensen, 1987) para a distribuição nula da função desvio pode ser utilizado:

$$\mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) \sim \chi^2_{n-p}, \quad \text{quando} \quad \phi \to \infty.$$

Isto é, quando a dispersão é pequena, fica razoável comparar os valores observados de $\mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ com os quantis da χ^2_{n-p}. Em particular, para o caso

normal linear, o resultado acima diz que $\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2/\sigma^2 \sim \chi^2_{n-p}$ quando $\sigma^2 \to 0$. No caso do modelo gama, o desvio estará bem aproximado por uma qui-quadrado com $n - p$ graus de liberdade à medida que o coeficiente de variação ficar próximo de zero.

2.4.3 Análise do desvio

Supor para o vetor de parâmetros $\boldsymbol{\beta}$ a partição $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, em que $\boldsymbol{\beta}_1$ é um vetor q-dimensional, enquanto $\boldsymbol{\beta}_2$ tem dimensão $p - q$ e ϕ é conhecido (ou fixo). Portanto, pode haver interesse em testar as hipóteses $\mathrm{H}_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ contra $\mathrm{H}_1 : \boldsymbol{\beta}_1 \neq \mathbf{0}$. As funções desvio correspondentes aos modelos sob H_0 e H_1 serão denotadas por $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}^0)$ e $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})$, respectivamente, em que $\hat{\boldsymbol{\mu}}^0$ é a estimativa de máxima verossimilhança sob H_0. A estatística do teste da razão de verossimilhanças fica nesse caso dada por

$$\xi_{RV} = \phi\{\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - \mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})\}, \tag{2.3}$$

isto é, a diferença entre dois desvios. Como é conhecido, sob a hipótese nula, $\xi_{RV} \sim \chi^2_q$ quando $n \to \infty$. De forma similar, pode-se definir a estatística

$$\mathrm{F} = \frac{\{\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - \mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})\}/q}{\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})/(n-p)}, \tag{2.4}$$

cuja distribuição nula assintótica é uma $\mathrm{F}_{q,(n-p)}$ quando o denominador de (2.4) é uma estimativa consistente de ϕ^{-1} (ver, por exemplo, Jørgensen, 1987). A vantagem em utilizar (2.4) em relação a (2.3) é que a estatística F não depende do parâmetro de dispersão. O resultado (2.4) também é verificado quando $\phi \to \infty$ e n é arbitrário. Quando ϕ é desconhecido a estatística do teste da razão de verossimilhanças assume uma expressão diferente de (2.3). A estatística F acima fica, no caso normal linear, reduzida à forma conhecida dada abaixo

$$\mathrm{F} = \frac{\sum_{i=1}^{n}(y_i - \hat{\mu}_i^0)^2 - \sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2}{qs^2},$$

em que $s^2 = \sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2/(n-p)$ é o erro quadrático médio do modelo com p parâmetros. A forma da estatística F dada em (2.4) pode ser obtida, em particular, quando tem-se uma hipótese de igualdades lineares num modelo de regressão normal linear. Como ilustração, supor o modelo

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\epsilon},$$

em que $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}_n)$, $\mathbf{X}$ é uma matriz $n \times p$, $\mathbf{I}_n$ é a matriz identidade de ordem n, $\mathbf{W}$ é aqui uma matriz $n \times q$, ambas de posto completo, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ e $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$. Considere as hipóteses

$$\mathrm{H}_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{0} \;\text{ contra }\; \mathrm{H}_1 : \mathbf{R}\boldsymbol{\theta} \neq \mathbf{0},$$

em que $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ e $\mathbf{R}$ é uma matriz $k \times (p+q)$ de posto completo. O acréscimo na soma de quadrados de resíduos devido às restrições em H_0 (Seção 1.4) é dado por

$$\mathrm{ASQ}(\mathbf{R}\boldsymbol{\theta} = \mathbf{0}) = (\mathbf{R}\hat{\boldsymbol{\theta}})^\top\{\mathbf{R}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{R}^\top\}^{-1}(\mathbf{R}\hat{\boldsymbol{\theta}}),$$

em que $\hat{\boldsymbol{\theta}} = (\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{y}$ e $\mathbf{Z} = (\mathbf{X}, \mathbf{W})$. A estatística F para testar H_0 fica então dada por

$$\mathrm{F} = \frac{\mathrm{ASQ}(\mathbf{R}\boldsymbol{\theta} = \mathbf{0})/k}{\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})/(n-p-q)},$$

em que $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})$ é o desvio do modelo completo com $p+q$ parâmetros e $\mathrm{ASQ}(\mathbf{R}\boldsymbol{\theta} = \mathbf{0}) = \mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - \mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})$, com $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}^0)$ sendo o desvio do modelo sob H_0. Portanto, F assume a forma

$$\mathrm{F} = \frac{\{\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - \mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})\}/k}{\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})/(n-p-q)},$$

e segue, sob H_0, uma distribuição $\mathrm{F}_{k,(n-p-q)}$. No caso de testar $\mathrm{H}_0 : \boldsymbol{\gamma} = \mathbf{0}$ contra $\mathrm{H}_1 : \boldsymbol{\gamma} \neq \mathbf{0}$, a matriz $\mathbf{R}$ tem dimensão $q \times (p+q)$ com a i-ésima linha tendo o valor 1 na posição $p+i$ e zeros nas demais posições. Essa

formulação pode também ser aplicada quando há interesse na inclusão de novas covariáveis num modelo de regressão normal linear.

Tabela 2.2
Análise do desvio (`ANODEV`) supondo dois fatores na parte sistemática.

Modelo	Desvio	Diferença	G.L.	Testando
Constante	D_0			
		$\mathrm{D}_0 - \mathrm{D}_A$	$n(A)-1$	A ignorando B
		$\mathrm{D}_0 - \mathrm{D}_B$	$n(B)-1$	B ignorando A
+A	D_A			
		$\mathrm{D}_A - \mathrm{D}_{A+B}$	$n(B)-1$	B|A ignorando AB
+B	D_B			
		$\mathrm{D}_B - \mathrm{D}_{A+B}$	$n(A)-1$	A|B ignorando AB
+A+B	D_{A+B}			
		$\mathrm{D}_{A+B} - \mathrm{D}_{AB}$	$\{n(A)-1\}\times$ $\{n(B)-1\}$	AB|A + B
+A+B+AB	D_{AB}			

Para ilustrar o uso das diferenças de desvios para hipóteses em modelos encaixados, supor um MLG com dois fatores, A e B. O fator A com n(A) níveis e o fator B com n(B) níveis. Na Tabela 2.2 tem-se os possíveis testes envolvendo os dois fatores. Em particular, se o interesse é testar a inclusão do fator B dado que o fator A já está no modelo, deve-se comparar a diferença $\phi\{\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}_A)-\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}_{A+B})\}$ com os níveis críticos da distribuição qui-quadrado com $\{n(\mathrm{B})-1\}$ graus de liberdade. Alternativamente, pode-se comparar o valor observado da estatística F correspondente com os níveis da distribuição F com $\{n(\mathrm{B})-1\}$ e $\{n-n(\mathrm{A})-n(\mathrm{B})+1\}$ graus de liberdade. No caso normal linear pode ser construída a tabela `ANOVA` utilizando a estatística F no lugar da diferença entre desvios. A vantagem disso é o fato do parâmetro de dispersão ϕ^{-1} não precisar ser estimado. Através do comando `anova()` o `R` fornece uma tabela `ANODEV` para os ajustes colocados como objetos. Por exemplo, supor que os objetos `fit1.reg`, `fit2.reg` e `fit3.reg` correspon-

dam aos ajustes de um MLG com um, dois e três fatores, respectivamente. Então, o comando

```
anova(fit1.reg,fit2.reg,fit3.reg)
```

fornece uma tabela `ANODEV` comparando os três fatores.

Como aplicação do `ANODEV`, considere o exemplo descrito na Seção 2.10.2 em que um modelo logístico com resposta Bernoulli é ajustado para explicar a ocorrência de câncer de pulmão numa amostra de 175 pacientes com processo infeccioso pulmonar, em que foram observadas as variáveis explicativas SEXO e IDADE e a intensidade das células HF e FF. A parte sistemática do modelo é representada abaixo

$$1 + \text{SEXO} + \text{IDADE} + \text{HL} + \text{FF},$$

em que 1 denota a presença de intercepto no modelo, SEXO (1:feminino, 0:masculino), IDADE (em anos) e HL e FF são dois fatores com 4 níveis cada um representando a intensidade de dois tipos de célula. A Tabela 2.3 resume alguns resultados.

Tabela 2.3
Análise do desvio referente ao exemplo sobre processo infeccioso pulmonar.

Modelo	Desvio	Diferença	G.L.	Testando
Constante	236,34	-	-	-
+ SEXO	235,20	1,14	1	SEXO
+ IDADE	188,22	46,98	1	IDADE | SEXO
+ HL	162,55	25,67	3	HL | SEXO + IDADE
+ FF	157,40	5,15	3	FF | SEXO + IDADE + HL

Para calcular os níveis descritivos das diferenças apresentadas na Tabela 2.3, pode-se aplicar o comando `pchisq(dv,q)` do `R`. Por exemplo, para calcular o nível descritivo referente ao efeito do fator SEXO, aplica-se

```
1 - pchisq(1.14,1)
```

obtendo-se P = 0, 285. Similarmente, para testar a inclusão de FF dado que já temos no modelo 1+SEXO+IDADE+HL, aplica-se

```
1 - pchisq(5.15,3)
```

e obtém-se P = 0, 1611, indicando que o fator FF é não significativo a 10%.

2.5 Função escore e informação de Fisher

2.5.1 Escore e Fisher para $\boldsymbol{\beta}$

Considere a partição $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$ e denote o logaritmo da função de verossimilhança por $\mathrm{L}(\boldsymbol{\theta})$. Para obter a função escore para o parâmetro $\boldsymbol{\beta}$ deriva-se inicialmente $\mathrm{L}(\boldsymbol{\theta})$ com relação a cada coeficiente

$$\begin{aligned}
\partial \mathrm{L}(\boldsymbol{\theta})/\partial\beta_j &= \sum_{i=1}^{n} \phi \left\{ y_i \frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial\eta_i}{\beta_j} - \frac{db(\theta_i)}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial\eta_i}{\partial\beta_j} \right\} \\
&= \sum_{i=1}^{n} \phi \left\{ y_i \frac{(d\mu_i/d\eta_i)}{V_i} x_{ij} - \mu_i \frac{(d\mu_i/d\eta_i)}{V_i} x_{ij} \right\} \\
&= \sum_{i=1}^{n} \phi \left\{ \sqrt{\frac{\omega_i}{V_i}} (y_i - \mu_i) x_{ij} \right\},
\end{aligned}$$

em que $\omega_i = (d\mu_i/d\eta_i)^2/V_i$. Logo, é possível escrever a função escore na forma matricial

$$\mathbf{U}_\beta(\boldsymbol{\theta}) = \frac{\partial \mathrm{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \phi \mathbf{X}^\top \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{y} - \boldsymbol{\mu}),$$

em que $\mathbf{X}$ é uma matriz $n \times p$ de posto completo cujas linhas serão denotadas por $\mathbf{x}_i^\top, i = 1, \ldots, n$, $\mathbf{W} = \mathrm{diag}\{\omega_1, \ldots, \omega_n\}$ é a matriz de pesos, $\mathbf{V} = \mathrm{diag}\{V_1, \ldots, V_n\}$, $\mathbf{y} = (y_1, \ldots, y_n)^\top$ e $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$.

A matriz de informação de Fisher para o parâmetro $\boldsymbol{\beta}$ é obtida derivando-se novamente $\mathrm{L}(\boldsymbol{\theta})$ com relação aos coeficientes

$$\begin{aligned}\partial^2 \mathrm{L}(\boldsymbol{\theta})/\partial\beta_j\partial\beta_\ell &= \phi\sum_{i=1}^{n}(y_i-\mu_i)\frac{d^2\theta_i}{d\mu_i^2}\left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij}x_{i\ell} \\ &\quad +\phi\sum_{i=1}^{n}(y_i-\mu_i)\frac{d\theta_i}{d\mu_i}\frac{d^2\mu_i}{d\eta_i^2}x_{ij}x_{i\ell} \\ &\quad -\phi\sum_{i=1}^{n}\frac{d\theta_i}{d\mu_i}\left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij}x_{i\ell},\end{aligned}$$

cujos valores esperados ficam dados por

$$\begin{aligned}\mathrm{E}\left\{\partial^2 \mathrm{L}(\boldsymbol{\theta})/\partial\beta_j\partial\beta_\ell\right\} &= -\phi\sum_{i=1}^{n}\frac{d\theta_i}{d\mu_i}\left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij}x_{i\ell} \\ &= -\phi\sum_{i=1}^{n}\frac{(d\mu_i/d\eta_i)^2}{V_i}x_{ij}x_{i\ell} \\ &= -\phi\sum_{i=1}^{n}\omega_i x_{ij}x_{i\ell}.\end{aligned}$$

Logo, a submatriz de informação de Fisher para $\boldsymbol{\beta}$ fica expressa na forma matricial

$$\mathbf{K}_{\beta\beta}(\boldsymbol{\theta}) = \mathrm{E}\left\{-\frac{\partial^2 \mathrm{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right\} = \phi\mathbf{X}^\top\mathbf{W}\mathbf{X}.$$

Em particular, para ligação canônica ($\theta_i = \eta_i$), essas quantidades tomam formas simplificadas

$$\mathbf{U}_\beta = \phi\mathbf{X}^\top(\mathbf{y}-\boldsymbol{\mu}) \;\text{ e }\; \mathbf{K}_{\beta\beta} = \phi\mathbf{X}^\top\mathbf{V}\mathbf{X},$$

respectivamente. Particionando o vetor de parâmetros tal que $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, a função escore e a matriz de informação de Fisher ficam para o parâmetro $\boldsymbol{\beta}_1$, respectivamente, dadas por $\mathbf{U}_{\beta_1} = \phi\mathbf{X}_1^\top\mathbf{W}^{\frac{1}{2}}\mathbf{V}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})$ e $\mathbf{K}_{\beta_1\beta_1} = \phi\mathbf{X}_1^\top\mathbf{W}\mathbf{X}_1$.

2.5.2 Escore e Fisher para ϕ

A função escore para o parâmetro ϕ fica dada por

$$\begin{aligned} \mathrm{U}_\phi(\boldsymbol{\theta}) &= \frac{\partial \mathrm{L}(\boldsymbol{\theta})}{\partial \phi} \\ &= \sum_{i=1}^{n}\{y_i\theta_i - b(\theta_i)\} + \sum_{i=1}^{n} c'(y_i;\phi), \end{aligned}$$

em que $c'(y_i;\phi) = dc(y_i,\phi)/d\phi$. Para obter a informação de Fisher para ϕ é preciso calcular $\partial^2 \mathrm{L}(\boldsymbol{\theta})/\partial\phi^2 = \sum_{i=1}^{n} c''(y_i;\phi)$, em que $c''(y_i;\phi) = d^2c(y_i,\phi)/d\phi^2$. Assim, a informação de Fisher para ϕ fica dada por

$$\mathrm{K}_{\phi\phi}(\boldsymbol{\theta}) = -\sum_{i=1}^{n} \mathrm{E}\{c''(Y_i;\phi)\}.$$

2.5.3 Ortogonalidade

Tem-se que $\partial^2 \mathrm{L}(\boldsymbol{\theta})/\partial\boldsymbol{\beta}\partial\phi = \sum_{i=1}^{n} \sqrt{\omega_i V_i^{-1}}(y_i - \mu_i)\mathbf{x}_i$. Portanto, verificamos facilmente que $\boldsymbol{\beta}$ e ϕ são ortogonais, isto é, $\mathbf{K}_{\beta\phi}(\boldsymbol{\theta}) = \mathrm{E}[-\partial^2 \mathrm{L}(\boldsymbol{\theta})/\partial\boldsymbol{\beta}\partial\phi] = \mathbf{0}$. Logo, segue que a matriz de informação de Fisher para $\boldsymbol{\theta}$ é bloco diagonal sendo dada por $\mathbf{K}_{\theta\theta} = \mathrm{diag}\{\mathbf{K}_{\beta\beta}, \mathrm{K}_{\phi\phi}\}$. A função escore para $\boldsymbol{\theta}$ fica dada por $\mathbf{U}_\theta = (\mathbf{U}_\beta^\top, \mathrm{U}_\phi)^\top$. A seguir são discutidos alguns casos particulares.

2.5.4 Casos particulares

Normal

A função de variância no caso normal é dada por $V(\mu) = 1$ ($d\mu/d\theta = 1$). Logo, $\omega = (d\theta/d\eta)^2$. Em particular para ligação canônica ($\theta = \eta$), obtém-se $\omega = 1$. Assim,

$$\mathbf{U}_\beta = \sigma^{-2}\mathbf{X}^\top(\mathbf{y} - \boldsymbol{\mu}) \;\text{ e }\; \mathbf{K}_{\beta\beta} = \sigma^{-2}\mathbf{X}^\top\mathbf{X},$$

como é conhecido. Segue ainda o resultado

$$\mathrm{U}_\phi = \sum_{i=1}^{n}\left(y_i\mu_i - \frac{\mu_i^2}{2}\right) + \sum_{i=1}^{n} c'(y_i;\phi),$$

em que $c'(y_i;\phi) = 1/2\phi - y_i^2/2$. Daí segue que $c''(y_i;\phi) = -1/2\phi^2$ e portanto $\mathrm{K}_{\phi\phi} = -\sum_{i=1}^{n}\mathrm{E}\{c''(Y_i;\phi)\} = n/2\phi^2$.

Poisson

Aqui a função de variância é dada por $V(\mu) = \mu$. Logo, $\omega = \mu(d\theta/d\eta)^2$. Para ligação canônica ($\log(\mu) = \eta$) os pesos são as próprias médias, isto é $\omega = \mu$. Em particular, para ligação raiz quadrada ($\sqrt{\mu} = \eta$), obtém-se $\omega = 4$. Assim, $\mathbf{U}_\beta = \mathbf{X}^\top\mathbf{V}^{-\frac{1}{2}}(\mathbf{y} - \boldsymbol{\mu})$ e $\mathbf{K}_{\beta\beta} = \mathbf{X}^\top\mathbf{X}$.

Binomial

No caso binomial, a função de variância é definida por $V(\mu) = \mu(1-\mu)$, em que $0 < \mu < 1$. Portanto, segue que $\omega = \mu(1-\mu)(d\theta/d\eta)^2$. Por convenção é assumido que $\omega = n\mu(1-\mu)(d\theta/d\eta)^2$ e $\phi = 1$. No caso de ligação canônica ($\log\{\mu/(1-\mu)\} = \eta$) os pesos são as variâncias das binomiais, isto é $\omega = n\mu(1-\mu)$. As matrizes $\mathbf{U}_\beta$ e $\mathbf{K}_{\beta\beta}$ ficam nesse caso dadas por

$$\mathbf{U}_\beta = \mathbf{X}^\top(\mathbf{y} - \boldsymbol{\mu}) \;\text{ e }\; \mathbf{K}_{\beta\beta} = \mathbf{X}^\top\mathbf{V}\mathbf{X},$$

em que $\mathbf{X}$ é uma matriz $k \times p$, $\mathbf{y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{\mu} = (n_1\mu_1, \ldots, n_k\mu_k)^\top$ e $\mathbf{V} = \mathrm{diag}\{n_1\mu_1(1-\mu_1), \ldots, n_k\mu_k(1-\mu_k)\}$.

Gama

Para o caso gama $V(\mu) = \mu^2$. Logo, $\omega = \mu^2(d\theta/d\eta)^2$. Em particular, para um modelo log-linear ($\log(\mu) = \eta$), obtém-se $d\mu/d\eta = \mu$, o que implica em $\omega = 1$. Assim, $\mathbf{U}_\beta = \phi\mathbf{X}^\top\mathbf{V}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$ e $\mathbf{K}_{\beta\beta} = \phi\mathbf{X}^\top\mathbf{X}$, similarmente ao

caso normal. Para ligação canônica, $\omega = \mu^2$. A função escore para ϕ fica dada por

$$\mathrm{U}_\phi = -\sum_{i=1}^{n}\left(\frac{y_i}{\mu_i} + \log(\mu_i)\right) + \sum_{i=1}^{n} c'(y_i;\phi),$$

em que $c'(y_i;\phi) = \log(y_i) + \log(\phi) + 1 - \psi(\phi)$ e $\psi(\phi) = \Gamma'(\phi)/\Gamma(\phi)$ é a função digama. Daí segue que $c''(y_i;\phi) = 1/\phi - \psi'(\phi)$ e portanto

$$\mathrm{K}_{\phi\phi} = -\sum_{i=1}^{n}\mathrm{E}\{c''(Y_i;\phi)\} = n\{\phi\psi'(\phi) - 1\}/\phi,$$

em que $\psi'(\phi) = d\psi(\phi)/d\phi$ é a função trigama.

Normal inversa

Neste caso a função de variância é dada por $V(\mu) = \mu^3$. Assim, $\omega = \mu^3(d\theta/d\eta)^2$. Pode ser muito razoável aplicar aqui um modelo log-linear, uma vez que as respostas são sempre positivas. No entanto, diferente dos modelos log-lineares com resposta de Poisson, os pesos aqui são inversamente proporcionais às médias, isto é $\omega = \mu^{-1}$. Em particular para ligação canônica, $\omega = \mu^3$, e portanto $\mathbf{U}_\beta = \phi\mathbf{X}^\top(\mathbf{y} - \boldsymbol{\mu})$ e $\mathbf{K}_{\beta\beta} = \phi\mathbf{X}^\top\mathbf{V}\mathbf{X}$. Tem-se ainda o resultado

$$\mathrm{U}_\phi = \sum_{i=1}^{n}\left(\frac{1}{\mu_i} - \frac{y_i}{2\mu_i^2}\right) + \sum_{i=1}^{n} c'(y_i;\phi),$$

em que $c'(y_i;\phi) = 1/2\phi - 1/2y_i$. Daí segue que $c''(y_i;\phi) = -1/2\phi^2$ e portanto $\mathrm{K}_{\phi\phi} = -\sum_{i=1}^{n}\mathrm{E}\{c''(Y_i;\phi)\} = n/2\phi^2$.

2.6 Estimação dos parâmetros

2.6.1 Estimação de $\boldsymbol{\beta}$

O processo iterativo de Newton-Raphson para a obtenção da estimativa de máxima verossimilhança de $\boldsymbol{\beta}$ é definido expandindo a função escore $\mathbf{U}_\beta$ em

torno de um valor inicial $\boldsymbol{\beta}^{(0)}$, tal que

$$\mathbf{U}_\beta \cong \mathbf{U}_\beta^{(0)} + \mathbf{U}_\beta^{'(0)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}),$$

em que $\mathbf{U}'_\beta$ denota a primeira derivada de $\mathbf{U}_\beta$ com respeito a $\boldsymbol{\beta}^\top$, sendo $\mathbf{U}_\beta{}'(0)$ e $\mathbf{U}_\beta^{(0)}$, respectivamente, essas quantidades avaliadas em $\boldsymbol{\beta}^{(0)}$. Assim, repetindo o procedimento acima, chega-se ao processo iterativo

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \{(-\mathbf{U}'_\beta)^{-1}\}^{(m)}\mathbf{U}_\beta^{(m)},$$

$m = 0, 1, \ldots$. Como a matriz $-\mathbf{U}'_\beta$ pode não ser positiva definida, a aplicação do método escore de Fisher substituindo a matriz $-\mathbf{U}'_\beta$ pelo correspondente valor esperado $\mathbf{K}_{\beta\beta}$ pode ser mais conveniente. Isso resulta no seguinte processo iterativo:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \{\mathbf{K}_{\beta\beta}^{-1}\}^{(m)}\mathbf{U}_\beta^{(m)},$$

$m = 0, \ldots$. Trabalhando um pouco o lado direito da expressão acima, chega-se a um processo iterativo de mínimos quadrados reponderados

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top\mathbf{W}^{(m)}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{W}^{(m)}\mathbf{z}^{(m)}, \tag{2.5}$$

$m = 0, 1, \ldots$, em que $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{-\frac{1}{2}}\mathbf{V}^{-\frac{1}{2}}(\mathbf{y} - \boldsymbol{\mu})$. A quantidade $\mathbf{z}$ desempenha o papel de uma variável dependente modificada, enquanto $\mathbf{W}$ é uma matriz de pesos que muda a cada passo do processo iterativo. A convergência de (2.5) ocorre em geral num número finito de passos, independente dos valores iniciais utilizados. É usual iniciar (2.5) com $\boldsymbol{\eta}^{(0)} = (g(y_1), \ldots, g(y_n))^\top$.

Apenas como ilustração, para o caso logístico binomial, tem-se que $\omega = n\mu(1-\mu)$ e variável dependente modificada dada por $z = \eta + (y - n\mu)/n\mu(1-\mu)$. Lembrando, para o modelo normal linear não é preciso recorrer ao processo iterativo (2.5) para a obtenção da estimativa de máxima verossimilhança. Nesse caso, $\hat{\boldsymbol{\beta}}$ assume a forma fechada

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}.$$

Observa-se que o lado direito de (2.5) não depende de ϕ. Portanto, para obter $\hat{\boldsymbol{\beta}}$ não é preciso conhecer ϕ.

2.6.2 Estimação de ϕ

Igualando a função escore U_ϕ a zero chega-se à seguinte solução:

$$\sum_{i=1}^{n} c'(y_i;\hat{\phi}) = \frac{1}{2}\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}) - \sum_{i=1}^{n}\{y_i\tilde{\theta}_i - b(\tilde{\theta}_i)\},$$

em que $\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}})$ denota o desvio do modelo sob investigação. Verifica-se que a estimativa de máxima verossimilhança para ϕ nos casos normal e normal inversa, igualando U_ϕ a zero, é dada por

$$\hat{\phi} = \frac{n}{\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}})}.$$

Para o caso gama, a estimativa de máxima verossimilhança de ϕ sai da equação

$$2n\{\log\hat{\phi} - \psi(\hat{\phi})\} = \mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}).$$

A equação acima pode ser resolvida diretamente pelo `R` através da biblioteca `MASS` (Venables e Ripley, 1999). Como ilustração, supor que os resultados do ajuste sejam guardados em `fit.model`. Então, para encontrar a estimativa de máxima verossimilhança de ϕ com o respectivo erro padrão aproximado deve-se aplicar os comandos

```
require(MASS)
gamma.shape(fit.model).
```

Um outro estimador consistente para ϕ (de momentos) que não envolve processo iterativo é baseado na estatística de Pearson, sendo dado por

$$\hat{\phi} = \frac{(n-p)}{\sum_{i=1}^{n}\{\frac{(y_i-\hat{\mu}_i)^2}{V(\hat{\mu}_i)}\}}.$$

A suposição aqui é que $\hat{\boldsymbol{\beta}}$ tem sido consistentemente estimado. O `R` solta a estimativa $\hat{\phi} = (n-p)/\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}})$ que não é consistente para ϕ.

2.6.3 Distribuição assintótica

Para mostrar heuristicamente que $\hat{\boldsymbol{\beta}}$ e $\hat{\phi}$ são assintoticamente normais e independentes, considere os resultados abaixo

$$\mathrm{E}(\mathbf{U}_\theta) = \mathbf{0} \;\text{ e }\; \mathrm{Var}(\mathbf{U}_\theta) = \mathbf{K}_{\theta\theta},$$

com as funções escore de $\boldsymbol{\beta}$ e ϕ sendo, respectivamente, expressas nas formas $\mathbf{U}_\beta = \sum_{i=1}^{n} \mathrm{U}_{i\beta}$, em que

$$\mathrm{U}_{i\beta} = \phi\sqrt{\omega_i V_i^{-1}}(y_i - \mu_i)\mathbf{x}_i \;\text{ e }\; \mathrm{U}_\phi = \sum_{i=1}^{n} \mathrm{U}_{i\phi},$$

com $\mathrm{U}_{i\phi} = \{y_i\theta_i - b(\theta_i)\} + c'(y_i;\phi)$. Portanto, para n grande, segue pelo Teorema Central do Limite que $\mathbf{U}_\theta \sim \mathrm{N}_{p+1}(\mathbf{0}, \mathbf{K}_{\theta\theta})$. Em particular, assintoticamente $\mathbf{U}_\beta \sim \mathrm{N}_p(\mathbf{0}, \mathbf{K}_{\beta\beta})$ e $\mathrm{U}_\phi \sim \mathrm{N}(0, \mathrm{K}_{\phi\phi})$ e $\mathbf{U}_\beta$ e U_ϕ são independentes.

Expandindo $\mathbf{U}_{\hat{\theta}}$ em série de Taylor em torno de $\boldsymbol{\theta}$ obtém-se

$$\mathbf{U}_{\hat{\theta}} \cong \mathbf{U}_\theta + \mathbf{U}'_\theta(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}),$$

em que $\mathbf{U}'_\theta = \partial\mathbf{U}_\theta/\partial\boldsymbol{\theta}^\top$. Assim, como $\hat{\boldsymbol{\theta}}$ é o estimador de máxima verossimilhança de $\boldsymbol{\theta}$ tem-se que $\mathbf{U}_{\hat{\theta}} = \mathbf{0}$ e daí segue a relação

$$\hat{\boldsymbol{\theta}} \cong \boldsymbol{\theta} + (-\mathbf{U}'_\theta)^{-1}\mathbf{U}_\theta.$$

Supondo que para n grande $-\mathbf{U}'_\theta \cong \mathbf{K}_{\theta\theta}$ (para ligação canônica $\mathbf{K}_{\beta\beta} = -\mathbf{U}'_\beta$), então obtém-se

$$\hat{\boldsymbol{\theta}} \cong \boldsymbol{\theta} + \mathbf{K}_{\theta\theta}^{-1}\mathbf{U}_\theta,$$

ou seja, para n grande $\hat{\boldsymbol{\theta}} \sim \mathrm{N}_{p+1}(\boldsymbol{\theta}, \mathbf{K}_{\theta\theta}^{-1})$. Como $\mathbf{K}_{\theta\theta} = \mathrm{diag}\{\mathbf{K}_{\beta\beta}, \mathrm{K}_{\phi\phi}\}$ então assintoticamente segue que $\hat{\boldsymbol{\beta}} \sim \mathrm{N}_p(\boldsymbol{\beta}, \mathbf{K}_{\beta\beta}^{-1})$ e $\hat{\phi} \sim \mathrm{N}(0, \mathrm{K}_{\phi\phi}^{-1})$ e $\hat{\boldsymbol{\beta}}$ e $\hat{\phi}$ são independentes. Demonstrações mais rigorosas desses resultados podem ser encontradas, por exemplo, em Fahrmeir e Kaufmann (1985) e Sen e Singer (1993, Cap. 7).

2.7 Teste de hipóteses

2.7.1 Hipóteses simples

Buse (1982) apresenta de uma forma bastante didática a interpretação geométrica dos testes da razão de verossimilhanças, escore e Wald para o caso de hipóteses simples. A seguir são apresentadas as generalizações para os MLGs. Supor, inicialmente, a seguinte situação de hipóteses simples:

$$\mathrm{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^0 \ \text{ contra } \ \mathrm{H}_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}^0,$$

em que $\boldsymbol{\beta}^0$ é um vetor p-dimensional conhecido e ϕ é também assumido conhecido.

Teste da razão de verossimilhanças

O teste da razão de verossimilhanças, no caso de hipóteses simples, é usualmente definido por

$$\xi_{RV} = 2\{\mathrm{L}(\hat{\boldsymbol{\beta}}) - \mathrm{L}(\boldsymbol{\beta}^0)\}.$$

Essa estatística pode também ser expressa, para os MLGs, como a diferença entre duas funções desvio

$$\xi_{RV} = \phi\{\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - \mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})\},$$

em que $\hat{\boldsymbol{\mu}}^0 = \mathbf{g}^{-1}(\hat{\boldsymbol{\eta}}^0)$, $\hat{\boldsymbol{\eta}}^0 = \mathbf{X}\boldsymbol{\beta}^0$. Em particular, para o caso normal linear, tem-se que $\xi_{RV} = \{\sum_{i=1}^n (y_i - \hat{\mu}_i^0)^2 - \sum_{i=1}^n (y_i - \hat{\mu}_i)^2\}/\sigma^2$.

Teste de Wald

O teste de Wald é definido, nesse caso, por

$$\xi_W = [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0]^\top \hat{\mathrm{V}}\mathrm{ar}^{-1}(\hat{\boldsymbol{\beta}})[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0],$$

em que $\hat{\text{Var}}(\hat{\boldsymbol{\beta}})$ denota a matriz de variância-covariância assintótica de $\hat{\boldsymbol{\beta}}$ estimada em $\hat{\boldsymbol{\beta}}$. Para os MLGs, $\hat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \mathbf{K}^{-1}(\hat{\boldsymbol{\beta}})$. Assim, a estatística de Wald fica reexpressa na forma

$$\xi_W = \phi[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0]^\top(\mathbf{X}^\top\hat{\mathbf{W}}\mathbf{X})[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0].$$

Em particular, para o caso de $p = 1$, o teste de Wald é equivalente ao teste z^2 usual

$$\xi_W = \frac{(\hat{\beta} - \beta^0)^2}{\hat{\text{Var}}(\hat{\beta})}.$$

Um problema com a estatística de Wald, especialmente quando $\boldsymbol{\eta}(\boldsymbol{\beta})$ é não linear em $\boldsymbol{\beta}$, é a dependência de ξ_W com a parametrização utilizada. Isto é, duas formas diferentes e equivalentes para $\boldsymbol{\eta}(\boldsymbol{\beta})$, podem levar a diferentes valores de ξ_W.

Teste de escore

O teste de escore, também conhecido como teste de Rao, é definido quando $\mathbf{U}_\beta(\hat{\boldsymbol{\beta}}) = \mathbf{0}$ por

$$\xi_{SR} = \mathbf{U}_\beta(\boldsymbol{\beta}^0)^\top\hat{\text{Var}}_0(\hat{\boldsymbol{\beta}})\mathbf{U}_\beta(\boldsymbol{\beta}^0),$$

em que $\hat{\text{Var}}_0(\hat{\boldsymbol{\beta}})$ denota que a variância assintótica de $\hat{\boldsymbol{\beta}}$ está sendo estimada sob H_0. Para os MLGs tem-se que

$$\xi_{SR} = \phi^{-1}\mathbf{U}_\beta(\boldsymbol{\beta}^0)^\top(\mathbf{X}^\top\hat{\mathbf{W}}_0\mathbf{X})^{-1}\mathbf{U}_\beta(\boldsymbol{\beta}^0),$$

em que $\hat{\mathbf{W}}_0$ é estimado sob H_0, embora tenha a forma do modelo em H_1. A estatística de escore pode ser muito conveniente em situações em que a hipótese alternativa é bem mais complexa do que a hipótese nula. Nesses casos, somente seria necessário estimar os parâmetros sob H_1 quando o modelo em H_0 fosse rejeitado. Novamente, ilustrando o caso normal linear, tem-se

que a estatística de escore fica expressa na forma

$$\xi_{SR} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0)^\top \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^0)/\sigma^2.$$

Observe que, nesse caso, as estatísticas ξ_{RV} e ξ_W coincidem com ξ_{SR}.

Teste F

A estatística F, que foi definida em (2.4), assume a seguinte forma para o caso de hipóteses simples:

$$\mathrm{F} = \frac{\{\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}^0) - \mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}})\}/p}{\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}})/(n-p)},$$

que para $\phi \to \infty$ e sob H_0 segue uma $\mathrm{F}_{p,(n-p)}$. Esse resultado vale também para $n \to \infty$ quando coloca-se no denominador da estatística F uma estimativa consistente para ϕ^{-1}. Uma propriedade interessante das estatísticas ξ_{RV}, ξ_{SR} e F é o fato de serem invariantes com reparametrizações. Isso pode ser muito útil na construção de regiões de confiança para os parâmetros. A estatística F tem a vantagem adicional de não depender do parâmetro de dispersão ϕ^{-1}. Como essa estatística pode ser obtida diretamente de funções desvio, talvez seja a mais conveniente para uso prático. Assintoticamente e sob a hipótese nula, segue que ξ_{RV}, ξ_W e $\xi_{SR} \sim \chi^2_p$.

Uma região assintótica de confiança para $\boldsymbol{\beta}$ baseada no teste de Wald e com coeficiente de confiança $(1-\alpha)$, é dada por

$$[\boldsymbol{\beta}; (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top(\mathbf{X}^\top\hat{\mathbf{W}}\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq \phi^{-1}\chi^2_p(1-\alpha)],$$

em que $\chi^2_p(1-\alpha)$ denota o quantil $(1-\alpha)$ de uma distribuição qui-quadrado com p graus de liberdade. Como essa região pode depender da parametrização utilizada quando η é não linear (ver, por exemplo, Ratkowsky, 1983), pode ser mais conveniente, nesses casos, construir a região utilizando uma

das estatísticas invariantes. Em particular, se a estatística da razão de verossimilhanças for escolhida, a região assintótica fica dada por

$$[\boldsymbol{\beta}; 2\{\mathrm{L}(\hat{\boldsymbol{\beta}}) - \mathrm{L}(\boldsymbol{\beta})\} \leq \chi^2_p(1-\alpha)].$$

Se há interesse num subconjunto $\boldsymbol{\beta}_1$ q-dimensional, a região assintótica de confiança utilizando as estatísticas de Wald e da razão de verossimilhanças ficam, respectivamente, dadas por

$$[\boldsymbol{\beta}; (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta})^\top \hat{\mathrm{V}}\mathrm{ar}^{-1}(\hat{\boldsymbol{\beta}}_1)(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}) \leq \phi^{-1}\chi^2_q(1-\alpha)]$$

e

$$[\boldsymbol{\beta}; 2\{\mathrm{L}(\hat{\boldsymbol{\beta}}) - \mathrm{L}(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}_2(\boldsymbol{\beta}))\} \leq \chi^2_q(1-\alpha)],$$

em que $\boldsymbol{\beta}$ é aqui q-dimensional e $\hat{\boldsymbol{\beta}}_2(\boldsymbol{\beta})$ é a estimativa de máxima verossimilhança de $\boldsymbol{\beta}_2$ dado $\boldsymbol{\beta}$ (ver, por exemplo, Seber e Wild, 1989).

2.7.2 Modelos encaixados

ϕ conhecido

Supor novamente a partição $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ definida na Seção 2.4.3 e as seguintes hipóteses: $\mathrm{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0$ contra $\mathrm{H}_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_1^0$. Para esse caso tem-se que

$$\xi_{RV} = \phi\{\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - \mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})\},$$

em que $\hat{\boldsymbol{\mu}}^0$ é a estimativa de máxima verossimilhança do MLG com parte sistemática $\eta = \hat{\eta}_1^0 + \eta_2$, em que $\hat{\eta}_1^0 = \sum_{j=1}^q x_j\beta_j^0$ e $\eta_2 = \sum_{j=q+1}^p x_j\beta_j$. A quantidade $\hat{\eta}_1^0$ desempenha o papel de um *offset* (parte conhecida no preditor linear), conforme a nomenclatura de modelos lineares generalizados. Para ilustrar a utilização do *offset*, supor um modelo de Poisson com ligação log-linear, resposta `resp`, covariáveis `cov1` e `cov2` e *offset* dado por `logt0`. Para

ajustar o modelo e armazenar os resultados em `fit1.poisson` deve-se aplicar o comando

`fit1.poisson = glm(resp ~ cov1 + cov2 + offset(logt0), family= poisson).`

Esse tipo de recurso é muito utilizado em estudos de seguimento em que cada indivíduo é observado durante um tempo diferente. Como ilustração, supor um MLG com distribuição normal inversa, ligação canônica e preditor linear dado por $\eta = \beta_1 + \beta_2 \text{cov}_2 + \beta_3 \text{cov}_3$ e que o interesse é testar $\text{H}_0 : \beta_2 = b$, em que b é uma constante diferente de zero, contra $\text{H}_1 : \beta_2 \neq b$. Os ajustes correspondentes a H_0 e H_1 são, respectivamente, dados por

```
fit1.ni = glm(resp ~ cov3 + offset(b*cov2),
family=inverse.gaussian)
fit2.ni = glm(resp ~ cov2+cov3, family=inverse.gaussian).
```

Logo, de (1.4), a estatística F para testar $\text{H}_0 : \beta_2 = b$ contra $\text{H}_1 : \beta_2 \neq b$ fica dada por

```
d1 = deviance(fit1.ni)
d2 = deviance(fit2.ni)
F = (d1 - d2)/(d2/(n-3)).
```

Em particular, o *offset* desaparece para $b = 0$. O ajuste, nesse caso, fica simplesmente dado por

```
fit1.ni = glm(resp ~ cov3, family=inverse.gaussian).
```

Teste de Wald

Para testar H_0, a estatística de Wald fica expressa na forma

$$\xi_W = [\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0]^\top \hat{\text{V}}\text{ar}^{-1}(\hat{\boldsymbol{\beta}}_1)[\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0],$$

em que $\hat{\boldsymbol{\beta}}_1$ sai do vetor $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1^\top, \hat{\boldsymbol{\beta}}_2^\top)^\top$. Usando resultados conhecidos de álgebra linear, mostra-se que a variância assintótica de $\hat{\boldsymbol{\beta}}_1$ é dada por

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_1) = \phi^{-1}[\mathbf{X}_1^\top \mathbf{W}^{\frac{1}{2}} \mathbf{M}_2 \mathbf{W}^{\frac{1}{2}} \mathbf{X}_1]^{-1},$$

em que $\mathbf{X}_1$ sai da partição $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, sendo portanto $n \times q$, $\mathbf{X}_2$ é $n \times (p-q)$, $\mathbf{M}_2 = \mathbf{I}_n - \mathbf{H}_2$ e $\mathbf{H}_2 = \mathbf{W}^{\frac{1}{2}} \mathbf{X}_2 (\mathbf{X}_2^\top \mathbf{W} \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{W}^{\frac{1}{2}}$ é a matriz de projeção ortogonal de vetores do $\mathcal{R}^n$ no subespaço gerado pelas colunas da matriz $\mathbf{W}^{\frac{1}{2}} \mathbf{X}_2$. Em particular, no caso normal linear, tem-se as simplificações $\mathbf{H}_2 = \mathbf{X}_2 (\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top$ e $\mathrm{Var}(\hat{\boldsymbol{\beta}}_1) = \sigma^2 [\mathbf{X}_1^\top (\mathbf{I}_n - \mathbf{H}_2) \mathbf{X}_1]^{-1}$.

Teste de escore

A função escore pode ser expressa na forma $\mathbf{U}_\beta = \phi^{\frac{1}{2}} \mathbf{X}^\top \mathbf{W}^{\frac{1}{2}} \mathbf{r}_P$, em que $\mathbf{r}_P = \phi^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{y} - \boldsymbol{\mu})$ é conhecido como resíduo de Pearson. Observe que $\mathbf{r}_P$ tem a mesma distribuição de $\mathbf{Y}$, no entanto, $\mathrm{E}(\mathbf{r}_P) = \mathbf{0}$ e $\mathrm{Var}(\mathbf{r}_P) = \mathbf{I}_n$. O teste de escore é definido por

$$\xi_{SR} = \mathbf{U}_{\beta_1}(\hat{\boldsymbol{\beta}}^0)^\top \hat{\mathrm{V}}\mathrm{ar}_0(\hat{\boldsymbol{\beta}}_1) \mathbf{U}_{\beta_1}(\hat{\boldsymbol{\beta}}^0),$$

em que $\mathbf{U}_{\beta_1}(\boldsymbol{\beta}) = \partial \mathrm{L}(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}_1 = \phi \mathbf{X}_1^\top \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{y} - \boldsymbol{\mu})$, $\hat{\boldsymbol{\beta}}^0 = (\boldsymbol{\beta}_1^{0\top}, \hat{\boldsymbol{\beta}}_2^{0\top})^\top$ e $\hat{\boldsymbol{\beta}}_2^0$ é a estimativa de máxima verossimilhança de $\boldsymbol{\beta}_2$ sob o modelo com parte sistemática $\boldsymbol{\eta} = \hat{\boldsymbol{\eta}}_1^0 + \boldsymbol{\eta}_2$, isto é, sob H_0, em que $\hat{\boldsymbol{\eta}}_1^0 = \mathbf{X}_1 \boldsymbol{\beta}_1^0$ e $\boldsymbol{\eta}_2 = \mathbf{X}_2 \boldsymbol{\beta}_2$. Trabalhando um pouco mais a expressão para $\mathrm{Var}(\hat{\boldsymbol{\beta}}_1)$, chega-se ao seguinte resultado:

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_1) = \phi^{-1} (\mathbf{R}^\top \mathbf{W} \mathbf{R})^{-1},$$

em que $\mathbf{R} = \mathbf{X}_1 - \mathbf{X}_2 \mathbf{C}$ e $\mathbf{C} = (\mathbf{X}_2^\top \mathbf{W} \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{W} \mathbf{X}_1$. Aqui $\mathbf{C}$ é uma matriz $n \times q$ cuja j-ésima coluna é o vetor de coeficientes da regressão linear (com pesos $\mathbf{W}$) da j-ésima coluna de $\mathbf{X}_1$ sobre $\mathbf{X}_2$. Assim, $\mathbf{R}$ pode ser interpretado como sendo uma matriz $n \times q$ de resíduos. A j-ésima coluna de $\mathbf{R}$ corresponde

aos resíduos ordinários da regressão linear (com pesos $\mathbf{W}$) da j-ésima coluna de $\mathbf{X}_1$ sobre $\mathbf{X}_2$. Assim, o teste de escore fica reexpresso na forma (vide Cordeiro, et al., 1993)

$$\xi_{SR} = \hat{\mathbf{r}}_{P_0}^{\top} \hat{\mathbf{W}}_0^{\frac{1}{2}} \mathbf{X}_1 (\hat{\mathbf{R}}_0^{\top} \hat{\mathbf{W}}_0 \hat{\mathbf{R}}_0)^{-1} \mathbf{X}_1^{\top} \hat{\mathbf{W}}_0^{\frac{1}{2}} \hat{\mathbf{r}}_{P_0},$$

com as quantidades $\hat{\mathbf{r}}_{P_0}$, $\hat{\mathbf{W}}_0$ e $\hat{\mathbf{R}}_0$ sendo avaliadas em $\hat{\boldsymbol{\beta}}^0$.

Para ilustrar o cálculo da estatística de escore, supor um MLG com preditor linear dado por $\eta = \beta_1 + \beta_2 \text{cov}_2 + \beta_3 \text{cov}_3 + \beta_4 \text{cov}_4$ e que o interesse é testar $\text{H}_0 : \beta_3 = \beta_4 = 0$. As matrizes $\mathbf{X}_1$ e $\mathbf{X}_2$ serão então dadas por $\mathbf{X}_1 = [\text{cov}_3 \ , \ \text{cov}_4]$ e $\mathbf{X}_2 = [1 \ , \ \text{cov}_2]$. Para um modelo de Poisson, por exemplo com ligação canônica, tem-se que $\omega = \mu$. Logo, $\hat{\mathbf{W}}_0 = \text{diag}\{\hat{\mu}_1^0, \ldots, \hat{\mu}_n^0\}$, em que $\hat{\mu}_1^0, \ldots, \hat{\mu}_n^0$ são os pesos sob H_0, ou seja, os pesos do modelo ajustado de Poisson com preditor linear $\eta = \beta_1 + \beta_2 \text{cov}_2$. Portanto, é preciso apenas fazer esse ajuste e computar $\hat{\mathbf{W}}_0$, $\hat{\mathbf{R}}_0$, $\hat{\mathbf{r}}_{P_0}$ e finalmente ξ_{SR}. Chamando no `R` os pesos por `w`, $\hat{\mathbf{W}}_0$ por `W`, $\hat{\mathbf{r}}_{P_0}$ por `rp` e $\hat{\mathbf{R}}_0$ por `R`, os passos para o cálculo de ξ_{SR} são dados abaixo

```
X1 = cbind(cov3 , cov4)
X2 = cbind(1 , cov2)
fit.poisson = glm(resp ~ cov2, family=poisson)
rp = resid(fit.poisson, type=``pearson'')
w = fit.poisson$weights
W = diag(w)
A = solve(t(X2)%*%W%*%X2)
C1 = A%*%t(X2)%*%W%*%cov3
C2 = A%*%t(X2)%*%W%*%cov4
C = cbind(C1 , C2)
R = X1 - X2%*%C
```

```
SR = solve(t(R)%*%W%*%R)
SR = t(rp)%*%sqrt(W)%*%X1%*%SR%*%t(X1)%*%sqrt(W)%*%rp.
```

Em particular, para o caso normal linear, $\mathbf{C} = (\mathbf{X}_2^\top\mathbf{X}_2)^{-1}\mathbf{X}_2^\top\mathbf{X}_1$ e $\mathbf{r}_P = (\mathbf{y} - \boldsymbol{\mu})/\sigma$. Logo, $\xi_{SR} = \sigma^{-2}(\mathbf{y} - \hat{\boldsymbol{\mu}}^0)^\top\mathbf{X}_1(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{X}_1\top(\mathbf{y} - \hat{\boldsymbol{\mu}}^0)$, em que $\mathbf{R} = \mathbf{X}_1 - \mathbf{X}_2(\mathbf{X}_2^\top\mathbf{X}_2)^{-1}\mathbf{X}_2^\top\mathbf{X}_1 = (\mathbf{I}_n - \mathbf{H}_2)\mathbf{X}_1$. Aqui, também as estatísticas da razão de verossimilhanças e de Wald coincidem com a estatística de escore. Isso em geral vale para o modelo normal linear.

A estatística de Wald fica, analogamente ao caso anterior, dada por

$$\xi_W = \phi[\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0]^\top[\hat{\mathbf{R}}^\top\hat{\mathbf{W}}\hat{\mathbf{R}}][\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0].$$

O cálculo de $\hat{\mathbf{R}}$ segue os mesmos passos descritos para o cálculo do teste de escore, com a única diferença de que os pesos sairão do ajuste do modelo com todos os parâmetros. As mudanças nos comandos são as seguintes:

```
fit1.poissom = glm( resp ~ cov2 + cov3 + cov4,
family=poisson)
w = fit1.poisson$weights
W = diag(w).
```

Sob H_0 e para grandes amostras, tem-se que ξ_{RV}, ξ_W e $\xi_{SR} \sim \chi_q^2$.

ϕ desconhecido

No caso de ϕ ser desconhecido e o interesse for testar $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0$ contra $H_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_1^0$, as estatísticas ξ_{RV}, ξ_{SR} e ξ_W assumem formas diferentes daquelas apresentadas para o caso de ϕ ser conhecido. Em particular, denote por $\hat{\phi}^0$ e $\hat{\phi}$ as estimativas de máxima verossimilhança de ϕ sob H_0 e H_1, respectivamente. Para facilitar a notação da estatística ξ_{RV} usa-se o resultado $c(y, \phi) = d(\phi) + \phi a(y) + u(y)$ válido para algumas distribuições da família exponencial dada em (2.1) (por exemplo normal, gama e normal inversa), em

que $a(\cdot), d(\cdot)$ e $u(\cdot)$ são funções diferenciáveis. Assim, a estatística da razão de verossimilhanças fica expressa na forma

$$\xi_{RV} = 2\{\hat{\phi}t(\hat{\boldsymbol{\mu}}) - \hat{\phi}^0 t(\hat{\boldsymbol{\mu}}^0)\} + 2n\{d(\hat{\phi}) - d(\hat{\phi}^0)\},$$

em que $t(\boldsymbol{\mu}) = \sum_{i=1}^{n}\{y_i\theta_i - b(\theta_i) + a(y_i)\}$ e $\theta_i = \theta(\mu_i)$. Para o modelo gama, por exemplo, tem-se que $t(\boldsymbol{\mu}) = \sum_{i=1}^{n}\{\log(y_i/\mu_i) - y_i/\mu_i\}$ e $d(\phi) = \phi\log(\phi) - \log\{\Gamma(\phi)\}$. A estatística de Wald fica, por sua vez, dada por

$$\begin{aligned}\xi_W &= [\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0]^\top \hat{\mathrm{V}}\mathrm{ar}^{-1}(\hat{\boldsymbol{\beta}}_1)[\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0] \\ &= \hat{\phi}[\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0]^\top(\hat{\mathbf{R}}^\top\hat{\mathbf{W}}\hat{\mathbf{R}})[\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0].\end{aligned}$$

Já a estatística de escore assume a forma

$$\begin{aligned}\xi_{SR} &= \mathbf{U}_{\beta_1}(\hat{\boldsymbol{\theta}}^0)^\top \hat{\mathrm{V}}\mathrm{ar}_0(\hat{\boldsymbol{\beta}}_1)\mathbf{U}_{\beta_1}(\hat{\boldsymbol{\theta}}^0) \\ &= \hat{\mathbf{r}}_{P_0}^\top \hat{\mathbf{W}}_0^{\frac{1}{2}}\mathbf{X}_1(\hat{\mathbf{R}}_0^\top\hat{\mathbf{W}}_0\hat{\mathbf{R}}_0)^{-1}\mathbf{X}_1^\top\hat{\mathbf{W}}_0^{\frac{1}{2}}\hat{\mathbf{r}}_{P_0},\end{aligned}$$

em que $\hat{\mathbf{r}}_{P_0} = \sqrt{\hat{\phi}^0}\hat{\mathbf{V}}_0^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}^0)$ e $\hat{\boldsymbol{\theta}}^0 = (\hat{\boldsymbol{\beta}}^{0\top}, \hat{\phi}^0)^\top$ é a estimativa de máxima verossimilhança de $\boldsymbol{\theta}$ sob H_0. As três estatísticas seguem assintoticamente e sob H_0 distribuição χ^2_q.

2.7.3 Modelo de análise de variância

Como ilustração supor o modelo de análise de variância balanceado com um fator e dois grupos

$$g(\mu_{ij}) = \alpha + \beta_i,$$

em que $i = 1, 2$, $j = 1, \ldots, m$, $\beta_1 = 0$, $\beta_2 = \beta$ e ϕ é conhecido. Considere as hipóteses $\mathrm{H}_0 : \beta = 0$ contra $\mathrm{H}_1 : \beta \neq 0$. Aqui $\mathbf{X}_2$ é um vetor $2m \times 1$ de $1's$ enquanto $\mathbf{X}_1$ é um vetor $2m \times 1$ com $0's$ nas m primeiras posições e $1's$ nas m restantes. Daí segue que $\mathbf{X}_2^\top\mathbf{W}\mathbf{X}_2 = m(\omega_1 + \omega_2)$, $\mathbf{X}_2^\top\mathbf{W}\mathbf{X}_1 = m\omega_2$,

$C = \omega_2/(\omega_1 + \omega_2)$ e consequentemente

$$\mathbf{R}^\top \mathbf{W}\mathbf{R} = \frac{m\omega_1\omega_2}{(\omega_1 + \omega_2)},$$

em que ω_1 e ω_2 são os pesos correspondentes aos dois grupos. A estatística de escore fica então dada por

$$\xi_{SR} = \frac{2}{m}\left(\sum_{j=1}^{m} \hat{r}^0_{P_{2j}}\right)^2,$$

em que $\hat{r}^0_{P_{2j}}$, $j = 1, \ldots, m$, são os resíduos estimados de Pearson, sob H_0, correspondentes ao segundo grupo, sendo dados por $\hat{r}^0_{P_{2j}} = \phi^{\frac{1}{2}}(y_{2j} - \hat{\mu}^0)/\hat{V}_0^{\frac{1}{2}}$. Em particular, sob a hipótese nula, $\hat{\mu}^0 = \bar{y}$. Assim, obtém-se a simplificação

$$\xi_{SR} = \frac{\phi m}{2\hat{V}_0}(\bar{y}_1 - \bar{y}_2)^2, \tag{2.6}$$

em que $\bar{y}_1$ e $\bar{y}_2$ são as médias amostrais correspondentes aos dois grupos e $\hat{V}_0 = V(\bar{y})$ é a função de variância sob a hipótese nula[1].

Tabela 2.4
Expressões para as estatísticas de escore e de Wald.

Distribuição	ξ_{SR}	ξ_W
Normal	$\frac{m}{2\sigma^2}(\bar{y}_1 - \bar{y}_2)^2$	$\frac{m}{2\sigma^2}\hat{\beta}^2$
Poisson	$\frac{m}{2\bar{y}}(\bar{y}_1 - \bar{y}_2)^2$	$\frac{m\bar{y}_1\bar{y}_2}{(\bar{y}_1+\bar{y}_2)}\hat{\beta}^2$
Binomial	$\frac{2m}{y(2m-y)}(y_1 - y_2)^2$	$\frac{\hat{\beta}^2}{m}\frac{y_1(m-y_1)y_2(m-y_2)}{y_1(m-y_1)+y_2(m-y_2)}$
Gama	$\frac{\phi m}{2\bar{y}^2}(\bar{y}_1 - \bar{y}_2)^2$	$\frac{\phi m(\bar{y}_1\bar{y}_2)^2}{(\bar{y}_1^2+\bar{y}_2^2)}\hat{\beta}^2$
Normal inversa	$\frac{\phi m}{2\bar{y}^3}(\bar{y}_1 - \bar{y}_2)^2$	$\frac{\phi m(\bar{y}_1\bar{y}_2)^3}{(\bar{y}_1^3+\bar{y}_2^3)}\hat{\beta}^2$

[1]no caso binomial tomar $\bar{y}_i = y_i/m$ e $V(\bar{y}) = \bar{y}(1 - \bar{y})$

Similarmente, pode-se mostrar que a estatística de Wald fica dada por

$$\xi_W = \frac{\phi m \hat{\omega}_1 \hat{\omega}_2}{(\hat{\omega}_1 + \hat{\omega}_2)} \hat{\beta}^2, \tag{2.7}$$

em que $\hat{\beta}$ denota a estimativa de máxima verossimilhança de β. Na Tabela 2.4 são apresentadas as expressões das estatísticas ξ_{SR} e ξ_W para alguns casos da família exponencial.

2.7.4 Regressão linear simples

Supor agora um MLG com parte sistemática na forma linear simples

$$g(\mu_i) = \alpha + \beta x_i, \ i = 1, \ldots, n,$$

e as hipóteses $\mathrm{H}_0 : \beta = 0$ contra $\mathrm{H}_1 : \beta \neq 0$ com ϕ conhecido. Nesse caso obtém-se $R_j = (x_j \sum_{i=1}^n \omega_i - \sum_{i=1}^n \omega_i x_i)/\sum_{i=1}^n \omega_i$ e $\mathbf{R}^\top \mathbf{W} \mathbf{R} = \sum_{i=1}^n \omega_i R_i^2$. Consequentemente, $\hat{R}_{0j} = x_j - \bar{x}$ e $\hat{\mathbf{R}}_0^\top \hat{\mathbf{W}}_0 \hat{\mathbf{R}}_0 = \hat{\omega}_0 \sum_{i=1}^n (x_i - \bar{x})^2$. Aqui, também obtém-se $\hat{\mu}^0 = \bar{y}$.

A estatística de escore fica, portanto, dada por

$$\xi_{SR} = \frac{\phi}{\hat{V}_0} \frac{\{\sum_{i=1}^n x_i (y_i - \bar{y})\}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \tag{2.8}$$

em que $\hat{V}_0 = V(\bar{y})$.

Similarmente, obtém-se para a estatística de Wald

$$\xi_W = \phi \hat{\beta}^2 \sum_{i=1}^n \hat{\omega}_i \hat{R}_i^2, \tag{2.9}$$

em que $\hat{\beta}$ é a estimativa de β sob H_1.

2.7.5 Hipóteses restritas

Pode ser de interesse, em algumas situações práticas, testar hipóteses na forma de igualdades lineares, isto é, $\mathrm{H}_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ contra $\mathrm{H}_1 : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{0}$, em

que $\mathbf{C}$ é uma matriz $k \times p$ de posto linha completo e $k \leq p$. A estimativa de máxima verossimilhança sob a hipótese alternativa coincide com a estimativa de máxima verossimilhança irrestrita $\hat{\boldsymbol{\beta}}$. No entanto, obter a estimativa de máxima verossimilhança sob H_0 pode ser mais complexo, requerendo o uso de algum procedimento iterativo. Nyquist (1991) propõe um processo iterativo para a obtenção da estimativa de máxima verossimilhança em MLGs com parâmetros restritos na forma $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$. O processo iterativo é dado abaixo

$$\boldsymbol{\beta}_c^{(m+1)} = \tilde{\boldsymbol{\beta}}^{(m+1)} - (\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{C}^\top \{\mathbf{C}(\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{C}^\top\}^{-1} \mathbf{C} \tilde{\boldsymbol{\beta}}^{(m+1)},$$

$m = 0, 1, \ldots$, em que $\tilde{\boldsymbol{\beta}}^{(m+1)}$ é (1.5) avaliado na estimativa restrita $\boldsymbol{\beta}_c^{(m)}$. A matriz de variância-covariância assintótica de $\hat{\boldsymbol{\beta}}_c$ fica dada por

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_c) = \phi^{-1} (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} [\mathbf{I}_n - \mathbf{C}^\top \{\mathbf{C}(\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{C}^\top\}^{-1} \mathbf{C} (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1}].$$

Os testes estatísticos assumem formas similares aos testes do caso irrestrito. Em particular, quando ϕ é conhecido, o teste da razão de verossimilhanças fica dado por

$$\xi_{RV} = \phi\{\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - \mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})\},$$

em que $\hat{\boldsymbol{\mu}}^0$ denota aqui a estimativa de máxima verossimilhança de $\boldsymbol{\mu}$ sob $\mathrm{H}_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$. Já, o teste de escore, assume a forma

$$\xi_{SR} = \phi^{-1} \mathbf{U}_\beta(\hat{\boldsymbol{\beta}}_c)^\top (\mathbf{X}^\top \hat{\mathbf{W}}_0 \mathbf{X})^{-1} \mathbf{U}_\beta(\hat{\boldsymbol{\beta}}_c),$$

em que $\hat{\mathbf{W}}_0$ é aqui avaliado em $\hat{\boldsymbol{\beta}}_c$. Finalmente, o teste de Wald fica dado por

$$\begin{aligned} \xi_W &= [\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{0}]^\top [\hat{\mathrm{V}}\mathrm{ar}(\mathbf{C}\hat{\boldsymbol{\beta}})]^{-1} [\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{0}] \\ &= \phi \hat{\boldsymbol{\beta}}^\top \mathbf{C}^\top [\mathbf{C}(\mathbf{X}^\top \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} \mathbf{C}\hat{\boldsymbol{\beta}}. \end{aligned}$$

Sob H_0 e para amostras grandes, as estatísticas ξ_{RV}, ξ_W e ξ_{SR} seguem uma distribuição χ^2_k. A distribuição nula assintótica dos testes acima para o caso

$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ contra $H_1 - H_0$, em que $H_1 : \mathbf{C}\boldsymbol{\beta} \geq \mathbf{0}$, é uma mistura de distribuições do tipo qui-quadrado. Fahrmeir e Klinger (1994) discutem esse tipo de teste em MLGs.

2.7.6 Bandas de confiança

Uma banda assintótica de confiança de coeficiente $1-\alpha$ pode ser também construída para $\mu(\mathbf{z}) = g^{-1}(\mathbf{z}\top\boldsymbol{\beta}), \forall \mathbf{z} \in I\!R^p$ (Piegorsch e Casella, 1988) generalizando os resultados da Seção 1.6. Assintoticamente tem-se que $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim \mathrm{N}_p(\mathbf{0}, \phi^{-1}(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1})$. Logo, uma banda assintótica de confiança de coeficiente $1-\alpha$ para o preditor linear $\mathbf{z}^\top\boldsymbol{\beta}, \forall \mathbf{z} \in I\!R^p$, fica dada por

$$\mathbf{z}^\top\hat{\boldsymbol{\beta}} \pm \sqrt{\phi^{-1}c_\alpha}\{\mathbf{z}^\top(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{z}\}^{\frac{1}{2}}, \quad \forall \mathbf{z} \in I\!R^p,$$

em que c_α é tal que $Pr\{\chi^2_p \leq c_\alpha\} = 1-\alpha$. Aplicando a transformação $g^{-1}(\cdot)$ tem-se, equivalentemente, uma banda assintótica de confiança de coeficiente $1-\alpha$ para $\mu(\mathbf{z})$, dada por

$$g^{-1}[\mathbf{z}^\top\hat{\boldsymbol{\beta}} \pm \sqrt{\phi^{-1}c_\alpha}\{\mathbf{z}^\top(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{z}\}^{\frac{1}{2}}] \quad \forall \mathbf{z} \in I\!R^p.$$

Lembrando que $\mathbf{z}$ é um vetor $p \times 1$ que varia livremente no $I\!R^p$, enquanto $\mathbf{X}$ é uma matriz fixa com os valores das variáveis explicativas. As quantidades $\mathbf{W}$ e ϕ devem ser estimadas consistentemente.

2.8 Técnicas de diagnóstico

2.8.1 Pontos de alavanca

Como já foi mencionado na Seção 1.7.1 a ideia principal que está por trás do conceito de ponto de alavanca é de avaliar a influência de y_i sobre o próprio valor ajustado $\hat{y}_i$. Essa influência pode ser bem representada pela derivada

$\partial\hat{y}_i/\partial y_i$ que coincide, como foi visto na Seção 1.7.1, com h_{ii} no caso normal linear. Wei et al.(1998) propuseram uma forma geral para a obtenção da matrix $(\partial\hat{\mathbf{y}}/\partial\mathbf{y}^\top)_{n\times n}$ quando a resposta é contínua e que pode ser aplicada em diversas situações de estimação. No caso de MLGs, para ϕ conhecido, a matriz $\partial\hat{\mathbf{y}}/\partial\mathbf{y}^\top$ pode ser obtida da forma geral

$$\widehat{\mathrm{GL}} = \frac{\partial\hat{\mathbf{y}}}{\partial\mathbf{y}^\top} = \{\mathbf{D}_\beta(-\ddot{\mathbf{L}}_{\beta\beta})^{-1}\ddot{\mathbf{L}}_{\beta y}\}|_{\hat{\beta}},$$

em que $\mathbf{D}_\beta = \partial\boldsymbol{\mu}/\partial\boldsymbol{\beta}$, $\ddot{\mathbf{L}}_{\beta\beta} = \partial^2\mathrm{L}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top$ e $\ddot{\mathbf{L}}_{\beta y} = \partial^2\mathrm{L}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\mathbf{y}^\top$. Tem-se que

$$\mathbf{D}_\beta = \mathbf{N}\mathbf{X} \;\text{ e }\; \ddot{\mathbf{L}}_{\beta y} = \phi\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{N},$$

em que $\mathbf{N} = \mathrm{diag}\{d\mu_1/d\eta_1, \ldots, d\mu_n/d\eta_n\}$. Substituindo $-\ddot{\mathbf{L}}_{\beta\beta}$ pelo seu valor esperado $\phi(\mathbf{X}^\top\mathbf{W}\mathbf{X})$, obtém-se aproximadamente

$$\widehat{\mathrm{GL}} \cong \hat{\mathbf{N}}\mathbf{X}(\mathbf{X}^\top\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^\top\hat{\mathbf{V}}^{-1}\hat{\mathbf{N}}.$$

Assim, o elemento $\widehat{\mathrm{GL}}_{ii}$ pode ser expresso na forma

$$\widehat{\mathrm{GL}}_{ii} \cong \hat{\omega}_i\mathbf{x}_i^\top(\mathbf{X}^\top\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{x}_i,$$

em que $\omega_i = (d\mu_i/d\eta_i)^2/V_i$. Em particular, para ligação canônica em que $-\ddot{\mathbf{L}}_{\beta\beta} = \phi(\mathbf{X}^\top\mathbf{V}\mathbf{X})$ obtém-se exatamente $\widehat{\mathrm{GL}} = \hat{\mathbf{V}}\mathbf{X}(\mathbf{X}^\top\hat{\mathbf{V}}\mathbf{X})^{-1}\mathbf{X}^\top$.

Outra definição de pontos de alavanca que tem sido utilizada na classe dos MLGs, embora não coincida exatamente com a expressão acima, exceto no caso de resposta contínua e ligação canônica, é construída fazendo uma analogia entre a solução de máxima verossimilhança para $\hat{\boldsymbol{\beta}}$ num MLG e a solução de mínimos quadrados de uma regressão normal linear ponderada. Considerando a expressão para $\hat{\boldsymbol{\beta}}$ obtida na convergência do processo iterativo dado em (2.5), tem-se que

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^\top\hat{\mathbf{W}}\hat{\mathbf{z}},$$

com $\hat{\mathbf{z}} = \hat{\boldsymbol{\eta}} + \hat{\mathbf{W}}^{-\frac{1}{2}}\hat{\mathbf{V}}^{-\frac{1}{2}}(\mathbf{y} - \hat{\boldsymbol{\mu}})$. Portanto, $\hat{\boldsymbol{\beta}}$ pode ser interpretado como sendo a solução de mínimos quadrados da regressão linear de $\hat{\mathbf{W}}^{\frac{1}{2}}\hat{\mathbf{z}}$ contra as colunas de $\hat{\mathbf{W}}^{\frac{1}{2}}\mathbf{X}$. A matriz de projeção da solução de mínimos quadrados da regressão linear de $\hat{\mathbf{z}}$ contra $\mathbf{X}$ com pesos $\hat{\mathbf{W}}$ fica dada por

$$\hat{\mathbf{H}} = \hat{\mathbf{W}}^{\frac{1}{2}}\mathbf{X}(\mathbf{X}^\top\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^\top\hat{\mathbf{W}}^{\frac{1}{2}},$$

que sugere a utilização dos elementos $\hat{h}_{ii}$ da diagonal principal de $\hat{\mathbf{H}}$ para detectar a presença de pontos de alavanca nesse modelo de regressão normal linear ponderada. Essa extensão para MLGs foi proposta por Pregibon (1981). Pode-se verificar facilmente que $\hat{h}_{ii} = \widehat{\mathrm{GL}}_{ii}$, ou seja, para grandes amostras $\widehat{\mathrm{GL}}$ e $\hat{\mathbf{H}}$ coincidem. No caso de ligação canônica essa igualdade vale para qualquer tamanho amostral. Como em geral $\hat{h}_{ii}$ depende de $\hat{\mu}_{ii}$ é sugerido para detectar pontos de alavanca o gráfico de $\hat{h}_{ii}$ contra os valores ajustados.

Moolgavkar et al.(1984) estendem a proposta de Pregibon para modelos não lineares e sugerem o uso dos elementos da diagonal principal da matriz de projeção no plano tangente à solução de máxima verossimilhança $\boldsymbol{\mu}(\hat{\boldsymbol{\beta}})$ para detectar pontos de alavanca. Hosmer et al. (2013) mostram, contudo, que o uso da diagonal principal da matriz de projeção $\hat{\mathbf{H}}$ deve ser feito com algum cuidado em regressão logística e que as interpretações são diferentes daquelas do caso normal linear (vice Seção 4.6.8).

2.8.2 Resíduos

A definição de um resíduo studentizado para os MLGs pode ser feita analogamente à regressão normal. Todavia, não necessariamente as propriedades continuam valendo. Assim, torna-se importante a definição de outros tipos de resíduo cujas propriedades sejam conhecidas ou pelo menos estejam mais próximas das propriedades de t_i^* definido na Seção 1.7.3.

Uma primeira proposta seria considerar o resíduo ordinário da solução de mínimos quadrados da regressão linear ponderada de $\hat{\mathbf{z}}$ contra $\mathbf{X}$, que é definido por

$$\mathbf{r}^* = \hat{\mathbf{W}}^{\frac{1}{2}}(\hat{\mathbf{z}} - \hat{\boldsymbol{\eta}}) = \hat{\mathbf{V}}^{-\frac{1}{2}}(\mathbf{y} - \hat{\boldsymbol{\mu}}).$$

Assumindo que $\text{Var}(\mathbf{z}) \cong \hat{\mathbf{W}}^{-1}\phi^{-1}$, tem-se aproximadamente

$$\text{Var}(\mathbf{r}^*) \cong \phi^{-1}(\mathbf{I}_n - \hat{\mathbf{H}}).$$

Logo, pode-se definir o resíduo padronizado

$$t_{S_i} = \frac{\sqrt{\phi}(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i(1 - \hat{h}_{ii})}},$$

em que h_{ii} é o i-ésimo elemento da diagonal principal da matriz $\mathbf{H}$. Fica fácil mostrar que $\mathbf{r}^* = (\mathbf{I}_n - \hat{\mathbf{H}})\hat{\mathbf{W}}^{\frac{1}{2}}\hat{\mathbf{z}}$, isto é, $\hat{\mathbf{H}}$ desempenha o papel de matriz de projeção ortogonal local, como na regressão normal linear em que $\mathbf{W}$ é identidade.

No entanto, na prática, $\hat{\boldsymbol{\eta}}$ não é fixo nem conhecido, bem como $\mathbf{z}$ não segue distribuição normal. Uma implicação disso é que as propriedades de t_i^* não são mais verificadas para t_{S_i}. Williams (1984) mostra através de estudos de Monte Carlo que a distribuição de t_{S_i} é em geral assimétrica, mesmo para grandes amostras.

Outros resíduos cujas distribuições poderiam estar mais próximas da normalidade têm sido sugeridos para os MLGs. Por exemplo, o resíduo de Anscombe

$$t_{A_i} = \frac{\sqrt{\phi}\{\psi(y_i) - \psi(\hat{\mu}_i)\}}{\psi'(\hat{\mu}_i)\sqrt{\hat{V}(\hat{\mu}_i)}},$$

em que $\psi(\cdot)$ é uma transformação utilizada para normalizar a distribuição de

Y. Para os MLGs essa transformação é definida por

$$\psi(\mu) = \int_0^{\mu} V^{-\frac{1}{3}}(t)dt.$$

Em particular, para os principais MLGs a transformação $\psi(\mu)$ é descrita na tabela dada abaixo.

	Distribuição				
	Normal	Binomial	Poisson	Gama	N. Inversa
$\psi(\mu)$	μ	$\int_0^{\mu} t^{-\frac{1}{3}}(1-t)^{-\frac{1}{3}}dt$	$\frac{3}{2}\mu^{\frac{2}{3}}$	$3\mu^{\frac{1}{3}}$	$\log(\mu)$

Contudo, um dos resíduos mais utilizados MLGs é definido a partir dos componentes da função desvio. A versão padronizada (ver McCullagh, 1987; Davison e Gigli, 1989) é a seguinte:

$$t_{D_i} = \frac{d^*(y_i;\hat{\mu}_i)}{\sqrt{1-\hat{h}_{ii}}} = \frac{\sqrt{\phi}d(y_i;\hat{\mu}_i)}{\sqrt{1-\hat{h}_{ii}}},$$

em que $d(y_i;\hat{\mu}_i) = \pm\sqrt{2}\{y_i(\tilde{\theta}_i - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\tilde{\theta}_i))\}^{\frac{1}{2}}$. O sinal de $d(y_i;\hat{\mu}_i)$ é o mesmo de $y_i - \hat{\mu}_i$. Williams (1984) verificou através de simulações que a distribuição de t_{D_i} tende a estar mais próxima da normalidade do que as distribuições dos demais resíduos. McCullagh (1987, p. 214) mostra para os MLGs que a distribuição de probabilidade de

$$\frac{d^*(Y_i;\mu_i) + \rho_{3i}/6}{\sqrt{1 + (14\rho_{3i}^2 - 9\rho_{4i})/36}}$$

é aproximadamente $\mathrm{N}(0,1)$, em que ρ_{3i} e ρ_{4i} são os coeficientes de assimetria e curtose de $\partial \mathrm{L}(\eta_i)/\partial\eta_i$, respectivamente, e $d^*(Y_i;\mu_i)$ é o i-ésimo componente do desvio $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}})$ avaliado no parâmetro verdadeiro. Pode-se mostrar usando resultados de Cox e Snell (1968) que $\mathrm{E}\{d^*(Y_i;\mu_i)\} = 0$ e $\mathrm{Var}\{d^*(Y_i;\mu_i)\} = 1 - h_{ii}$, em que os termos negligenciados são de $O(n^{-1})$. Esses resultados reforçam o uso da padronização $\sqrt{1-\hat{h}_{ii}}$ para $d^*(y_i;\hat{\mu}_i)$. Na prática deve-se substituir ϕ por uma estimativa consistente.

Um quarto resíduo foi definido por Williams (1987) e pode ser interpretado como uma média ponderada entre t_{S_i} e t_{D_i},

$$t_{G_i} = \text{sinal}(y_i - \hat{\mu}_i)\{(1 - \hat{h}_{ii})t_{D_i}^2 + \hat{h}_{ii}t_{S_i}^2\}^{\frac{1}{2}}.$$

Williams (1987) verificou através de simulações e para alguns MLGs que t_{G_i} tem esperança ligeiramente diferente de zero, variância excedendo um, assimetria desprezível e alguma curtose.

O `R` solta os resíduos $d_i = d(y_i; \hat{\mu}_i)$ e $\hat{r}_{P_i}$ sem o termo $\sqrt{\phi}$. Precisa, para padronizá-los, calcular os correspondentes $\hat{h}'_{ii}s$ bem como extrair $\hat{\phi}$ nos casos em que $\phi \neq 1$. Inicialmente, é ilustrado como calcular $\hat{h}_{ii}$. Supor um modelo com duas covariáveis e dois fatores e que os resultados do ajuste são armazenados em `fit.model`. A matriz $\mathbf{X}$ é obtida com um dos comandos abaixo

```
X = model.matrix( ~ cov1 + cov2 + A + B)
X = model.matrix(fit.model).
```

Em `V` pode-se armazenar a matriz $\hat{\mathbf{V}}$. Os elementos da diagonal principal de `V` devem ser obtidos dos valores ajustados do modelo, os quais por sua vez são extraídos através do comando `fitted(fit.model)`. Como exemplo, a matriz com as funções de variância estimadas seria obtida para um modelo de Poisson da forma seguinte:

```
V = fitted(fit.model)
V = diag(V).
```

Em particular, a matriz $\hat{\mathbf{W}}$ também depende dos valores ajustados, no entanto, como é a matriz de pesos, pode ser obtida diretamente fazendo

```
w = fit.model$weights
W = diag(w).
```

Assim, uma vez obtida a matriz $\hat{\mathbf{W}}$ pode-se obter os elementos $\hat{h}_{ii}$ com os comandos

```
H = solve(t(X)%*%W%*%X)
H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
h = diag(H).
```

Armazenando em `fit` a estimativa $\hat{\phi}$ (o `R` solta $\hat{\phi}^{-1}$), os componentes do desvio e os resíduos studentizados são obtidos da seguinte maneira:

```
rd = resid(fit.model, type= ‘‘deviance’’)
td = rd*sqrt(fi/(1-h))
rp = resid(fit.model, type= ‘‘pearson’’)
rp = sqrt(fi)*rp
ts = rp/sqrt(1 - h).
```

Lembrando que para ligações canônicas $\mathbf{W}$ e $\mathbf{V}$ coincidem.

Por fim, tem-se o resíduos quantílico (Dunn e Smyth, 1996) que é definido para variáveis contínuas por

$$r_{q_i} = \Phi^{-1}\{F(y_i; \hat{\boldsymbol{\theta}})\},$$

em que $\Phi(\cdot)$ e $F(y_i; \boldsymbol{\theta})$ denotam, respectivamente, as funções de distribuição acumuladas da $N(0,1)$ e da distribuição postulada para a resposta, $i = 1, \ldots, n$. Para n grande os resíduos $r_{q_1}, \ldots, r_{q_n}$ são independentes e igualmente distribuídos $N(0,1)$. Assim, o gráfico entre os quantis amostrais $r_{q_{(1)}} \leq \cdots \leq r_{q_{(n)}}$ contra os quantis teóricos da normal padrão é recomendado para avaliar afastamentos da distribuição postulada para a resposta bem como a presença de observações aberrantes. Esse resíduo é estendido para o caso discreto, contudo o resíduo não é único, e a sugestão é trabalhar com resíduos aleatorizados.

O resíduo quantílico é disponibilizado na biblioteca `GAMLSS` do R (ver, por exemplo, Stasinopoulos et al., 2017) através dos comandos

```
require(gamlss)
plot(ajuste).
```

Aqui `ajuste` é o nome do objeto referente ao ajuste do modelo. Além desse painel gráfico, o GAMLSS também disponibiliza o `worm plot` que é o gráfico entre $r_{q_{(i)}} - \mathrm{E}(Z_{(i)})$ contra $\mathrm{E}(Z_{(i)})$. Esse gráfico pode ser interpretado como um refinamento do gráfico normal de probabilidades, podendo ser acionado para variáveis contínuas através do comando

```
wp(ajuste).
```

No caso de variáveis discretas, a sugestão é gerar m gráficos, que são avaliados conjuntamente. Por exemplo para $m = 8$ o gráfico pode ser realizado através do comando

```
rqres.plot(ajuste, howmany=8, type=‘‘wp’’),
```

Ou construir um único gráfico com todos os resíduos, por exemplo $m = 50$, cujo comando é dado por

```
rqres.plot(ajuste, howmany=50, type=‘‘all’’).
```

A construção de bandas empíricas de confiança para o gráfico normal de probabilidades com o resíduo quantílico seria recomendada no caso de amostras pequenas e moderadas, uma vez que os resíduos são correlacionados. Embora o resíduo quantílico tenha uma distribuição assintótica conhecida, sob o modelo postulado, tendo portanto aplicação direta em modelagem de regressão, o resíduo componente do desvio pode continuar sendo aplicado de forma complementar por tratar-se de um resíduo condicional. Ou seja, tem-se o componente do desvio para a localização fixando a dispersão, e de forma similar pode-se ter o resíduo componente do desvio para a dispersão fixando a localização. Na Seção 4.9.2 é derivado o resíduo componente do desvio para avaliar a qualidade do ajuste no componente de dispersão em MLGs. Essa ideia se estende para outros modelos de regressão em que há mais de dois tipos de parâmetros para serem modelados.

2.8.3 Influência

A ideia de influência é avaliar o impacto de perturbações no modelo ou dados nas estimativas dos parâmetros, através de alguma medida de influência apropriada. Supondo ϕ conhecido, a medida de influência mais utilizada é denominada afastamento pela verossimilhança (Cook, 1986), sendo definda por

$$\mathrm{LD}(\boldsymbol{\delta}) = 2\{\mathrm{L}(\hat{\boldsymbol{\theta}}) - \mathrm{L}(\hat{\boldsymbol{\theta}}_\delta)\},$$

em que $\hat{\boldsymbol{\theta}}_\delta$ denota a estimativa de máxima verossimilhança sob o modelo com perturbação $\boldsymbol{\delta}$. Contudo, o caso mais usual de perturbação é avaliar o impacto nas estimativas dos parâmetros eliminando individualmetne cada observação. Nesse caso, o afastamento pela verossimilhança fica definido por

$$\mathrm{LD}_i = 2\{\mathrm{L}(\hat{\boldsymbol{\beta}}) - \mathrm{L}(\hat{\boldsymbol{\beta}}_{(i)})\},$$

em que $\boldsymbol{\beta}_{(i)}$) denota a estimativa de máxima verossimilhança de $\boldsymbol{\beta}$ quando a i-ésima observação é eliminada. O caso mais geral de perturbação será discutido mais à frente em influência local.

Não sendo possível obter uma forma analítica para LD_i, é usual utilizar a segunda aproximação por série de Taylor de $\mathrm{L}(\boldsymbol{\beta})$ em torno de $\hat{\boldsymbol{\beta}}$, obtendo-se $\mathrm{L}(\boldsymbol{\beta}) \cong \mathrm{L}(\hat{\boldsymbol{\beta}}) + \frac{1}{2}(\beta - \hat{\beta})^\top\{-\ddot{\mathbf{L}}_{\beta\beta}(\hat{\boldsymbol{\beta}})\}(\beta - \hat{\beta})$. Essa expansão leva ao seguinte resultado:

$$\mathrm{LD}_i \cong (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top\{-\ddot{\mathbf{L}}_{\beta\beta}(\hat{\boldsymbol{\beta}})\}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}).$$

Substituindo $-\ddot{\mathbf{L}}_{\beta\beta}(\hat{\boldsymbol{\beta}})$ pelo correspondente valor esperado e $\boldsymbol{\beta}$ por $\hat{\boldsymbol{\beta}}_{(i)}$, obtém-se

$$\mathrm{LD}_i \cong \phi(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})^\top(\mathbf{X}^\top\hat{\mathbf{W}}\mathbf{X})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}). \tag{2.10}$$

Assim, tem-se uma boa aproximação para LD_i quando $\mathrm{L}(\boldsymbol{\beta})$ for aproximadamente quadrática em torno de $\hat{\boldsymbol{\beta}}$.

Como em geral não é possível obter uma forma fechada para $\hat{\boldsymbol{\beta}}_{(i)}$, a aproximação de um passo tem sido utilizada (ver, por exemplo, Cook e Weisberg, 1982), que consiste em tomar a primeira iteração do processo iterativo pelo método escore de Fisher quando o mesmo é iniciado em $\hat{\boldsymbol{\beta}}$.

Essa aproximação, introduzida por Pregibon (1981), é dada por

$$\hat{\boldsymbol{\beta}}^{1}_{(i)} = \hat{\boldsymbol{\beta}} + (\mathbf{X}^\top \boldsymbol{\Delta} \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Delta} \hat{\mathbf{W}}^{\frac{1}{2}} \hat{\mathbf{V}}^{-\frac{1}{2}} (\mathbf{y} - \hat{\boldsymbol{\mu}}),$$

em que $\boldsymbol{\Delta} = \text{diag}\{\delta_1, \ldots, \delta_n\}$ com $\delta_i = 0$ e $\delta_j = 1$ para $j \neq i$. Após algumas manipulações algébricas obtém-se

$$\hat{\boldsymbol{\beta}}^{1}_{(i)} = \hat{\boldsymbol{\beta}} - \frac{\hat{r}_{P_i} \sqrt{\hat{\omega}_i \phi^{-1}}}{(1 - \hat{h}_{ii})} (\mathbf{X}^\top \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i \tag{2.11}$$

e, finalmente, substituindo a expressão acima em (2.17) tem-se que

$$\text{LD}_i \cong \left\{ \frac{\hat{h}_{ii}}{(1 - \hat{h}_{ii})} \right\} t^2_{S_i}.$$

A distância de Cook aproximada fica facilmente obtida com o comando

`LD = h*(ts^2)/(1 - h).`

A validade da aproximação de um passo tem sido investigada por alguns pesquisadores. A constatação é que a mesma em geral subestima o verdadeiro valor de LD_i, no entanto é suficiente para chamar a atenção dos pontos influentes.

2.8.4 Influência local

Um dos métodos mais modernos de diagnóstico foi proposto por Cook (1986). A ideia básica consiste em estudar o comportamento de alguma medida particular de influência segundo pequenas perturbações (`influência local`) nos dados ou no modelo. Isto é, verificar a existência de pontos que sob

modificações modestas no modelo causam variações desproporcionais nos resultados.

Pode-se, por exemplo, querer avaliar a influência que pequenas mudanças nas variâncias das observações causam nas estimativas dos parâmetros. Nesse caso, pode-se utilizar a distância de Cook como medida de referência. Por outro lado, se o interesse é estudar a influência local das observações no ajuste, a sugestão de Cook é perturbar as covariáveis ou a variável resposta e utilizar alguma medida adequada para quantificar a influência das observações. Como ilustração, supor que uma variável explicativa que representa uma distância particular é perturbada localmente e detecta-se através de uma medida de influência que pontos com distâncias altas produzem variações acentuadas na medida adotada. Isso sugere que a variável explicativa sob estudo é bastante sensível para valores altos, podendo não ser uma boa preditora nesses casos. A seguir é descrito o procedimento de influência local.

Curvatura normal

Para formalizar o método de influência local denote por $\mathrm{L}(\boldsymbol{\theta})$ o logaritmo da função de verossimilhança do modelo postulado e $\boldsymbol{\theta}$ um vetor r-dimensional. No caso de MLGs pode-se ter $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$ e $r = p+1$ ou simplesmente $\boldsymbol{\theta} = \boldsymbol{\beta}$ quando ϕ for conhecido. Seja $\boldsymbol{\delta}$ um vetor $q \times 1$ de perturbações, restritas a um conjunto aberto $\boldsymbol{\Omega} \subset I\!R^q$. Em geral tem-se $q = n$. As perturbações são feitas no logaritmo da verossimilhança de modo que o mesmo assume a forma $\mathrm{L}(\boldsymbol{\theta}|\boldsymbol{\delta})$. Denotando o vetor de não perturbação por $\boldsymbol{\delta}_0$, tem-se que $\mathrm{L}(\boldsymbol{\theta}|\boldsymbol{\delta}_0) = \mathrm{L}(\boldsymbol{\theta})$. A fim de verificar a influência das perturbações na estimativa de máxima verossimilhança $\hat{\boldsymbol{\theta}}$, considere o afastamento pela verossimilhança

$$\mathrm{LD}(\boldsymbol{\delta}) = 2\{\mathrm{L}(\hat{\boldsymbol{\theta}}) - \mathrm{L}(\hat{\boldsymbol{\theta}}_\delta)\},$$

em que $\hat{\boldsymbol{\theta}}_\delta$ denota a estimativa de máxima verossimilhança sob o modelo $\mathrm{L}(\boldsymbol{\theta}|\boldsymbol{\delta})$. Com a definição acima tem-se que $\mathrm{LD}(\boldsymbol{\delta}) \geq 0$.

A ideia de influência local consiste basicamente em estudar o comportamento da função $\mathrm{LD}(\boldsymbol{\delta})$ em torno de $\boldsymbol{\delta}_0$. O procedimento procura selecionar uma direção unitária $\boldsymbol{\ell}$, $\parallel \boldsymbol{\ell} \parallel = 1$, e então estudar o gráfico de $\mathrm{LD}(\boldsymbol{\delta}_0 + a\boldsymbol{\ell})$ contra a, em que $a \in I\!R$. Esse gráfico é conhecido como linha projetada. Em particular, tem-se que $\mathrm{LD}(\boldsymbol{\delta}_0) = 0$, assim $\mathrm{LD}(\boldsymbol{\delta}_0 + a\boldsymbol{\ell})$ tem um mínimo local em $a = 0$. Cada linha projetada pode ser caracterizada por uma curvatura normal $\mathrm{C}_\ell(\boldsymbol{\theta})$ em torno de $a = 0$. Essa curvatura é interpretada como sendo o inverso do raio do melhor círculo ajustado em $a = 0$. Uma sugestão é considerar a direção $\boldsymbol{\ell}_{max}$ que corresponde à maior curvatura denotada por $\mathrm{C}_{\ell_{\max}}$. Por exemplo, o gráfico de $|\boldsymbol{\ell}_{\max}|$ contra a ordem das observações pode revelar quais observações que sob pequenas perturbações exercem uma influência desproporcional em $\mathrm{LD}(\boldsymbol{\delta})$. Cook(1986) usa conceitos de geometria diferencial para mostrar que a curvatura normal na direção $\boldsymbol{\ell}$ assume a forma

$$\mathrm{C}_\ell(\boldsymbol{\theta}) = 2|\boldsymbol{\ell}^\top \boldsymbol{\Delta}^\top \ddot{\mathbf{L}}^{-1}_{\hat{\theta}\hat{\theta}} \boldsymbol{\Delta}\boldsymbol{\ell}|,$$

em que $-\ddot{\mathbf{L}}_{\hat{\theta}\hat{\theta}}$ é a matriz de informação observada enquanto $\boldsymbol{\Delta}$ é uma matriz $r \times q$ com elementos $\Delta_{ij} = \partial^2 \mathrm{L}(\boldsymbol{\theta}|\boldsymbol{\delta})/\partial\theta_i\partial\delta_j$, avaliados em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\delta} = \boldsymbol{\delta}_0$, $i = 1, \ldots, r$ e $j = 1, \ldots, q$.

Tem-se que o máximo de $\boldsymbol{\ell}^\top \mathbf{B} \boldsymbol{\ell}$, em que $\mathbf{B} = \boldsymbol{\Delta}^\top (-\ddot{\mathbf{L}}_{\hat{\theta}\hat{\theta}})^{-1} \boldsymbol{\Delta}$, corresponde ao maior autovalor (em valor absoluto) de $\mathbf{B}$. Portanto, $\mathrm{C}_{\ell_{\max}}$ corresponde ao maior autovalor da matriz $\mathbf{B}$ e $\boldsymbol{\ell}_{max}$ denota o autovetor correspondente.

Assim, o gráfico de $|\boldsymbol{\ell}_{\max}|$ contra a ordem das observações pode revelar aqueles pontos com maior influência na vizinhança de $\mathrm{LD}(\delta_0)$. Tais pontos podem ser responsáveis por mudanças substanciais nas estimativas dos parâmetros sob pequenas perturbações no modelo ou nos dados. Seria, portanto, prudente olhar com mais cuidado esses pontos a fim de entender me-

lhor a influência dos mesmos e consequentemente tentar propor uma forma segura de usar o modelo ajustado. Quando $\mathrm{C}_{\ell_{\max}}$ não for muito maior do que o segundo autovalor, pode ser informativo olhar também os componentes do segundo autovetor. É provável, nesse caso, que o segundo autovetor destaque algum tipo de influência particular das observações nas estimativas. O maior autovalor da matriz $\mathbf{B}$ pode ser obtido pelo comando abaixo

```
Cmax = eigen(B)$val[1].
```

De forma similar, o autovetor correspondente padronizado e em valor absoluto é obtido com os comandos

```
lmax = eigen(B)$vec[,1]
lmax = abs(lmax).
```

Gráficos alternativos, tais como de C_{ℓ_i} contra a ordem das observações, em que $\boldsymbol{\ell}_i$ denota um vetor $n \times 1$ de zeros com um na i-ésima posição têm sido sugeridos (ver, por exemplo, Lesaffre e Verbeke, 1998). Nesse caso deve-se padronizar $\mathrm{C}_i = \mathrm{C}_i / \sum_{j=1}^{n} \mathrm{C}_j$. Uma sugestão é olhar com mais atenção aqueles pontos tais que $\mathrm{C}_i > \bar{\mathrm{C}} + k\mathrm{DP}\{\mathrm{C}_i)\}$, para $k = 1, 2, 3$ dependendo do tamanho amostral, em que $\bar{\mathrm{C}} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{C}_i$.

Por outro lado, se o interesse está num subvetor $\boldsymbol{\theta}_1$ de $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, então a curvatura normal na direção $\boldsymbol{\ell}$ fica dada por

$$\mathrm{C}_\ell(\boldsymbol{\theta}_1) = 2|\boldsymbol{\ell}^\top \boldsymbol{\Delta}^\top (\ddot{\mathbf{L}}_{\hat{\theta}\hat{\theta}}^{-1} - \mathbf{B}_1)\boldsymbol{\Delta}\boldsymbol{\ell}|,$$

sendo

$$\mathbf{B}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathbf{L}}_{\hat{\theta}_2\hat{\theta}_2}^{-1} \end{pmatrix},$$

com $-\ddot{\mathbf{L}}_{\hat{\theta}_2\hat{\theta}_2}$ denotando a matriz de informação observada para $\boldsymbol{\theta}_2$. O gráfico do maior autovetor de $\boldsymbol{\Delta}^\top(\ddot{\mathbf{L}}_{\hat{\theta}\hat{\theta}}^{-1} - \mathbf{B}_1)\boldsymbol{\Delta}$ contra a ordem das observações pode revelar os pontos com maior influência local em $\hat{\boldsymbol{\theta}}_1$.

Poon e Poon (1999) propoêm uma variação da medida de curvatura normal de Cook, a qual denominam curvatura normal conformal, que é invariante com mudanças de escala e é definida no intervalo unitário. Vários gráficos novos de influência são propostas, em particular uma forma de agregar as direções de maior curvatura em medidas resumo de influência.

Ponderação de casos

Para ilustrar uma aplicação particular considere o modelo normal linear com σ^2 conhecido e esquema de perturbação ponderação de casos, em que

$$\mathrm{L}(\boldsymbol{\beta}|\boldsymbol{\delta}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\delta_i(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2$$

com $0 \leq \delta_i \leq 1$. A matriz $\boldsymbol{\Delta}$ nesse caso fica dada por $\mathbf{X}^\top\mathrm{D}(\mathbf{r})/\sigma^2$ em que $\mathrm{D}(\mathbf{r}) = \mathrm{diag}\{r_1, \ldots, r_n\}$ com $r_i = y_i - \hat{y}_i$. Logo, desde que $\ddot{\mathbf{L}}_{\beta\beta} = -\sigma^{-2}(\mathbf{X}^\top\mathbf{X})$ a curvatura normal na direção unitária $\boldsymbol{\ell}$ fica dada por

$$\mathrm{C}_\ell(\boldsymbol{\beta}) = \frac{2}{\sigma^2}|\boldsymbol{\ell}^\top\mathrm{D}(\mathbf{r})\mathbf{H}\mathrm{D}(\mathbf{r})\boldsymbol{\ell}|,$$

com $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$. Portanto, $\boldsymbol{\ell}_{\max}$ é o autovetor correspondente ao maior autovalor (em valor absoluto) da matriz $\mathbf{B} = \mathrm{D}(\mathbf{r})\mathbf{H}\mathrm{D}(\mathbf{r})$. Se for considerada a direção $\boldsymbol{\ell}_i$ correspondente à i-ésima observação, a curvatura normal assume a forma simplificada $\mathrm{C}_i = \frac{2}{\sigma^2}h_{ii}r_i^2$. Os gráficos de índices de $\boldsymbol{\ell}_{max}$ e C_i podem revelar aquelas observações mais sensíveis ao esquema de perturbação adotado.

Cálculos similares para σ^2 desconhecido levam ao seguinte $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1^\top, \boldsymbol{\Delta}_2^\top)^\top$ em que $\boldsymbol{\Delta}_1 = \mathbf{X}^\top\mathrm{D}(\mathbf{r})/\hat{\sigma}^2$ e $\boldsymbol{\Delta}_2 = \mathbf{r}^{(2)\top}/2\hat{\sigma}^4$ com $\mathbf{r}^{(2)\top} = (r_1^2, \ldots, r_n^2)$ e $-\ddot{\mathbf{L}}_{\hat{\theta}\hat{\theta}} = \mathrm{diag}\{\mathbf{X}^\top\mathbf{X}/\hat{\sigma}^2, n/2\hat{\sigma}^4\}$. Logo, a curvatura normal na direção unitária $\boldsymbol{\ell}$ fica dada por

$$\mathrm{C}_\ell(\boldsymbol{\theta}) = \frac{2}{\hat{\sigma}^2}|\boldsymbol{\ell}^\top\{\mathrm{D}(\mathbf{r})\mathbf{H}\mathrm{D}(\mathbf{r}) + \mathbf{r}^{(2)}\mathbf{r}^{(2)\top}/2n\hat{\sigma}^2\}\boldsymbol{\ell}|.$$

Quando o interesse é verificar a influência local das observações na estimativa de um coeficiente particular β_1 deve-se considerar a curvatura normal $\mathrm{C}_\ell(\beta_1) = 2|\boldsymbol{\ell}^\top \mathbf{B}\boldsymbol{\ell}|$, em que

$$\mathbf{B} = \mathrm{D}(\mathbf{r})\mathbf{X}\{(\mathbf{X}^\top\mathbf{X})^{-1} - \mathbf{B}_1\}\mathbf{X}^\top\mathrm{D}(\mathbf{r})$$

sendo $\mathbf{B}_1 = \mathrm{diag}\{\mathbf{0}, (\mathbf{X}_2^\top\mathbf{X}_2)^{-1}\}$ com $\mathbf{X}_2$ saindo da partição $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$. Aqui $\mathbf{X}_1$ é um vetor $n \times 1$ correspondente à variável explicativa sob estudo e $\mathbf{X}_2$ é uma matriz $n \times (p-1)$ correspondente às demais variáveis explicativas. Cook (1986) mostra que $\boldsymbol{\ell}_{\max}$, nesse caso, assume a forma

$$\boldsymbol{\ell}_{\max}^\top = \left(\frac{v_1 r_1}{\sqrt{\mathrm{C}_{\ell_{\max}}}}, \ldots, \frac{v_n r_n}{\sqrt{\mathrm{C}_{\ell_{\max}}}}\right),$$

em que $v_1, \ldots, v_n$ são os resíduos ordinários da regressão linear de $\mathbf{X}_1$ sobre as colunas de $\mathbf{X}_2$, ou seja, o vetor $\mathbf{v} = (v_1, \ldots, v_n)^\top$ é dado por $\mathbf{v} = (\mathbf{I}_n - \mathbf{H}_2)\mathbf{X}_1$, $\mathbf{H}_2 = \mathbf{X}_2(\mathbf{X}_2^\top\mathbf{X}_2)^{-1}\mathbf{X}_2^\top$. Aqui, a matriz $\mathbf{B}$ tem posto $m = 1$. Logo, há apenas um autovalor diferente de zero. Nesse caso, pode-se tanto utilizar o procedimento descrito acima para calcular $\boldsymbol{\ell}_{\max}$ como obtê-lo diretamente sem precisar calcular a matriz $\mathbf{H}_2$. Como ilustração, supor que os resultados do ajuste estão armazenados em `fit.model`. Para extrair o vetor $\mathbf{r}$ pode-se aplicar o comando

```
r = resid(fit.model).
```

Se o modelo tem as covariáveis `cov1` e `cov2` além dos fatores `A` e `B`, o vetor $\boldsymbol{\ell}_{\max}$ correspondente, por exemplo à covariável `cov1`, sai de

```
fit = lm(cov1 ~ A + B + cov2 - 1)
v = resid(fit)
lmax = v*r
tot = t(lmax)%*%lmax
lmax = lmax/sqrt(tot)
```

```
lmax = abs(lmax).
```

Extensão para os MLGs

A metodologia de influência local pode ser facilmente estendida para a classe de MLGs. Em particular, considerando ϕ conhecido e perturbação de casos em que $\mathrm{L}(\boldsymbol{\beta}|\boldsymbol{\delta}) = \sum_{i=1}^{n} \delta_i \mathrm{L}_i(\boldsymbol{\beta})$ com $0 \leq \delta_i \leq 1$, a matriz $\boldsymbol{\Delta}$ assume a forma

$$\boldsymbol{\Delta} = \sqrt{\phi}\mathbf{X}^\top \hat{\mathbf{W}}^{\frac{1}{2}} \mathbf{D}(\hat{\mathbf{r}}_P),$$

em que $\mathbf{D}(\hat{\mathbf{r}}_P) = \mathrm{diag}\{\hat{r}_{P_1}, \ldots, \hat{r}_{P_n}\}$ e $\hat{r}_{P_i} = \sqrt{\phi}(y_i - \hat{\mu}_i)/\sqrt{\hat{V}_i}$ é o i-ésimo resíduo de Pearson estimado. Assim, substituindo $-\ddot{\mathbf{L}}_{\beta\beta}$ por $\phi(\mathbf{X}^\top \mathbf{W}\mathbf{X})$ tem-se que a curvatura normal na direção unitária $\boldsymbol{\ell}$ assume a forma

$$\mathrm{C}_\ell(\boldsymbol{\beta}) = 2|\boldsymbol{\ell}^\top \mathbf{D}(\hat{\mathbf{r}}_P)\hat{\mathbf{H}}\mathbf{D}(\hat{\mathbf{r}}_P)\boldsymbol{\ell}|.$$

Se o interesse é calcular a curvatura normal na direção $\boldsymbol{\ell}_i$ da i-ésima observação, então pode-se avaliar o gráfico de índices de $\mathrm{C}_i = 2\hat{h}_{ii}\hat{r}^2_{P_i}$.

Em particular, o vetor $\boldsymbol{\ell}_{max}$ para avaliar a influência local das observações nas estimativas dos parâmetros é o autovetor correspondente ao maior autovalor da seguinte matriz $n \times n$:

$$\mathbf{B} = \mathbf{D}(\hat{\mathbf{r}}_P)\hat{\mathbf{H}}\mathbf{D}(\hat{\mathbf{r}}_P).$$

Para obter $\boldsymbol{\ell}_{max}$, a maneira mais simples é construir a matriz $\mathbf{B}$ e extrair o seu autovetor correspondente ao maior autovalor. Os comandos são os seguintes:

```
B = diag(rp)%*% H %*% diag(rp)
Cmax = eigen(B)$val[1]
lmax = eigen(B)$vec[,1]
lmax = abs(lmax).
```

Por outro lado, se há interesse em detectar observações influentes na estimativa de um coeficiente particular, associado por exemplo à variável explicativa $\mathbf{X}_1$, o vetor $\boldsymbol{\ell}_{max}$ fica dado por

$$\boldsymbol{\ell}_{\max}^{\top} = \left(\frac{v_1 \hat{r}_{P_1}}{\sqrt{\mathrm{C}_{\ell_{\max}}}}, \ldots, \frac{v_n \hat{r}_{P_n}}{\sqrt{\mathrm{C}_{\ell_{\max}}}} \right),$$

em que $v_1, \ldots, v_n$ são agora obtidos da regressão linear de $\mathbf{X}_1$ contra as colunas de $\mathbf{X}_2$ com matriz de pesos $\hat{\mathbf{V}}$, isto é $\mathbf{v} = \hat{\mathbf{V}}^{\frac{1}{2}}\mathbf{X}_1 - \hat{\mathbf{V}}^{\frac{1}{2}}\mathbf{X}_2(\mathbf{X}_2^{\top}\hat{\mathbf{V}}\mathbf{X}_2)^{-1}\mathbf{X}_2^{\top}\hat{\mathbf{V}}\mathbf{X}_1$.

Para ligação não canônica os resultados continuam valendo desde que a matriz observada seja substituída pela matriz de informação de Fisher.

2.8.5 Gráfico da variável adicionada

A seguir é apresentada a versão do gráfico da variável adicionada para os MLGs. Supor um MLG com p parâmetros, $\beta_1, \ldots, \beta_p$, ϕ conhecido, e que um coeficiente adicional γ relacionado a uma variável quantitativa Z está sendo incluído no modelo. O interesse é testar $\mathrm{H}_0 : \gamma = 0$ contra $\mathrm{H}_1 : \gamma \neq 0$.

Seja $\boldsymbol{\eta}(\boldsymbol{\beta}, \gamma)$ o preditor linear com $p + 1$ parâmetros, isto é

$$\boldsymbol{\eta}(\boldsymbol{\beta}, \gamma) = \mathbf{X}^{\top}\boldsymbol{\beta} + \gamma\mathbf{Z}.$$

A função escore para γ é dada por

$$\mathrm{U}_{\gamma} = \frac{\partial \mathrm{L}(\boldsymbol{\beta}, \gamma)}{\partial \gamma} = \phi^{\frac{1}{2}}\mathbf{Z}^{\top}\mathbf{W}^{\frac{1}{2}}\mathbf{r}_P,$$

em que $\mathbf{Z} = (z_1, \ldots, z_n)^{\top}$. De resultados anteriores segue que

$$\mathrm{Var}(\hat{\gamma}) = \phi^{-1}[\mathbf{Z}^{\top}\mathbf{W}^{\frac{1}{2}}\mathbf{M}\mathbf{W}^{\frac{1}{2}}\mathbf{Z}]^{-1},$$

em que $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$. Logo, $\mathrm{Var}(\hat{\gamma}) = \phi^{-1}(\mathbf{R}^{\top}\mathbf{W}\mathbf{R})^{-1}$ com $\mathbf{R} = \mathbf{Z} - \mathbf{X}\mathbf{C}$ e $\mathbf{C} = (\mathbf{X}^{\top}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{W}\mathbf{Z}$.

Portanto, a estatística de escore para testar $\mathrm{H}_0 : \gamma = 0$ contra $\mathrm{H}_1 : \gamma \neq 0$ fica dada por

$$\xi_{SR} = (\hat{\mathbf{r}}_P^\top \hat{\mathbf{W}}^{\frac{1}{2}} \mathbf{Z})^2 / (\mathbf{Z}^\top \hat{\mathbf{W}}^{\frac{1}{2}} \hat{\mathbf{M}} \hat{\mathbf{W}}^{\frac{1}{2}} \mathbf{Z}),$$

em que $\hat{\mathbf{W}}$, $\hat{\mathbf{r}}_P$ e $\hat{\mathbf{M}}$ são avaliados em $\hat{\boldsymbol{\beta}}$ (sob H_0). Sob H_0, $\xi_{SR} \sim \chi^2_1$ quando $n \to \infty$.

Wang (1985) mostra que a estatística de escore acima coincide com a estatística F de uma regressão linear ponderada para testar a inclusão da variável Z no modelo. Nessa regressão linear, o gráfico da variável adicionada é formado pelos resíduos $\hat{\mathbf{r}}_P$ e $\boldsymbol{v} = \phi^{\frac{1}{2}}(\mathbf{I}_n - \hat{\mathbf{H}})\hat{\mathbf{W}}^{\frac{1}{2}}\mathbf{Z}$. O resíduo $\boldsymbol{v}$ pode ser obtido facilmente após a regressão linear ponderada (com pesos $\hat{\mathbf{W}}$) de $\mathbf{Z}$ contra $\mathbf{X}$. Tem-se que $\hat{\gamma} = (\boldsymbol{v}^\top \boldsymbol{v})^{-1} \boldsymbol{v}^\top \mathbf{r}$.

Logo, o gráfico de $\hat{\mathbf{r}}_P$ contra $\boldsymbol{v}$ pode revelar quais observações estão contribuindo mais na significância de γ. A principal dificuldade para construir o gráfico da variável adicionada em MLGs é a obtenção do resíduo $\boldsymbol{v}$, uma vez que o resíduo $\hat{\mathbf{r}}_P$ é obtido facilmente como visto anteriormente. Para ilustrar o cálculo de $\boldsymbol{v}$ num modelo particular, supor duas covariáveis e dois fatores e que o interesse é construir o gráfico da variável adicionada correspondente à covariável `cov1`. É preciso inicialmente ajustar o modelo com os dois fatores e a outra covariável e calcular a matriz $\hat{\mathbf{W}}$ cujos valores serão armazenados em $\mathbf{W}$. Lembrando que $\hat{\mathbf{W}}$ é a matriz estimada de pesos. Supondo, por exemplo, que tem-se um modelo de Poisson com ligação canônica, os passos para construir o gráfico são os seguintes:

```
fit.poisson = glm( resp ~ cov2 + A + B, family=poisson)
w = fit.poisson$weights
W = diag(w)
rp = resid(fit.poisson, type =‘‘pearson’’)
X = model.matrix(fit.poisson)
```

```
H = solve(t(X)%*%W%*%X)
H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
v = sqrt(W)%*%cov1 - H%*%sqrt(W)%*%cov1
plot(v, rp, xlab=``Residuo v'', ylab=``Residuo rp'').
```

2.8.6 Técnicas gráficas

As técnicas gráficas mais recomendadas para os MLGs são as seguintes: (i) pontos aberrantes, gráfico de t_{D_i} ou r_{q_i} (ou algum outro resíduo) contra a ordem das observações ou gráfico normal de probabilidades de cada resíduo com banda de confiança; (ii) variabilidade, gráfico de t_{D_i} ou r_{q_i} contra o valor ajustado $\hat{\mu}_i$; (iii) correlação, gráfico de t_{D_i} ou r_{q_i} contra o tempo ou alguma ordem em que há suspeita de correlação entre as observações; (iv) afastamento da distribuição postulada para a resposta, gráfico normal de probabilidades para t_{D_i} (com envelope) ou r_{q_i} com o `worm plot`; (v) adequação da ligação, gráfico de $\hat{z}_i$ contra $\hat{\eta}_i$ (uma tendência linear indica adequação da ligação) (exceto para o caso binomial); (vi) pontos influentes, gráficos de LD_i, C_i ou $|\boldsymbol{\ell}_{max}|$ contra a ordem das observações e (vii) falta de alugm termo extra numa variável explicativa quantitativa, gráfico da variável adicionada. Os envelopes, no caso de MLGs com distribuições diferentes da normal, são construídos com os resíduos gerados a partir do modelo ajustado (ver, por exemplo, Williams, 1987). No Apêndice B são relacionados programas para gerar envelopes em alguns MLGs.

2.9 Seleção de modelos

Os métodos de seleção de modelos descritos na Seção 1.13 podem ser estendidos diretamente para os MLGs. Algumas observações, contudo, são

necessárias. Nos casos de regressão logística e de Poisson o teste da razão de verossimilhanças, pelo fato de ser obtido pela diferença de duas funções desvio, aparece como o mais indicado. Para os casos de regressão normal, normal inversa e gama o teste F, por não exigir a estimativa de máxima verossimilança do parâmetro de dispersão, é o mais indicado. Isso não impede que outros testes sejam utilizados.

Já o método de Akaike pode ser expresso numa forma mais simples em função do desvio do modelo. Nesse caso, o critério consiste em encontrar o modelo tal que a quantidade abaixo seja minimizada

$$\text{AIC} = \text{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) + 2p,$$

em que $\text{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ denota o desvio do modelo e p o número de parâmetros. Os métodos `stepwise` e de Akaike estão disponíveis no `R`. O método `stepwise` está disponível apenas para modelos normais lineares. O comando `stepwise` é definido por `stepwise(Xvar, resposta)`, em que `Xvar` denota a matriz com os valores das variáveis explicativas e `resposta` denota o vetor com as respostas.

Para rodar o critério de Akaike é preciso usar antes o comando `require(MASS)`. Uma maneira de aplicar o critério de Akaike é partindo do maior modelo cujos resultados são guardados no objeto `fit.model`. Daí, então, deve-se usar o comando `stepAIC(fit.model)`.

2.10 Aplicações

2.10.1 Estudo entre renda e escolaridade

O conjunto de dados descrito na Tabela 2.5, extraído do censo do IBGE de 2000, apresenta para cada unidade da federação o número médio de anos de estudo e a renda média mensal (em reais) do chefe ou chefes do domicílio.

Esses dados estão também armazenados no arquivo **censo.txt**. O arquivo pode ser lido no `R` através do comando

```
censo= read.table(‘‘censo.txt’’, header=TRUE).
```

Propor inicialmente um modelo normal linear simples em que Y denota a renda e X a escolaridade. O modelo fica portanto dado por

$$y_i = \alpha + \beta x_i + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, para $i = 1, \ldots, 27$. Supor que a amostra da Tabela 2.5 é um corte transversal, isto é, as informações sobre renda e escolaridade das unidades da federação são referentes a um determinado instante do tempo. Nesse caso, os erros são devidos à variabilidade da renda (dada a escolaridade) nos diversos instantes do tempo. Assume-se que a relação funcional entre y_i e x_i é a mesma num determinado intervalo do tempo.

Tabela 2.5
Escolaridade e renda média domiciliar no Brasil em 2000.

RR	5,7	685	AP	6,0	683
AC	4,5	526	RO	4,9	662
PA	4,7	536	AM	5,5	627
TO	4,5	520	PB	3,9	423
MA	3,6	343	RN	4,5	513
SE	4,3	462	PI	3,5	383
BA	4,1	460	PE	4,6	517
AL	3,7	454	CE	4,0	448
SP	6,8	1076	RJ	7,1	970
ES	5,7	722	MG	5,4	681
SC	6,3	814	RS	6,4	800
PR	6,0	782	MT	5,4	775
GO	5,5	689	MS	5,7	731
DF	8,2	1499			

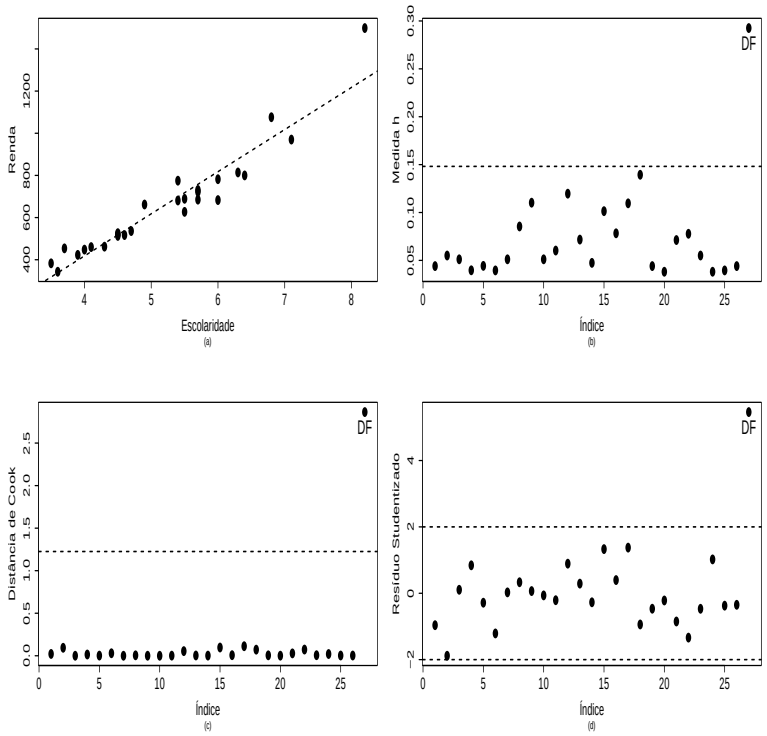

Figura 2.4: Reta ajustada do modelo normal linear e gráficos de diagnóstico para o exemplo sobre renda e escolaridade.

As estimativas dos parâmetros (erro padrão) são dadas por $\hat{\alpha} = -381,28$ $(69,40)$ e $\hat{\beta} = 199,82$ $(13,03)$, indicando que o coeficiente angular da reta é altamente significativo. Essa estimativa pode ser interpretada como o incremento esperado na renda média domiciliar de uma unidade da federação se o tempo de escolaridade médio domiciliar naquela unidade for acrescido de um ano. A estimativa de σ é dada por $s = 77,22$, enquanto que o coeficiente de determinação foi de $R^2 = 0,904$. O ajuste do modelo e a exibição dos resultados podem ser obtidos com os comandos abaixo

```
attach(censo)
fit1.censo = lm(renda ~ escolar)
```

```
summary(fit1.censo).
```

Ou, alternativamente, transformando o arquivo **censo** num arquivo do tipo `data.frame`, através dos comandos

```
censo = data.frame(censo)
fit1.censo = lm(renda ~ escolar, data=censo)
summary(fit1.censo).
```

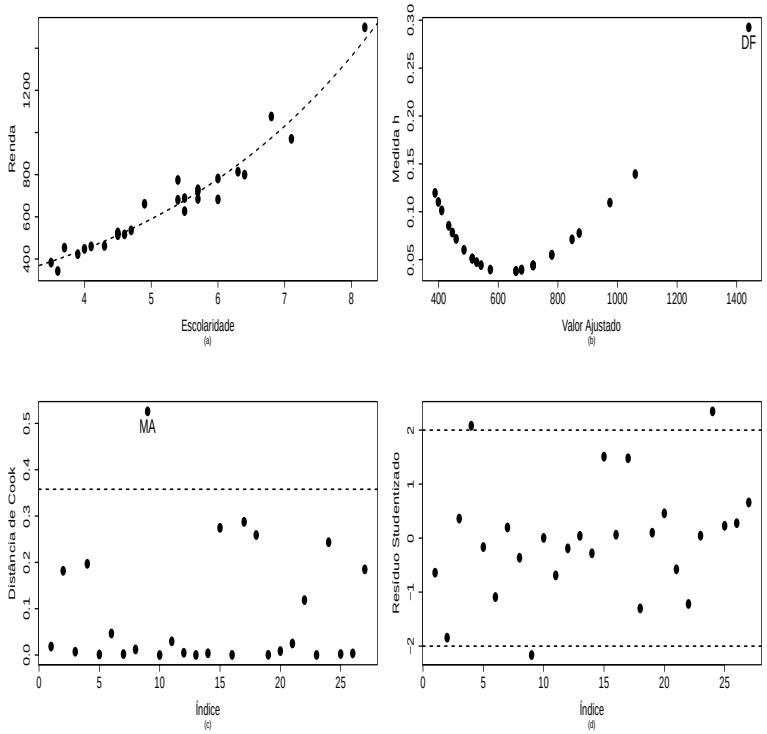

Figura 2.5: Curva ajustada do modelo gama log-linear e gráficos de diagnóstico para o exemplo sobre renda e escolaridade.

Pela Figura 2.4, onde são apresentados alguns gráficos de diagnóstico, além da reta ajustada aos dados, nota-se uma forte discrepância do Distrito Federal que aparece como ponto de alavanca, influente e aberrante. Além disso, nota-se pela Figura 2.4d indícios de variância não constante, ou seja,

um aumento da variabilidade com o aumento da escolaridade. Isso pode também ser notado na Figura 2.4a. Assim, pode-se propor um modelo alternativo, por exemplo, com efeitos multiplicativos conforme dado abaixo

$$\mu_i = e^{\alpha+\beta x_i} e^{\epsilon_i},$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{G}(1, \phi)$, $i = 1, \ldots, 27$. Pode-se ajustar esse modelo no `R` através dos comandos

```
fit2.censo = glm(renda ~ escolar, family=Gamma(link=log))
summary(fit1.censo).
```

Tabela 2.6
Estimativas de algumas quantidades com todos os pontos e quando as observações mais discrepantes são excluídas do modelo gama.

Estimativa	Com todos os pontos	Excluído DF	Excluído MA	Excluídos DF e MA
$\hat{\alpha}$	4,98 (0,068)	5,00 (0,078)	5,03 (0,067)	5,06 (0,077)
$\hat{\beta}$	0,28 (0,013)	0,27 (0,015)	0,27 (0,012)	0,26 (0,015)
$\hat{\phi}$	192(52)	188(52)	223(62)	223(63)

Na Figura 2.5 tem-se o ajuste do modelo gama aos dados, bem como alguns gráficos de diagnóstico que destacam DF como ponto de alavanca e MA como ponto influente, enquanto na Tabela 2.6 tem-se uma análise confirmatória em que verifica-se poucas variações nas estimativas dos parâmetros com a eliminação dessas unidades da federação. Finalmente, na Figura 2.6 tem-se o gráfico normal de probabilidades para o modelo normal linear e para o modelo gama log-linear. Nota-se uma melhor acomodação e distribuição dos pontos dentro do envelope gerado no segundo modelo. Pelo valor da estimativa do parâmetro de dispersão conclui-se que o modelo gama log-linear aproxima-se bem de um modelo normal de média μ e variância $\phi^{-1}\mu^2$.

Portanto, o modelo final ajustado fica dado por

$$\hat{y} = e^{4,98+0,28x}.$$

Desse modelo pode-se extrair a seguinte intrepretação: $e^{\hat{\beta}} = e^{0,28} = 1,32(32\%)$ é o aumento relativo esperado para a renda aumentando-se em 1 ano a escolaridade média.

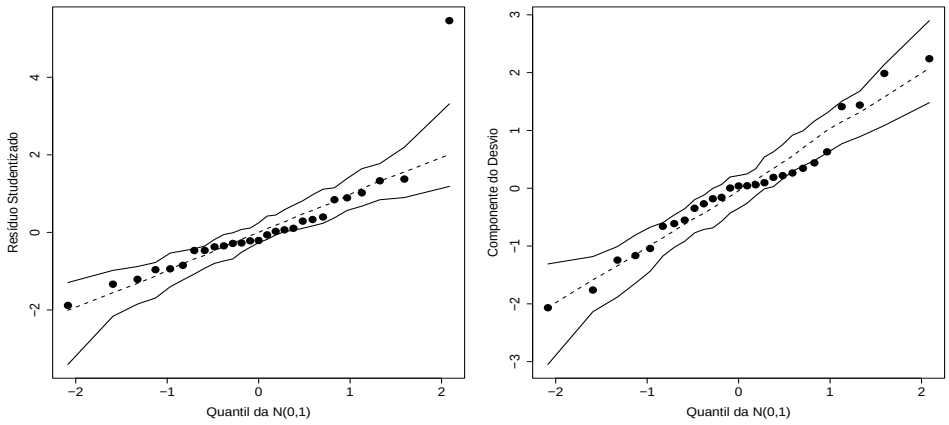

Figura 2.6: Gráfico normal de probabilidades para os modelos ajustados normal linear (esquerda) e gama log-linear (direita) aos dados sobre renda e escolaridade.

2.10.2 Processo infeccioso pulmonar

A seguir serão utilizados os dados referentes a um estudo de caso-controle realizado no Setor de Anatomia e Patologia do Hospital Heliópolis em São Paulo, no período de 1970 a 1982 (Paula e Tuder, 1986) (ver arquivo **canc3.txt**). Um total de 175 pacientes com processo infecioso pulmonar atendido no hospital no período acima foi classificado segundo as seguintes variáveis: Y,

tipo de tumor (1: maligno, 0: benigno); IDADE, idade em anos; SEXO (0: masculino, 1: feminino); HL, intensidade da célula histiócitos-linfócitos (1: ausente, 2: discreta, 3: moderada, 4: intensa) e FF, intensidade da célula fibrose-frouxa (1: ausente, 2: discreta, 3: moderada, 4: intensa). O arquivo pode ser lido no `R` através do comando

```
canc3 = read.table(‘‘canc3.txt’’, header=TRUE).
```

Deve-se informar o sistema que as variáveis SEXO, HL e FF são qualitativas, isto é, deve-se transformá-las em fatores. Os comandos são os seguintes:

```
attach(canc3)
sexo = factor(sexo)
sexo = C(sexo,treatment)
hl = factor(hl)
hl = C(hl,treatment)
ff = factor(ff)
ff = C(ff,treatment).
```

O comando `C(sexo,treatment)`, que é optativo, cria uma variável binária que assume valor zero para o sexo masculino e valor um para o sexo feminino. Analogamente, o comando `C(hl,treatment)` cria variáveis binárias para os níveis discreto, moderado e intenso do fator HL. O mesmo faz o comando `C(ff,treatment)` para o fator FF. Essa maneira de transformar todo fator de k níveis em $k-1$ variáveis binárias, denominado casela de referência, é padrão em MLGs, porém pode não ser a modelagem mais conveniente em outras situações de interesse prático. A casela de referência seria, nesses dois casos, o nível ausente.

Considere, como exemplo, a aplicação do modelo logístico com resposta Bernoulli apenas com os efeitos principais, em que

$$\Pr\{Y = 1 \mid \eta\} = \{1 + \exp(-\eta)\}^{-1},$$

com $\eta = \beta_1 + \beta_2 \text{IDADE} + \beta_3 \text{SEXO} + \sum_{i=1}^{4} \beta_{4i} \text{HL}_i + \sum_{i=1}^{4} \beta_{5i} \text{FF}_i$, SEXO, HL_i e FF_i sendo variáveis binárias correspondentes aos níveis de SEXO, HL e FF, respectivamente. Assume-se que $\beta_{41} = \beta_{51} = 0$. Uma observação importante é que devido ao fato da amostragem ter sido retrospectiva, o uso do modelo acima para fazer previsões somente é válido se a estimativa do intercepto (β_1) ser corrigida (ver, por exemplo, McCullagh e Nelder, 1989, p. 113). Isso será discutido na Seção 4.6.6. Para ajustar o modelo acima, os passos são dados abaixo

```
   fit1.canc3 = glm( tipo ~ sexo + idade + hl + ff,
family=binomial)
   summary(fit1.canc3).
```

Tabela 2.7
Estimativas dos parâmetros referentes ao modelo logístico ajustado aos dados sobre processo infeccioso pulmonar.

Efeito	Estimativa	Efeito	Estimativa	Efeito	Estimativa
Constante	-1,850(1,060)	HL(2)	-0,869(0,945)	FF(2)	-0,687(0,502)
Sexo	0,784(0,469)	HL(3)	-2,249(0,968)	FF(3)	-1,025(0,525)
Idade	0,065(0,013)	HL(4)	-3,295(1,466)	FF(4)	0,431(1,123)

As estimativas dos parâmetros (erro padrão aproximado) são apresentadas na Tabela 2.7. O desvio do modelo foi de $\text{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 157,40$ (166 graus de liberdade), indicando um ajuste adequado. Como pode-se observar, há indícios de que a chance de processso infecioso maligno seja maior para o sexo feminino do que para o sexo masculino. Nota-se também que a chance de processo maligno aumenta significativamente com a idade e há indicações de que tanto para a célula FF quanto para HL a chance de processo maligno diminui à medida que aumenta a intensidade da célula. Esse exemplo será reanalizado no Capítulo 3.

2.10.3 Sobrevivência de bactérias

Na Tabela 2.8, extraída de Montgomery et al.(2001, pgs. 201-202), tem-se o número de bactérias sobreviventes em amostras de um produto alimentício segundo o tempo (em minutos) de exposição do produto a uma temperatura de 300^oF. Na Figura 2.7a é apresentado o gráfico do número de bactérias sobreviventes contra o tempo de exposição. Nota-se uma tendência decrescente e quadrática.

Supondo que as amostras do produto enlatado submetidos à temperatura de 300^oF têm o mesmo tamanho, pode-se pensar, em princípio, que $Y_i \overset{\text{ind}}{\sim} \text{P}(\mu_i)$, com Y_i denotando o número de bactérias sobreviventes na i-ésima amostra $i = 1, \ldots, 12$. Para μ_i grande é razoável supor que Y_i se aproxima de uma distribuição normal (ver Seção 5.3.1). Assim, tem-se como proposta inicial, os seguintes modelos:

$$y_i = \alpha + \beta \text{tempo}_\text{i} + \epsilon_i \quad \text{e}$$

$$y_i = \alpha + \beta \text{tempo}_\text{i} + \gamma \text{tempo}_\text{i}^2 + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, $i = 1, \ldots, 12$.

Tabela 2.8
Número de bactérias sobreviventes e tempo de exposição.

Número	175	108	95	82	71	50	49	31	28	17	16	11
Tempo	1	2	3	4	5	6	7	8	9	10	11	12

As estimativas dos parâmetros são apresentadas na Tabela 2.9. Pelos gráficos de envelope (Figuras 2.7b e 2.7c) nota-se indícios de que a distribuição dos erros pode estar incorrretamente especificada. A maioria dos resíduos assume valor negativo. Nota-se a presença de um ponto aberrante, observação #1.

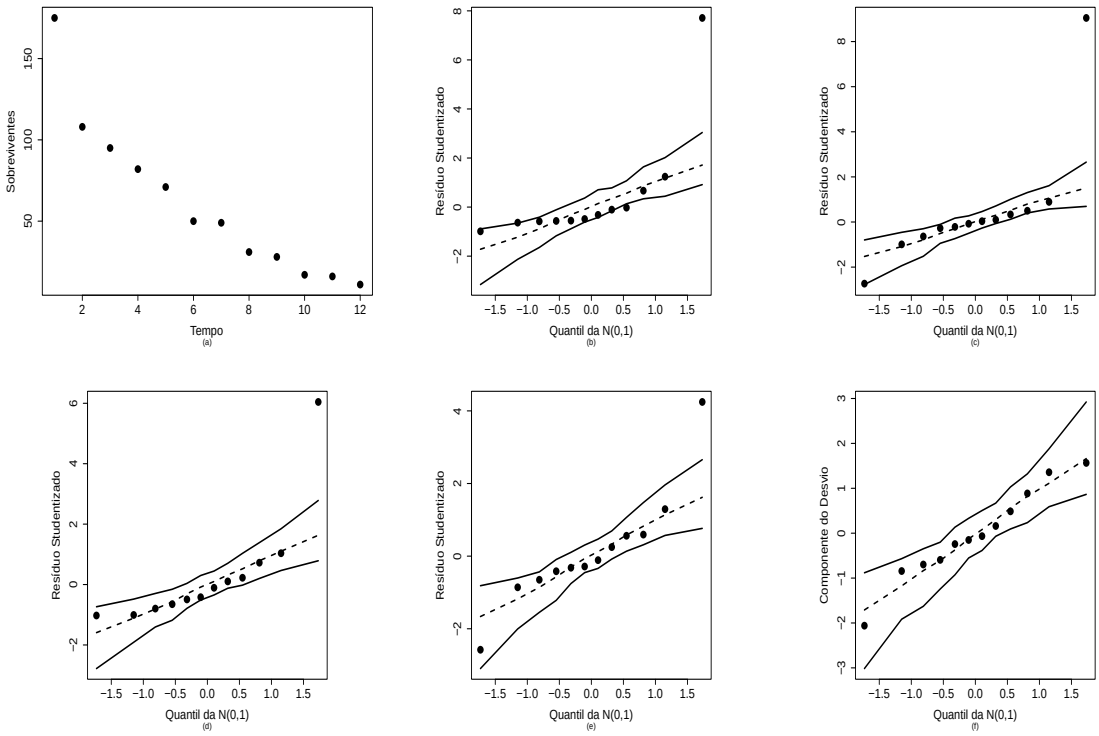

Figura 2.7: Diagrama de dispersão e gráficos normais de probabilidade referentes aos modelos ajustados aos dados sobre sobrevivência de bactérias.

Uma outra tentativa seria aplicar à resposta a transformação raiz quadrada que é conhecida no caso da Poisson como estabilizadora da variância, além de manter a aproximação normal (ver Seção 5.3.1). Logo, pode-se pensar em adotar os seguintes modelos alternativos:

$$\sqrt{y_i} = \alpha + \beta \text{tempo}_\text{i} + \epsilon_i \quad \text{e}$$

$$\sqrt{y_i} = \alpha + \beta \text{tempo}_\text{i} + \gamma \text{tempo}_\text{i}^2 + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, $i = 1, \ldots, 12$. As estimativas dos parâmetros são apresentadas na Tabela 2.9.

Tabela 2.9
Estimativas de algumas quantidades para os modelos com resposta transformada ajustados aos dados sobre sobrevivência de bactérias.

Parâmetro	Linear-Y	Quadrático-Y	Linear-$\sqrt{Y}$	Quadrático-$\sqrt{Y}$
α	142,20(11,26)	181,20(11,64)	12,57(0,38)	13,64(0,51)
β	-12,48(1,53)	-29,20(4,11)	-0,82(0,05)	-1,27(0,18)
γ		1,29(0,31)		0,04(0,01)
R^2	86,9%	95,5%	96,1%	97,8%

Nota-se uma melhora na qualidade do ajuste, particularmente no segundo caso. Porém, ainda há indícios pelos gráficos de envelope (Figuras 2.7d e 2.7e) de violação nas suposições para os modelos, além da presença da observação #1 como ponto aberrante. Finalmente, propôem-se um modelo log-linear de Poisson, em que

- $Y_i|\text{tempo}_\text{i} \overset{\text{ind}}{\sim} \text{P}(\mu_i)$
- $\log(\mu_i) = \alpha + \beta\text{tempo}_\text{i}$,

$i = 1, \ldots, 12$. As estimativas dos parâmetros são apresentadas na Tabela 2.10. Pelo gráfico de envelope (Figura 2.7f) não há evidências de que o modelo esteja mal ajustado. Nota-se também que a observação #1 foi acomodada dentro do envelope gerado. Parece, portanto, que esse último modelo é o que melhor se ajusta aos dados dentre os modelos propostos.

Tabela 2.10
Estimativas dos parâmetros do modelo de Poisson ajustado aos dados sobre sobrevivência de bactérias.

Parâmetro	Estimativa	E/E.Padrão
α	5,30	88,34
β	-0,23	-23,00
Desvio		8,42 (10 g.l.)

O modelo Poisson log-linear ajustado aos dados fica então dado por

$$\hat{\mu}(x) = e^{5{,}30-0{,}23x},$$

em que x denota o tempo de exposição. Logo, diminuindo de uma unidade o tempo de exposição a variação no valor esperado fica dada por

$$\frac{\hat{\mu}(x-1)}{\hat{\mu}(x)} = e^{0{,}23} = 1{,}259.$$

Ou seja, o número esperado de sobreviventes aumenta 25,9%.

2.10.4 Consumo de combustível

No arquivo **reg2.txt** (Gray, 1989) são apresentadas as siglas dos 48 estados norte-americanos contíguos juntamente com as seguintes variáveis: (i) taxa (taxa do combustível no estado em USD), (ii) licença (proporção de motoristas licenciados), (iii) renda (renda per capita em USD), (iv) estradas (ajuda federal para as estradas em mil USD) e (v) consumo (consumo de combustível por habitante). O interesse nesse estudo é tentar explicar o consumo médio de combustível pelas variáveis taxa, licença, renda e estradas. O arquivo pode ser lido no `R` através do comando

`reg2 = read.table(‘‘reg2.txt’’, header=TRUE).`

O modelo proposto é o seguinte:

$$y_i = \alpha + \beta_1 \text{taxa}_\text{i} + \beta_2 \text{licenca}_\text{i} + \beta_3 \text{renda}_\text{i} + \beta_4 \text{estradas}_\text{i} + \epsilon_i,$$

em que y_i denota o consumo anual de combustível (por habitante) no i-ésimo estado, enquanto $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, $i = 1, \ldots, 48$.

O modelo acima é ajustado no `R` com os resultados sendo colocados no objeto `fit1.reg2`. Daí então é aplicado o método de Akaike para selecionar o submodelo com menor AIC. Para tal, aplica-se os comandos

```
require(MASS)
stepAIC(fit1.reg2).
```

A variável estradas foi eliminada. Os resultados do modelo selecionado são apresentados na Tabela 2.11. Portanto, pode-se dizer que para cada aumento de uma unidade na renda, o consumo médio de combustível diminui 0,07 unidades. Para cada aumento de 1% na porcentagem de motoristas licenciados o consumo médio de combustível aumenta 13,75 unidades, e para cada aumento de 1% no imposto do combustível o consumo médio diminui 29,48 unidades.

Tabela 2.11
Estimativas dos parâmetros referentes ao modelo normal linear ajustado aos dados sobre consumo de combustível.

Efeito	Estimativa	E/E.Padrão
Constante	307,33	1,96
Taxa	-29,48	-2,78
Licença	1374,77	7,48
Renda	-0,07	-4,00
R^2	0,675	
s	8,12	

Na Figura 2.8 tem-se alguns gráficos de diagnóstico e como pode-se notar há um forte destaque para o estado de `WY`, que aparece como influente (Figura 2.8b) e aberrante (Figura 2.8c). Outros estados, tais como `CT, NY, SD, TX` e `NV` (Figura 2.8a) aparecem como remotos no subespaço gerado pelas colunas da matrix $\mathbf{X}$, embora não sejam confirmados como influentes. Não há indícios pela Figura 2.8d de variância não constante.

Pelo gráfico normal de probabilidades descrito na Figura 2.9 (esquerda) não há indícios fortes de afastamentos da suposição de normalidade para os erros, apesar da influência no gráfico do estado de `WY`. O gráfico sem esse estado apresentado na Figura 2.9 (direita) confirma esse suposição.

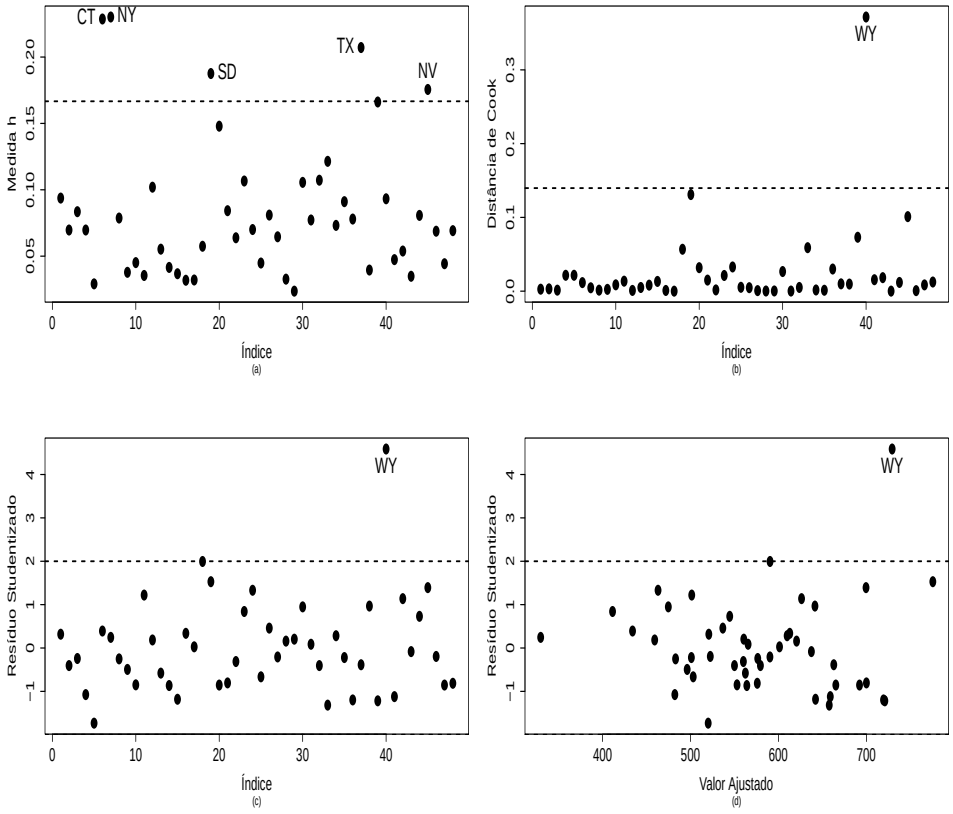

Figura 2.8: Gráficos de diagnóstico referentes ao modelo normal linear ajustado aos dados sobre consumo de combustível.

Analisando os dados referentes ao estado de `WY` nota-se que o mesmo tem uma taxa de 7% (abaixo da média de 7,67%), uma renda per capita anual de USD 4345 (ligeiramente acima da média de USD 4241,83), uma proporção de motoristas licenciados de 0,672 (acima da média de 0,570), porém um consumo médio de combustível muito alto 968 (média nacional de 576,77). Talvez as longas distâncias do estado tenham obrigado os motoristas a um consumo alto de combustível. A eliminação desse estado muda substacialmente algumas estimativas, embora não mude a inferência. A estimativa da

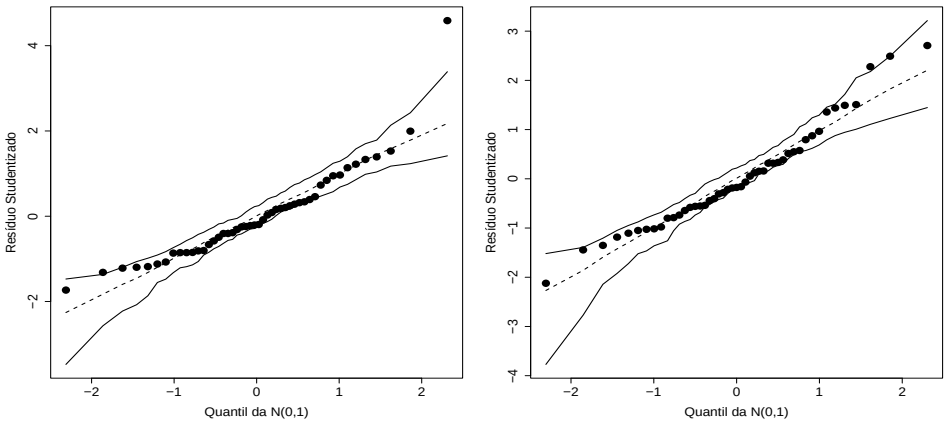

Figura 2.9: Gráfico normal de probabilidades com todos os pontos (esquerda) e sem o estado de WY (direita), referentes ao modelo normal linear ajustado aos dados sobre consumo de combustível.

variável licença cai 13,2%, a estimativa do intercepto aumenta 27,8%, o s^2 cai 17,1% e o R^2 aumenta 4,1%. As demais estimativas não sofrem grandes variações.

Assim, a inclusão de alguma variável que leve em conta a densidade demográfica de cada estado, conforme sugerido por Gray (1989), poderia explicar melhor o estado de `WY` pelo modelo proposto. Uma outra possibilidade seria a inclusão no modelo de uma variável binária referente a esse estado.

2.10.5 Crédito bancário

Considere o arquivo **credit** da biblioteca `Fahrmeir` do R, em que são descritas as seguintes variáveis referentes a empréstimos concedidos a $n = 1000$ clientes de um banco alemão: (i) `Y`: classificação do cliente com relação ao empréstimo (bom pagador, mal pagador), (ii) `Cuenta`: qualidade da conta do cliente (sem classificação, boa, ruim), (iii) `Mes`: duração do empréstimo em meses, (iv) `Ppag`: informação prévia do cliente (bom pagador, mal pagador), (v) `Uso`:

finalidade do empréstimo (privado, profissional), (vi) `DM`: valor do empréstimo em DM, (vii) `Sexo`: gênero do cliente (masculino, feminino) e (viii) `Estc`: estado civil do cliente (vive sozinho, não vive sozinho). O objetivo do estudo é ajustar a probabilidade de um cliente ser bom pagar dadas as demais variáveis explicativas.

A seguir é apresentada uma análise descritiva dos dados com tabelas de contingência entre a resposta e as variáveis explicativas categóricas e boxplots para as variáveis explicativas contínuas.

Tabela 2.12
Classificação do cliente segundo qualidade da conta.

	Sem Info	Boa	Ruim
Bom Pagador	139 (50,7%)	348 (88,3%)	213 (64,2%)
Mal Pagador	135 (49,3%)	46 (11,7%)	119 (35,8%)
Total	274 (100%)	394 (100%)	332 (100%)

A porcentagem de clientes classificados como bom pagador é bastante alta entre clientes com qualidade boa da conta, sendo menor para as demais categorias.

Tabela 2.13
Classificação do cliente segundo informação prévia. do cliente.

	Pré Bom Pagador	Pré Mal Pagador
Bom Pagador	664 (72,9%)	36 (44,4%)
Mal Pagador	247 (27,1%)	53 (55,6%)
Total	911 (100%)	89 (100%)

Clientes com informação prévia de bom pagador são classificados com porcentagem maior como bom pagador em relação a clientes com informação prévia de mal pagador.

Tabela 2.14
Classificação do cliente segundo finalidade. do empréstimo.

	Uso Privado	Uso Profissional
Bom Pagador	485 (73,8%)	215 (62,7%)
Mal Pagador	172 (26,2%)	128 (37,3%)
Total	657 (100%)	343 (100%)

A porcentagem de clientes classificados como bom pagador no grupo que pediu empréstimo para uso privado é maior em relação ao grupo cujo empréstimo foi para uso profissional.

Tabela 2.15
Classificação do cliente segundo sexo.

	Feminino	Masculino
Bom Pagador	268 (66,7%)	432 (72,2%)
Mal Pagador	134 (33,3%)	166 (27,8%)
Total	402 (100%)	598 (100%)

A porcentagem de homens classificados como bom pagador é ligeiramnte superior ao grupo feminino.

Tabela 2.16
Classificação do cliente segundo estado civil.

	Não Vive Sozinho	Vive Sozinho
Bom Pagador	469 (73,3%)	231 (64,2%)
Mal Pagador	171 (26,7%)	129 (35,8%)
Total	640 (100%)	360 (100%)

No grupo de clientes que não vivem sozinhos a porcentagem de bom pagador é maior com relação ao grupo de clientes que vivem sozinhos.

Clientes classificados como bom pagar em geral têm empréstimos em períodos mais curtos e com valores menores.

Denote por Y a classificação do cliente (=1:bom pagador,=0:mal pagador). Supor o modelo logístico binomial para explicar a probabilidade de ser

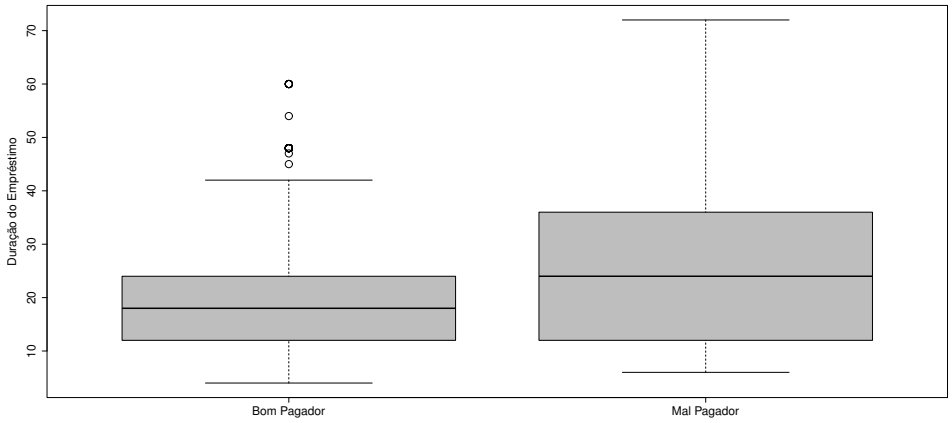

Figura 2.10: Boxplots da duração do empréstimo (em meses) segundo a classificação do cliente.

bom pagador $Y_i|\mathbf{x}_i \overset{\text{ind}}{\sim} \text{Be}(\pi_i)$ com componente sitemático

$$\begin{aligned} \log\left(\frac{\pi_i}{1-\pi_i}\right) &= \beta_1 + \beta_2 \text{CuentaB}_i + \beta_3 \text{CuentaR}_i + \beta_4 \text{Mes}_i + \\ &\quad \beta_5 \text{Ppag}_i + \beta_6 \text{Uso}_i + \beta_7 \text{DM}_i + \beta_8 \text{Sexo}_i + \beta_9 \text{Estc}_i, \end{aligned}$$

em que π_i denota a probabilidade do i-ésimo cliente ser bom pagador, para $i = 1, \ldots, 1000$.

Para ajustar o modelo no GAMLSS aplicar os comandos

```
resp = as.numeric(Y)
resp = abs(resp-2)
require(gamlss)
fit1.credito = gamlss(resp ~ Cuenta + Mes + Ppag + Uso + DM +
Sexo + Estc, family=BI)
summary(fit1.credito)
fit2.credito = stepGAIC(fit1.credito, direction=‘‘both’’)
summary(fit2.credito).
```

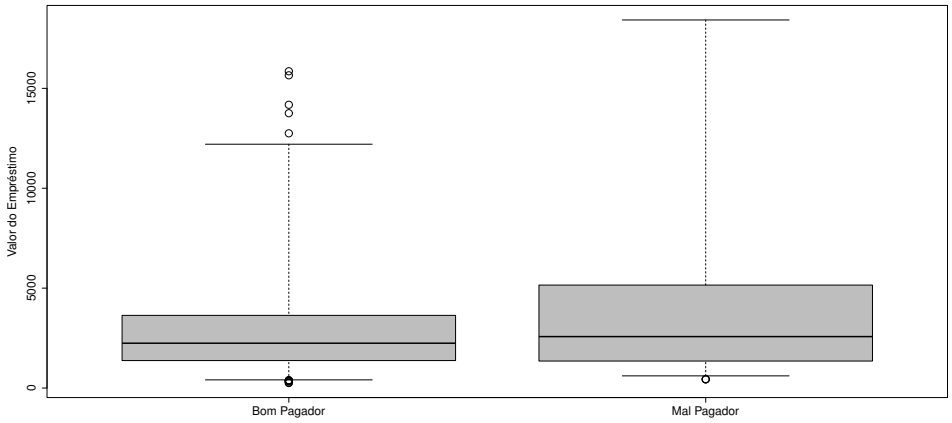

Figura 2.11: Boxplots do valor do empréstimo (em DM) segundo a classificação do cliente.

Após aplicação do comando `StepGAIC` do GAMLSS foram excluídas as variáveis explicativas Sexo e Valor do Empréstimo. As estimativas dos parâmetros do modelo final ajustado estão descritas na Tabela 2.17. Todos os efeitos altamente significativos. O desvio do modelo foi de $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 1019.1$ com 993 graus de liberdade. Nota-se que a probabilidade da ser bom pagador é maior para clientes com conta classificada como boa, diminui com a qunatidade de meses do empréstimo, é menor para clientes classificados previamente como mal pagador, é menor para clientes cujo empréstimo é para uso profissional e para clientes que moram sozinho. Estimativas de razões de chances sãp apresentdas na Tabela 2.17, em geral de acordo com as tabelas da análise descritiva.

Nota-se pelas estimativas *bootstrap* BC_a (Figura 2.12) concordância com a seleção pelo método de Akaike. As únicas estimativas intervalares de 95% que cobrem o valor zero são justamente dos coeficientes das variáveis explicativas Sexo e Valor do Empréstimo. Nota-se também excelente aproximação para

a distribuição normal de todas as estimativas dos parâmetros.

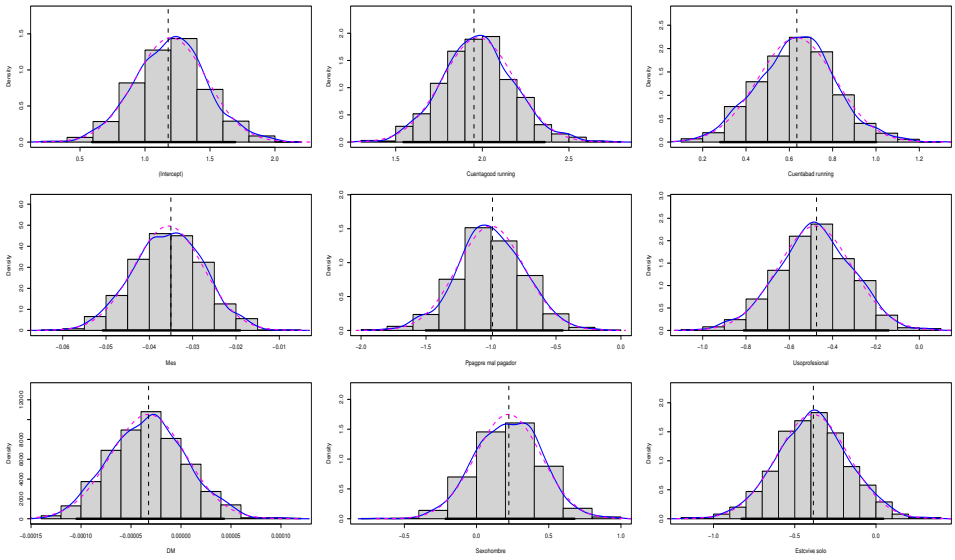

Figura 2.12: Distribuição amostral através de *boostrap* das estimativas dos parâmetros do modelo logístico binomial ajustado aos dados de crédito bancário.

Tabela 2.17
Estimativas dos parâmetros referentes ao modelo logístico binomial ajustado aos dados sobre crédito bancário.

Efeito	Estimativa	E/E.Padrão
Constante	1,346	6,40
CuentaB	1,938	9,43
CuentaR	0,617	3,51
Mes	-0,039	-6,17
PpagMalPag	-0,988	-3,91
UsoProfiss	-0,469	-2,94
EstcViveSo	-0,533	-3,35

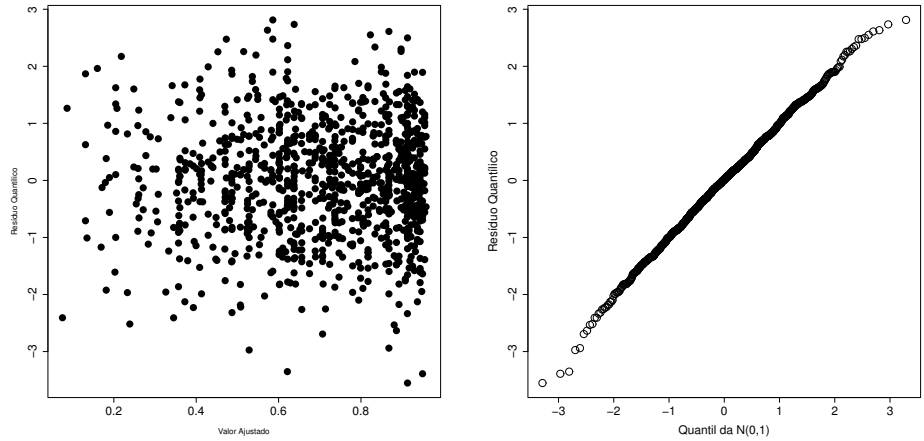

Figura 2.13: Resíduo quantílico contra o valor ajustado (esquerda) e gráfico normal de probabilidades com o resíduo quantílico (direita) do ajuste do modelo logístico binomial aos dados de crédito bancário.

Tabela 2.17
Estimativas pontual e intervalar de 95% para razões de chances referente ao ajuste do modelo logístico binomial aos dados de crédtio bancário.

Efeitos	E. Pontual	E. Intervalar 95%
CuentaB/CuentaR	3,75	[2,52;5,57]
Mes/(Mes + 1)	1,04	[1,03;1,05]
PreBomP/PreMalP	2,69	[1,64;4,41]
UsoPriv/UsoProf	1,60	[1,17;2,19]
NViveSo/ViveSo	1,70	[1,25;2,33]

Por exemplo, a razão de chances (de ser bom pagador) entre cliente de qualidade da conta boa e cliente de qualidade da conta ruim é estimada por $\widehat{\psi} = \exp(1,938 - 0,617) = 3,75$ com estimativa intervalar $[2,52; 5,57]$ de 95%. Os gráficos com o resíduo quantílico aleatoriazado pelo GAMLSS (Figuras 2.13 e 2.14) confimam a adequação do modelo ajustado. Para gerar esses gráficos usar os comandos

```
plot(fitted(fit2.credito), resid(fit2.credito),main=‘‘’’,
```

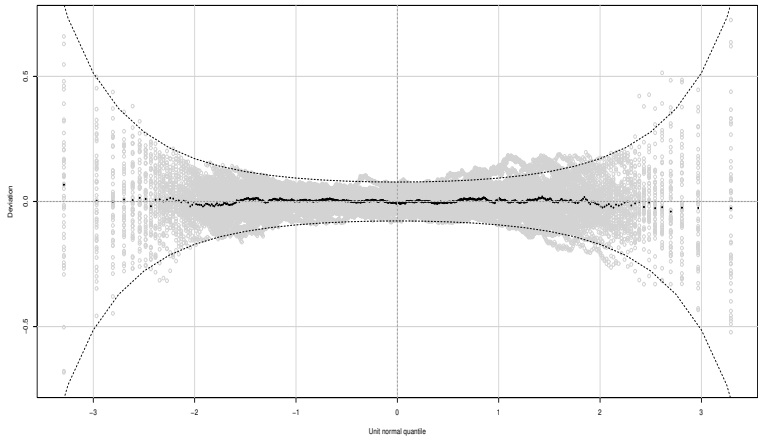

Figura 2.14: Worm plot do ajuste do modelo logístico binomial aos dados de crédito bancário.

```
xlab=‘‘Valor Ajustado’’, ylab=‘‘Resíduo Quantíliuco’’,
cex=2, cex.axis=1.5, cex.axis=1.5, pch=16)
qqnorm(resid(fit2.credito), cex=2, cex.lab=1.5, cex.axis=1.5,
main=‘’’, xlab=‘‘Quantil da N(0,1)’’, ylab=‘‘Resíduo Quantílico’’)
rqres.plot(fit2.credito, howmany=50, plot=‘‘all’’).
```

2.11 Exercícios

1. Se Y pertence à família exponencial de distribuições, então a função densidade ou função de probabilidades de Y pode ser expressa na forma

$$f(y;\theta,\phi) = \exp[\phi\{y\theta - b(\theta)\} + c(y;\phi)],$$

em que $b(\cdot)$ e $c(\cdot;\cdot)$ são funções diferenciáveis. Supondo ϕ conhecido seja $\mathrm{L}(\theta) = \log\{f(y;\theta,\phi)\}$ o logaritmo da função de verossimilhança.

Se $\mathrm{L}(\theta)$ é pelo menos duas vezes diferenciável em θ mostre que

$$\mathrm{E}\left(\frac{\partial \mathrm{L}(\theta)}{\partial \theta}\right) = 0 \;\text{ e }\; \mathrm{E}\left(\frac{\partial^2 \mathrm{L}(\theta)}{\partial \theta^2}\right) = -\mathrm{E}\left\{\left(\frac{\partial \mathrm{L}(\theta)}{\partial \theta}\right)^2\right\}.$$

2. Seja $Y \sim \mathrm{ES}(\mu, \phi)$ (distribuição estável) cuja função densidade de probabilidade é dada por

$$f(y;\theta,\phi) = \mathrm{a}(y,\phi)\exp[\phi\{\theta(y+1) - \theta\log(\theta)\}],$$

em que $\theta > 0$, $-\infty < y < \infty$, $\phi^{-1} > 0$ é o parâmetro de dispersão e $a(\cdot,\cdot)$ é uma função normalizadora. Mostre que essa distribuição pertence à família exponencial de distribuições. Encontre a função de variância e os componentes da função desvio $d^{*2}(y_i;\hat{\mu}_i)$.

3. Supor agora que $Y_{ij} \overset{\text{ind}}{\sim} \mathrm{ES}(\mu_i, \phi)$, para $i = 1, 2$ e $j = 1, \ldots, m$, em que $\mu_1 = \eta_1 = \alpha - \Delta$ e $\mu_2 = \eta_2 = \alpha + \Delta$. Mostre que $\hat{\mu}_1 = \bar{y}_1$ e $\hat{\mu}_2 = \bar{y}_2$. Como ficam as matrizes $\mathbf{X}$ e $\mathbf{W}$? Obter as variâncias e covariância assintóticas $\mathrm{Var}(\hat{\alpha})$, $\mathrm{Var}(\hat{\Delta})$ e $\mathrm{Cov}(\hat{\alpha}, \hat{\Delta})$. Mostre que a estatística do teste de Wald para testar $\mathrm{H}_0 : \alpha - \Delta = 0$ contra $\mathrm{H}_1 : \alpha - \Delta \neq 0$ pode ser expressa na forma

$$\xi_W = m\hat{\phi}\bar{y}_1^2 e^{\bar{y}_1}.$$

Qual a distribuição nula assintótica da estatística do teste?

4. Seja Y o número de ensaios independentes até a ocorrência do r-ésimo sucesso, em que π é a probabilidade de sucesso em cada ensaio. Denote $Y \sim \mathrm{Pascal}(r, \pi)$ (distribuição de Pascal) cuja função de probabilidade é dada por

$$f(y;r,\pi) = \binom{y-1}{r-1}\pi^r(1-\pi)^{(y-r)},$$

para $y = r, r+1, \ldots$ e $0 < \pi < 1$. Mostre que $Y^* = \frac{y}{r}$ pertence à família exponencial de distribuições. Encontre a função de variância $V(\mu)$, em

que $\mu = \mathrm{E}(Y^*)$. Supor agora que $Y_i \overset{\text{ind}}{\sim} \text{Pascal}(r, \pi_i)$ para $i = 1, \ldots, n$. Obtenha os componentes $d^{*2}(y_i; \hat{\pi}_i)$ da função desvio.

5. Considere a função densidade de probabilidade da distribuição hiperbólica:

$$f(y; \theta, \phi) = a(y; \phi)\exp[\phi\{y\theta + (1 - \theta^2)^{\frac{1}{2}}\}],$$

em que $0 < \theta < 1$, $-\infty < y < \infty$, $\phi^{-1} > 0$ é o parâmetro de dispersão,

$$a(y; \phi) = \frac{\phi K_1(\phi\sqrt{1 + y^2})}{\pi\sqrt{1 + y^2}}$$

é uma função normalizadora e $K_1(\cdot)$ é a função de Bessel de ordem 1. Mostre que essa distribuição pertence á família exponencial. Encontre a função de variância $V(\mu)$. Obtenha os componentes do desvio $d^{*2}(y_i; \hat{\mu}_i)$ supondo uma amostra de n variáveis aleatórias independentes de médias μ_i e parâmetro de dispersão ϕ^{-1}, para $i = 1, \ldots, n$. Expresse também a medida R^2.

6. Mostre que a distribuição logarítmica, com função de probabilidade

$$f(y; \rho) = \rho^y / \{-y\log(1 - \rho)\},$$

em que $y = 1, 2, \ldots$ e $0 < \rho < 1$, pertence à família exponencial. Calcule μ e $V(\mu)$. Obtenha a função desvio supondo uma amostra de n variáveis aleatórias independentes de parâmetros ρ_i, $i = 1, \ldots, n$. Obter R^2.

7. Supor que $Y_i \overset{\text{ind}}{\sim} \text{LG}(\rho_i)$, para $i = 1, \ldots, n$, em que $\rho_i = e^\alpha/(1 + e^\alpha)$ e LG denota distribuição logarítmica. Mostre que a variância assintótica de $\hat{\alpha}$ pode ser expressa na forma $\text{Var}(\hat{\alpha}) = \tau^2(\alpha)/ne^\alpha\{\tau(\alpha) - e^\alpha\}$, em que $\tau(\alpha) = (1 + e^\alpha)\log(1 + e^\alpha)$. Como fica a estatística do teste de escore para testar $\mathrm{H}_0 : \alpha = 0$ contra $\mathrm{H}_1 : \alpha \neq 0$? Qual a distribuição nula assintótica da estatística do teste?

8. Supor $Y_i \overset{\text{iid}}{\sim} \text{Ge}(\pi)$ em que $\pi = \exp(\alpha)/\{1+\exp(\alpha)\}$, para $i = 1, \ldots, k$. Obter a estimativa de máxima verossimilhança $\widehat{\alpha}$ e a respectiva variância assintótica $\text{Var}(\widehat{\alpha})$. Mostre que a estatística do teste da razão de verossimilhanças para testar $\text{H}_0 : \alpha = 0$ contra $\text{H}_1 : \alpha \neq 0$ pode ser expressa na forma

$$\xi_{RV} = 2n \left\{ \widehat{\alpha} + \bar{y} \log \left(\frac{2}{1+e^{\widehat{\alpha}}} \right) \right\}.$$

Qual a distribuição nula assintótica da estatístca do teste? A função de probabilidade de Y_i é dada por $f(y_i; \pi) = \pi(1-\pi)^{(y_i-1)}$, para $y_i = 1, 2, \ldots$, $0 < \pi < 1$ e tem-se que $\text{E}(Y_i) = 1/\pi = \{1+\exp(\alpha)\}/\exp(\alpha)$. Sugestão: expressar inicialmente o logaritmo da função de verossimilhança em função de α, denote por $\text{L}(\alpha)$. Obter U_α e $\text{K}_{\alpha\alpha}$.

9. Suponha o MLG em que $Y_i|\mathbf{x}_i \overset{\text{ind}}{\sim} \text{FE}(\mu_i, \phi)$ e parte sistemática dada por $g(\mu_i; \lambda) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$, com λ escalar desconhecido. Encontre as funções escore $\mathbf{U}_\beta$ e U_λ, as funções de informação de Fisher $\mathbf{K}_{\beta\beta}$, $\mathbf{K}_{\beta\lambda}$ e $\text{K}_{\lambda\lambda}$ e descreva o processo iterativo escore de Fisher para obter a estimativa de máxima verossimilhança de $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \lambda)^\top$. Como iniciar o processo iterativo? Sugestão de notação: $\boldsymbol{\Lambda} = \partial\boldsymbol{\eta}/\partial\lambda$, em que $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$.

10. Suponha agora o modelo de regressão normal linear simples

$$y_i = \alpha + \beta x_i + \epsilon_i,$$

em que $\epsilon_i \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2)$, $i = 1, \ldots, n$. Mostre a equivalência entre as estatísticas ξ_{RV}, ξ_W e ξ_{SR} para testar $\text{H}_0 : \beta = 0$ contra $\text{H}_1 : \beta \neq 0$. Supor σ^2 conhecido.

11. Suponha um MLG com ligação canônica e parte sistemática dada por $g(\mu_{1j}) = \alpha_1 + \beta x_j$ e $g(\mu_{2j}) = \alpha_2 + \beta x_j$, $j = 1, \ldots, r$. Interprete esse tipo

de modelo. Obtenha a matriz $\mathbf{X}$ correspondente. Como fica o teste de escore para testar $\mathrm{H}_0 : \beta = 0$? O que significa testar H_0?

12. Sejam Y_{ij}, $i = 1, 2, 3$ e $j = 1, \ldots, m$, variáveis aleatórias mutuamente independentes pertencentes à família exponencial tais que $\mathrm{E}(Y_{ij}) = \mu_{ij}$, $\mathrm{Var}(Y_{ij}) = V_{ij}\phi^{-1}$ e parte sistemática dada por $g(\mu_{1j}) = \alpha$, $g(\mu_{2j}) = \alpha + \Delta$ e $g(\mu_{3j}) = \alpha - \Delta$. Responda às seguintes questões:

    (i) como fica a matriz modelo $\mathbf{X}$?

    (ii) O que significa testar $\mathrm{H}_0 : \Delta = 0$? Qual a distribuição nula assintótica das estatísticas ξ_{RV}, ξ_W e ξ_{SR}?

    (iii) Calcular a variância assintótica de $\hat{\Delta}$, $\mathrm{Var}(\hat{\Delta})$.

    (iv) Mostre que a estatística do teste de escore para testar $\mathrm{H}_0 : \Delta = 0$ contra $\mathrm{H}_1 : \Delta \neq 0$ fica dada por

$$\xi_{SR} = \frac{\phi m(\bar{y}_2 - \bar{y}_3)^2}{2\hat{V}_0}.$$

13. Mostre que a estatística de escore para testar que o i-ésimo ponto é aberrante num MLG com ϕ conhecido e parte sistemática $g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}$ é dada por $t_{S_i}^2$, em que

$$t_{S_i} = \frac{\sqrt{\phi}(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i(1 - \hat{h}_{ii})}},$$

    sendo $\hat{\mu}_i$, $\hat{V}_i$ e $\hat{h}_{ii} = \hat{\omega}_i \mathbf{x}_i^\top (\mathbf{X}^\top \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i$ avaliados em $\hat{\boldsymbol{\beta}}$ (Pregibon, 1982). Qual a distribuição nula assintótica de $t_{S_i}^2$? Como seria interpretado o gráfico de $t_{S_i}^2$ contra a ordem das observações? Sugestão : chame $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \gamma \mathbf{z}$, em que $\mathbf{z}$ é um vetor $n \times 1$ de zeros com 1 na i-ésima posição, calcule $\mathrm{Var}(\hat{\gamma})$ e U_γ e teste $\mathrm{H}_0 : \gamma = 0$ contra $\mathrm{H}_1 : \gamma \neq 0$.

14. No arquivo **trees.txt** (Ryan e Joiner, 1994) é apresentado um conjunto de dados que tem sido analisado sob diversos pontos de vista por vários pesquisadores (ver, por exemplo, Jørgensen, 1989). As variáveis observadas são o diâmetro (d), a altura (h) e o volume (v) de uma amostra de 31 cerejeiras numa floresta do estado da Pensilvânia, EUA. Para ler o arquivo no R use o comando

`trees = read.table(‘‘trees.txt’’, header=TRUE).`

A relação entre diâmetro, altura e volume de uma árvore depende da forma da mesma e pode-se considerar duas possibilidades

$$v = \frac{1}{4}\pi d^2 h$$

para forma cilíndrica e

$$v = \frac{1}{12}\pi d^2 h$$

para forma cônica. Em ambos os casos a relação entre $\log(v)$, $\log(d)$ e $\log(h)$ é dada por $\log(v) = a + b\log(d) + c\log(h)$. Supor inicialmente um modelo linear em que $\epsilon \sim \mathrm{N}(0, \sigma^2)$. Faça uma análise de diagnóstico e verifique se é possível melhorar o modelo, por exemplo incluindo algum termo quadrático.

15. No arquivo **perch.txt** (Keen, 2010, Cap.12) são apresentadas 6 medidas realizadas em um amostra de $n = 56$ peixes da espécie Perch, pescados no lago Laengelmavesi próximo à cidade de Tampere na Finlândia. Essas medidas são descritas a seguir: (i) `Weight` (peso do peixe, em gramas), (ii) `Length1` (comprimento do peixe do nariz até o início da cauda, em cm), (iii) `Length2` (comprimento do peixe do nariz até o entalhe da cauda, em cm), (iv) `Length3` (comprimento do peixe do nariz até o final da cauda, em cm), (v) `Heightpc` (altura máxima, como

porcentagem de `Length3`) e (vi) `Widthpc` (largura máxima, como porcentagem de `Length3`). Para ler o arquivo no R use os comandos

```
perch = read.table("perch.txt", herder=TRUE)
summary(perch)
attach(perch).
```

Fazer inicialmente uma análise descritiva dos dados, tais como boxplots e densidade da variável resposta Weight e diagramas de dispersão (com tendência) entre a variável resposta e as demais variáveis explicativas.

Tentar ajustar um modelo com resposta gama no `gamlss`

```
fit1.perch = gamlss(Weight ~ ·, data=perch, family=GA)
fit2.perch = stepGAIC(fit1.perch, direction="both")
summary(fit2.perch).
```

Quais variáveis explicativas ficaram no modelo? Verifique se é possível incluir termo quadrátcio para as variáveis explicativas selecionadas. Por exemplo, se foram selecionadas Lenght2 e Heightpc, tente incluir termo quadrático em Lenght2 e fazer análise de resíduos

```
fit3.perch = gamlss(Weight ~ Lenght2 + I(Lenght2^2) + Heightpc,
family=GA)
summary(fit3.perch)
plot(fit3.perch)
wp(fit3.perch).
```

Se o modelo está bem ajustado comparar com ajustes de modelos competitivos aplicando em Lenght2 splines cúbicos ou P-splines

```
fit4.perch = gamlss(Weight ~ cs(Lenght2) + Heightpc, family=GA)
```

```
summary(fit4.perch)

plot(fit4.perch)

wp(fit4.perch)

fit5.perch = gamlss(Weight ~ pb(Lenght2) + Heightpc, family=GA)

summary(fit5.perch)

plot(fit5.perch)

wp(fit5.perch).
```

Qual modelo escolher? Apresentar o `term.plot(·)` para o modelo selecionado.

16. No arquivo **aids** do `gamlss` são descritas as seguintes variáveis: (i) `y`, número de casos trimestrais de aids na Inglaterra e País de Gales, (ii) `x`, tempo (em meses) desde janeiro de 1983 e (iii) `qrt`, trimestre para controlar a sazonalidade. O arquivo pode ser disponibilizado diretamente no gamlss através dos comandos

```
require(gamlss)

attach(aids).
```

Fazer inicialmente uma análise descritiva dos dados, tais como densidade e boxplot da variável resposta, diagrama de dispersão (com tendência) entre o tempo e a resposta e boxplots da variável resposta segundo o trimestre.

Comparar os ajustes entre modelos com respostas de Poisson e binomial negativa supondo trimestre como variável categórica (fator) e o tempo em forma quadrática. Posteriormente, comparar modelos com as mesmas respostas, porém substituindo o termo quadrático por um P-*spline* ($pb(x)$). Escolher o melhor modelo através de análise de resíduos.

Por exemplo, para ajustar um modelo com resposta binomial negativa e P-*spline*, aplicar os comandos

`fit.aids = gamlss(y ~ qrt + pb(`x`), family=NBI)`

`summary(fit.aids)`

`plot(fit.aids)`

`rqres.plot(fit.aids, howmany=50, type=‘‘all’’)`

`term.plot(fit.aids, pages=1).`

Para o modelo selecionado interpretar os resultados respondendo se há efeito de trimestre. Apresentar também o gráfico da curva ajustada ao longo do tempo.

17. Considere o arquivo **burn1000** da biblioteca `aplore3` do R, em que 1000 pacientes que sofreram algum tipo de queimadura foram atendidos em 40 centros especializados. Para cada paciente foram observadas as seguintes variáveis: (i) `age`, idade do paciente (em anos) no momento de admissão, (ii) `gender`, gênero do paciente (feminino, masculino), (iii) `race`, raça do paciente (não branco, branco), (iv) `tbsa`, área total afetada pela queimadura (em porcentagem), (v) `inh_inj`, se houve inalação pelo paciente (sim, não), (vi) `flame`, se a queimadura foi provocada por chamas e (vii) `death`, se houve óbito do paciente (sim, não).

    Para disponibilizar e visualizar um resumo dos dados use na sequência os seguintes comandos do R:

    `require(aplore3)`

    `attach(burn1000)`

    `summary(burn1000).`

Fazer inicialmente uma análise descritiva procurando relacionar a variável resposta (`death`) com as demais variáveis explicativas. Use tabelas de contingência e boxplots. Comente. Transforme a variável resposta em variável numérica binária. Por exemplo, usando os comandos

```
resp = as.numeric(death)
```

```
resp = abs(resp-2).
```

Ajustar um modelo logístico binomial no GAMLSS para explicar a probabilidade de óbito do paciente dadas as demais variáveis explicativaas. Por exemplo, através do comando

```
fit1.burn = gamlss(resp ~ age + gender + race + tbsa + inh_inj
+ flame, family=BI).
```

Use o comando `stepGAIC` para selecionar um submodelo

```
fit2.burn = stepGAIC(fit1.burn, direction="both")
```

```
summary(fit2.burn).
```

Fazer uma análise de resíduos através dos comandos

```
plot(fit2.burn)
```

```
rqres.plot(fit2.burn, howmany=8, type="wp")
```

para o submodelo selecionado. Apresentar estimativas intervalares de 95% para as razões de chances.

18. Supor $Y_{ij} \overset{\text{ind}}{\sim} \text{HPB}(\mu_i, \phi)$, em que $i = 1, 2$ e $j = 1, \ldots, m$. Supor $\mu_1 = \eta_1 = \alpha$ e $\mu_2 = \eta_2 = \alpha + \Delta$. Mostrar que $\hat{\mu}_1 = \bar{y}_1$ e $\hat{\mu}_2 = \bar{y}_2$. Obter as matrizes $\mathbf{X}$ e $\mathbf{W}$ e consequentemente as variâncias e covariância assintóticas $\text{Var}(\widehat{\alpha})$, $\text{Var}(\widehat{\Delta})$ e $\text{Cov}(\widehat{\alpha}, \widehat{\Delta})$. Expresse a estatística do

teste de Wald para testar $\mathrm{H}_0 : \Delta = 0$ contra $\mathrm{H}_1 : \Delta \neq 0$ na forma

$$\xi_W = \frac{m}{\widehat{\phi}} \frac{(\bar{y}_2 - \bar{y}_1)^2}{\{V(\bar{y}_1) + V(\bar{y}_2)\}},$$

em que $V(\bar{y}_1) = (1+\bar{y}_1^2)^{\frac{3}{2}}$ e $V(\bar{y}_2) = (1+\bar{y}_2^2)^{\frac{3}{2}}$. Qual a distribuição nula assintótica da estatística do teste? Lembre que a função densidade de probabilidade da distribuição hiperbólica é dada por

$$f(y; \theta, \phi) = a(y; \phi) \exp[\phi\{y\theta + (1-\theta^2)^{\frac{1}{2}}\}],$$

em que $0 < \theta < 1$, $-\infty < y < \infty$, $\phi^{-1} > 0$ é o parâmetro de dispersão.

19. Seja Y variável aleatória com distribuição BNT1(κ) cuja função de probabilidade é expressa na forma

$$f(y; \mu) = \exp\left\{ y \log\left(\frac{\kappa}{\kappa+1}\right) - \log(\kappa) \right\},$$

em que $\kappa > 0$ e $y = 1, 2, \ldots$. Mostre que essa distribuição pertence à família exponencial. Obter V(μ), em que $\mu = \mathrm{E}(Y)$. Supor agora $Y_i \overset{\text{iid}}{\sim} \text{BNT1}(\kappa)$, em que $\log(\kappa) = \gamma$, para $i = 1, \ldots, n$. Obter L(γ), U_γ e $\mathrm{K}_{\gamma\gamma}$. Mostre que a estatística do teste de escore para testar $\mathrm{H}_0 : \gamma = 1$ contra $\mathrm{H}_1 : \gamma \neq 1$ pode ser expressa na forma

$$\xi_{SR} = \frac{n}{e(1+e)}(\bar{y} - e - 1)^2.$$

Qual a distribuição nula assintótica da estatística do teste?

20. Supor $Y_i \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \phi)$, para $i = 1, \ldots, n$. Considere a transformação $\sigma = 1/\sqrt{\phi}$. Obter $\widehat{\sigma}$ e $\mathrm{K}_{\sigma\sigma}$. Mostre que a estatística do teste de Wald para testar $\mathrm{H}_0 : \sigma = 1$ contra $\mathrm{H}_1 : \sigma \neq 1$ pode ser expressa na forma

$$\frac{2n}{D}(\sqrt{D} - \sqrt{n})^2,$$

em que $D = \sum_{i=1}^n (y_i - \bar{y})^2$. Qual a distribuição nula assintótica da estatística do teste? Use os resultados para $\widehat{\phi}$ e $\mathrm{K}_{\phi\phi}$.

Capítulo 3

Modelos para Dados Positivos Assimétricos

3.1 Introdução

A classe de modelos para a análise de dados positivos assimétricos é bastante ampla incluindo distribuições conhecidas para os erros, tais como gama, normal inversa, Weibull, Pareto, log-normal e Birnbaum-Saunders, dentre outras. Essas distribuições têm sido particularmente aplicadas na análise de tempos de sobrevivência (ou duração) com forte ênfase nas áreas médica e de engenharia (ver, por exemplo, Lawless, 2003). Todavia, dados positivos assimétricos têm sido também comuns em outras áreas do conhecimento, como por exemplo pesca, meteorologia, finanças, seguros e atuária (ver, por exemplo, de Jong e Heller, 2008). Um componente importante no estudo de dados de sobrevivência é a possibilidade de incorporação nas análises de observações para as quais não foi possível observar a falha (dados censurados). Com os recentes avanços tecnológicos ocorridos principalmente na fabricação de equipamentos, os tempos até a ocorrência de falhas estão ficando cada vez mais longos, aumentando assim a porcentagem de dados censurados. Isso também pode ser notado na área médica com os avanços nos tratamentos

e medicamentos. Todavia, a inclusão de dados censurados nos modelos envolve um tipo de análise mais específica que está além das metodologias discutidas neste texto. Sugere-se ao leitor mais interessado consultar textos de análise de dados de sobrevivência, como por exemplo os livros de Cox e Oakes (1984), Lawless (2003), Collett (2003) e Colosimo e Giolo (2024). Portanto, neste capítulo será discutido apenas dados positivos assimétricos não censurados sob modelos com resposta gama e normal inversa, os quais já foram introduzidos no Capítulo 2.

3.2 Distribuição gama

Conforme assumido na Seção 2.2.1, supor que Y é uma variável aleatória com distribuição gama de média μ e coeficiente de variação $\phi^{-\frac{1}{2}}$, denota-se $Y \sim \mathrm{G}(\mu, \phi)$, e cuja função densidade é expressa na forma

$$\begin{aligned} f(y_i;\mu,\phi) &= \frac{1}{\Gamma(\phi)}\left(\frac{\phi y}{\mu}\right)^{\phi}\exp\left(-\frac{\phi y}{\mu}\right)\frac{1}{y} \\ &= \exp[\phi\{(-y/\mu)-\log(\mu)\}-\log\{\Gamma(\phi)\}+\phi\log(\phi y)-\log(y)], \end{aligned}$$

em que $y > 0$, $\phi > 0$, $\mu > 0$ e $\Gamma(\phi) = \int_0^\infty t^{\phi-1}e^{-t}dt$ é a função gama. Na Figura 3.1 tem-se a densidade da distribuição gama variando o parâmetro de precisão para μ fixado. Pode-se notar que à medida que ϕ aumenta a distribuição gama fica mais simétrica em torno da média. Pode ser mostrado que à medida que ϕ aumenta Y se aproxima de uma distribuição normal de média μ e variância $\mu^2\phi^{-1}$. Portanto, a distribuição gama torna-se atrativa para o estudo de variáveis aleatórias assimétricas e também simétricas em que a variância depende de forma quadrática da média. Os momentos centrais de Y são expressos na seguinte forma:

$$\mathrm{E}(Y-\mu)^r = \frac{(r-1)!\mu^r}{\phi^{(r-1)}},$$

para $r = 1, 2, \ldots$. Assim, expandindo $\log(Y)$ em série de Taylor em torno de μ até $2^{\underline{a}}$ ordem, obtém-se

$$\log(Y) \cong \log(\mu) + \frac{1}{\mu}(Y - \mu) - \frac{1}{2\mu^2}(Y - \mu)^2.$$

Portanto, para ϕ grande tem-se que

$$\begin{aligned} \mathrm{E}\{\log(Y)\} &\cong \log(\mu) - \frac{1}{2\mu^2}\mathrm{E}(Y - \mu)^2 \\ &= \log(\mu) - \frac{1}{2\mu^2}\frac{\mu^2}{\phi} \\ &= \log(\mu) - (2\phi)^{-1} \quad \text{e} \\ \mathrm{Var}\{\log(Y)\} &\cong \phi^{-1}. \end{aligned}$$

Ou seja, a transformação $\log(Y)$ estabiliza a variância à medida que o coeficiente de variação de Y fica pequeno. Uma outra transformação dada por

$$3\left\{\left(\frac{Y}{\mu}\right)^{\frac{1}{3}} - 1\right\}$$

se aproxima da distribuição normal padrão no caso gama (vide McCullagh e Nelder, 1989, p. 289).

A função de sobrevivência e a função de risco são quantidades usuais na análise de dados de sobrevivência sendo definidas, respectivamente, por

$$\begin{aligned} S(t) &= Pr\{Y \geq t\} \quad \text{e} \\ h(t) &= \lim_{\delta \to 0} \frac{Pr\{t \leq Y < t + \delta | Y \geq t\}}{\delta}. \end{aligned}$$

Em particular, tem-se que a função de risco pode ser expressa na forma $h(t) = f(t)/S(t)$ com $f(y)$ denotando a função densidade de Y. No caso

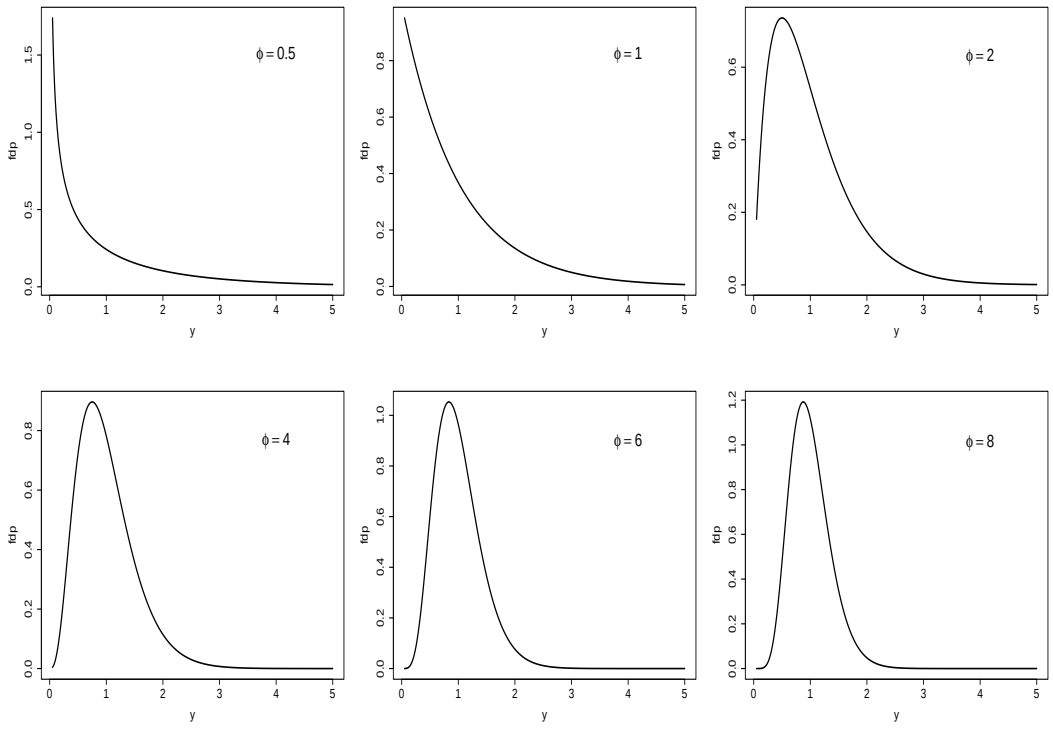

Figura 3.1: Densidades da distribuição gama para alguns valores do parâmetro de precisão e supondo $\mu = 1$.

da distribuição gama de média μ e parâmetro de dispersão ϕ^{-1} a função de sobrevivência é expressa (ver, por exemplo, Collett, 2003, pgs. 197-198) na forma

$$S(t) = 1 - I_{\lambda t}(\phi),$$

em que $I_{\lambda t}(\phi)$ é a função gama incompleta, dada por

$$I_{\lambda t}(\phi) = \frac{1}{\Gamma(\phi)} \int_0^{\lambda t} u^{\phi-1} e^{-u} du,$$

com $\lambda = \frac{\phi}{\mu}$. A função de risco $h(t)$ para a distribuição gama é crescente para $\phi > 1$ e decrescente para $\phi < 1$. Em particular, quando $t \to \infty$ tem-se que $h(t) \to \lambda$.

3.3 Modelos com resposta gama

Supor $Y_1, \ldots, Y_n$ variáveis aleatórias independentes tais que $Y_i \sim \mathrm{G}(\mu_i, \phi)$. Ou seja, está sendo assumido que essas variáveis possuem médias diferentes e mesmo coeficiente de variação $\phi^{-\frac{1}{2}}$. Ademais, supor que $g(\mu_i) = \eta_i$ com $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ contendo valores de variáveis explicativas e $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ sendo o vetor de parâmetros de interesse. As ligações mais usadas no caso gama são identidade ($\mu_i = \eta_i$), logarítmica ($\log(\mu_i) = \eta_i$) e recíproca ($\mu_i = \eta_i^{-1}$), esta última sendo a ligação canônica.

O processo iterativo para estimação de $\boldsymbol{\beta}$, como foi visto na Seção 1.6.1, é dado por

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{z}^{(m)},$$

$m = 0, 1, \ldots$, variável dependente modificada $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{-\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{y} - \boldsymbol{\mu})$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$, $\mathbf{y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, $\mathbf{V} = \mathrm{diag}\{\mu_1, \ldots, \mu_n\}$ e $\mathbf{W} = \mathrm{diag}\{\omega_1, \ldots, \omega_n\}$ com $\omega_i = (d\mu_i/d\eta_i)^2/\mu_i$, $i = 1, \ldots, n$.

É interessante notar que sob ligação logarítmica os pesos do processo iterativo para a obtenção de $\hat{\boldsymbol{\beta}}$ ficam dados por $\omega_i = \frac{\mu_i^2}{\mu_i^2} = 1$, de modo que o processo iterativo assume a forma simplificada

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{z}^{(m)},$$

em que $\mathbf{z} = (z_1, \ldots, z_n)^\top$ com $z_i = \eta_i = (y_i - \mu_i)/\mu_i$ e $\mu_i = \exp(\eta_i)$, $i = 1, \ldots, n$. A variância assintótica de $\hat{\boldsymbol{\beta}}$ fica dada por $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \phi^{-1}(\mathbf{X}^\top \mathbf{X})^{-1}$. Em particular, se as colunas da matriz $\mathbf{X}$ são ortogonais, isto é $\mathbf{X}^\top \mathbf{X} = \mathbf{I}_p$, em que $\mathbf{I}_p$ é a matriz identidade de ordem p, então $\mathrm{Var}(\hat{\beta}_j) = \phi^{-1}$ e $\mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_\ell) = 0$, para $j \neq \ell$, ou seja, β_j e $\hat{\beta}_\ell$ são assintoticamente independentes.

Portanto, a ligação logarítmica tem um atrativo especial de possibilitar o desenvolvimento de experimentos ortogonais como são bem conhecidos

em modelos de regressão normal linear. Pode-se escolher formas apropriadas para a matriz $\mathbf{X}$, de modo que $\mathbf{X}^\top\mathbf{X} = \mathbf{I}_p$, e assim obter estimativas mutuamente independentes e de variância constante para os coeficientes do preditor linear. Myers et al.(2002, Cap.6) discutem experimentos ortogonais em MLGs e apresentam alguns exemplos. As ligações identidade ($\mu = \eta$), raiz quadrada ($\sqrt{\mu} = \eta$) e arcoseno ($\text{sen}^{-1}\sqrt{\mu} = \eta$) produzem o mesmo efeito em MLGs com resposta normal, Poisson e binomial, respectivamente.

Aplicando, para ϕ suficientemente grande, a transformação logarítmica na resposta e ajustando $\text{E}\{\log(Y_i)\} = \mathbf{x}_i^\top\boldsymbol{\beta}$, tem-se de forma equivalente $\log(\mu_i) = \mathbf{x}_i^\top\boldsymbol{\beta} + 2\phi^{-1}$, ou seja, a menos da constante $2\phi^{-1}$ obtém-se as mesmas estimativas para $\boldsymbol{\beta}$ de um modelo com resposta gama e ligação logarítmica.

3.3.1 Qualidade do ajuste

Como foi visto na Seção 2.4 o desvio de um modelo gama é dado por $\text{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = \phi\text{D}(\mathbf{y};\hat{\boldsymbol{\mu}})$, em que

$$\text{D}(\mathbf{y};\hat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\{\log(\hat{\mu}_i/y_i) + (y_i - \hat{\mu}_i)/\hat{\mu}_i\}, \tag{3.1}$$

com $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$ e $\hat{\eta}_i = \mathbf{x}_i^\top\hat{\boldsymbol{\beta}}$. Pode ser mostrado facilmente para ligação logarítmica que o termo $\sum_{i=1}^{n}(y_i - \hat{\mu}_i)/\hat{\mu}_i = 0$ se a parte sistemática η_i contém um intercepto. Nesse caso, a função desvio fica dada por $\text{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = 2\phi\sum_{i=1}^{n}\log(\hat{\mu}_i/y_i)$. O parâmetro ϕ pode ser estimado por máxima verossimilhança, que equivale a resolver a seguinte equação:

$$2n\{\log(\hat{\phi}) - \psi(\hat{\phi})\} = \text{D}(\mathbf{y};\hat{\boldsymbol{\mu}}),$$

em que $\psi(\phi) = \Gamma'(\phi)/\Gamma(\phi)$ é a função digama (vide Seção 1.6.2)¿ Outra opção é utilizar a estimativa consistente $\hat{\phi}^{-1} = (n-p)^{-1}\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2/\hat{\mu}_i^2$

que será discutida no Capítulo 6. Supondo que o modelo postulado está correto tem-se, para ϕ grande, que o desvio $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}})$ segue distribuição qui-quadrado com $(n-p)$ graus de liberdade. Assim, valores altos para o desvio podem indicar inadequação do modelo ou falta de ajuste.

Quando todas as observações são positivas o desvio $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}})$ deve ser utilizado para avaliar a qualidade do ajuste e estimação de ϕ. Contudo, se pelo menos uma observação for igual a zero $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}})$ torna-se inapropriado. A estimativa para ϕ nesse caso fica indeterminada. Como foi mencionado na Seção 2.4, McCullagh e Nelder (1989) sugerem substituir $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}})$ por

$$\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = 2\phi\{C(\mathbf{y}) + \sum_{i=1}^{n}\log(\hat{\mu}_i) + \sum_{i=1}^{n} y_i/\hat{\mu}_i\},$$

em que $C(\mathbf{y})$ é uma função arbitrária, porém limitada. Se a parte sistemática do modelo contém um intercepto o desvio acima fica dado por $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = 2\phi\{n + C(\mathbf{y}) + \sum_{i=1}^{n}\log(\hat{\mu}_i)\}$. Na prática ϕ deve ser estimado.

3.3.2 Técnicas de diagnóstico

O resíduo componente do desvio padronizado assume para os modelos gama a forma

$$t_{D_i} = \pm\sqrt{\frac{2\hat{\phi}}{1-\hat{h}_{ii}}}\{\log(\hat{\mu}_i/y_i) - (y_i-\hat{\mu}_i)/\hat{\mu}_i\}^{\frac{1}{2}},$$

em que $y_i > 0$ e h_{ii} é o i-ésimo elemento da diagonal principal da matriz $\mathbf{H} = \mathbf{W}^{\frac{1}{2}}\mathbf{X}(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{W}^{\frac{1}{2}}$ com $\omega_i = (d\mu_i/d\eta_i)^2/\mu_i^2$, $i = 1,\ldots,n$. Em particular quando há um intercepto em η_i o resíduo componente do desvio t_{D_i} assume a forma reduzida

$$t_{D_i} = \pm\sqrt{\frac{2\hat{\phi}}{1-\hat{h}_{ii}}}\{\log(\hat{\mu}_i/y_i)\}^{\frac{1}{2}}.$$

Estudos de simulação indicam que o resíduo t_{D_i} se aproxima da normalidade, particularmente para ϕ grande.

Quando a i-ésima observação é excluída a distância de Cook aproximada fica dada por

$$\mathrm{LD}_i = \frac{\hat{\phi}\hat{h}_{ii}}{(1-\hat{h}_{ii})^2}\frac{(y_i-\hat{\mu}_i)^2}{\hat{\mu}_i^2}.$$

Gráficos de t_{D_i} e $\hat{h}_{ii}$ contra os valores ajustados $\hat{\mu}_i$ como também gráficos de índices de LD_i são recomendados para a análise de diagnóstico.

3.4 Aplicações

3.4.1 Comparação de cinco tipos de turbina de avião

Na Tabela 3.1 são descritos os resultados de um experimento conduzido para avaliar o desempenho de cinco tipos de turbina de alta velocidade para motores de avião (ver Lawless 1982, p. 201). Foram considerados dez motores de cada tipo nas análises e foi observado para cada um o tempo (em unidades de milhões de ciclos) até a perda da velocidade. Esses dados estão disponíveis no arquivo **turbina.txt**.

Tabela 3.1
Tempo até a perda da velocidade de cinco tipos de turbina de avião.

Tipo de turbina				
Tipo I	Tipo II	Tipo III	Tipo IV	Tipo V
3,03	3,19	3,46	5,88	6,43
5,53	4,26	5,22	6,74	9,97
5,60	4,47	5,69	6,90	10,39
9,30	4,53	6,54	6,98	13,55
9,92	4,67	9,16	7,21	14,45
12,51	4,69	9,40	8,14	14,72
12,95	5,78	10,19	8,59	16,81
15,21	6,79	10,71	9,80	18,39
16,04	9,37	12,58	12,28	20,84
16,84	12,75	13,41	25,46	21,51

Denote por T_{ij} o tempo até a perda da velocidade para o j-ésimo motor de tipo i, $i = 1, \ldots, 5$ e $j = 1, \ldots, 10$. Na tabela abaixo são apresentadas as médias, desvios padrão e coeficientes de variação amostrais para os cinco tipos de turbina. Nota-se que os coeficientes de variação parecem variar menos do que os desvios padrão.

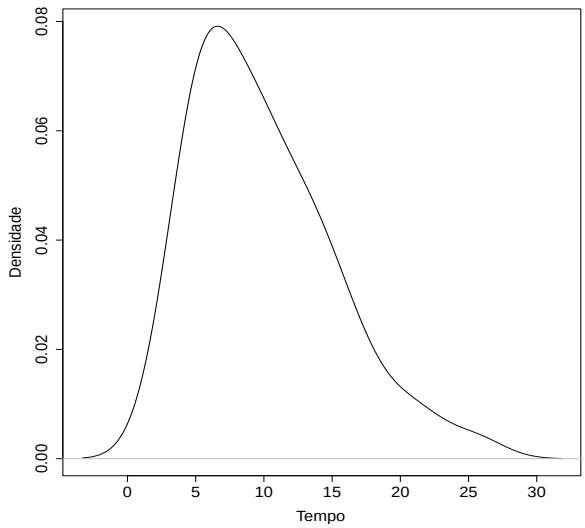

Figura 3.2: Densidade aproximada para o tempo até a perda da velocidade para todos os tipos de turbina de avião.

Estatística	Tipo I	Tipo II	Tipo III	Tipo IV	Tipo V
Média	10,69	6,05	8,64	9,80	14,71
D.Padrão	4,82	2,91	3,29	5,81	4,86
C. Variação	45,09%	48,10%	38,08%	59,29%	33,04%

Ignorando o tipo de turbina tem-se na Figura 3.2 a densidade aproximada para o tempo até a perda da velocidade. Assumindo que $T_{ij} \overset{\text{iid}}{\sim} \text{G}(\mu, \phi)$

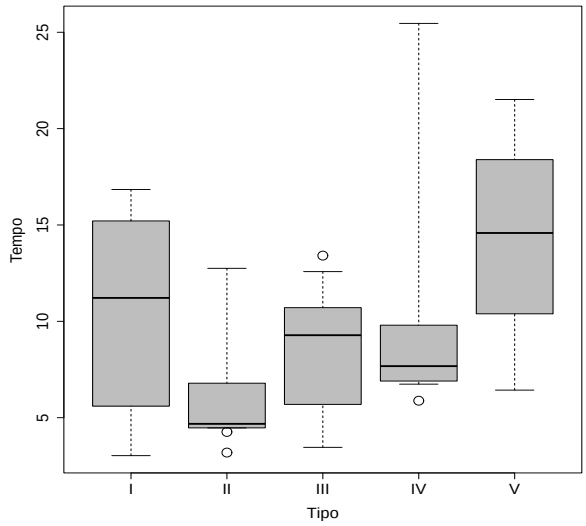

Figura 3.3: `Boxplots robustos` do tempo até a perda de velocidade para os 5 tipos de turbina de avião.

obtém-se as estimativas de máxima verossimilhança $\hat{\mu} = 9,98\ (0,73)$ e $\hat{\phi} = 4,01\ (0,77)$, confirmando pela estimativa de ϕ a assimetria à direita para a distribuição do tempo até a perda da velocidade. Contudo, pelos `boxplots robustos` (Hubert e Vandervierin, 2008) correspondentes aos tempos dos cinco grupos (ver Figura 3.3), nota-se distribuições mais assimétricas para os tipos II, III e IV e medianas e variabilidades distintas com algumas observações destoando como aberrantes. Assim, como o coeficiente de variação parece ser o menos heterogêneo dentre as medidas de variabilidade, sugere-se inicialmente distribuição gama de médias diferentes e coeficiente de variação constante para explicar o tempo médio até a perda da velocidade.

Assume-se então para o componente aleatório do modelo que $T_{ij} \overset{\text{ind}}{\sim} \text{G}(\mu_i, \phi)$, $i = 1, \ldots, 5$ e $j = 1, \ldots, 10$. A fim de facilitar as interpretações dos resultados ou mesmo fazer comparações com o modelo normal linear,

propõem-se um modelo gama com ligação identidade, sendo a parte sistemática dada por

$$\mu_i = \mu + \beta_i,$$

em que $\beta_1 = 0$ (casela de referência). Para ler os dados no R e ajustar o modelo gama deve-se aplicar os comandos

```
turbina = read.table("turbina.txt", header=TRUE)
attach(turbina)
tipo = factor(tipo)
fit1.turbina = glm(tempo ~ tipo, family=Gamma(link=identity))
summary(fit1.turbina)
require(MASS)
gamma.shape(fit1.turbina).
```

As estimativas de máxima verossimilhança ficam dadas por $\hat{\mu} = 10,693$ $(1,543)$, $\hat{\beta}_2 = -4,643$ $(1,773)$, $\hat{\beta}_3 = -2,057$ $(1,983)$, $\hat{\beta}_4 = -0,895$ $(2,093)$ e $\hat{\beta}_5 = 4,013$ $(2,623)$ indicando para o tipo II um tempo médio de sobrevivência significativamente menor do que o tipo I ao nível de 5%. Para o tipo V nota-se um tempo médio maior do que o tipo I, enquanto que os outros três tipos apresentam tempos médios pouco diferentes do tipo I. Esses resultados confirmam a análise descritiva apresentada na Figura 3.3. O desvio do modelo foi de $\mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 8,862 \times 5,804 = 51,43$, com 45 graus de liberdade, que leva a $\mathrm{P} = 0,236$ e indica um ajuste adequado.

Tem-se que $\mathrm{D}^*(\mathbf{y}; \bar{\mathbf{y}}) = 12,945$, logo o coeficiente de determinação fica dado por $\mathrm{R}^2 = 1 - \frac{8,862}{12,945} = 0,3154$. Levando-se em conta que é raro encontrar MLGs (exceto caso normal) com $\mathrm{R}^2 > 0,40$, tem-se indicação de um ajuste adequado.

A estimativa de máxima verossimilhança (erro padrão aproximado) do

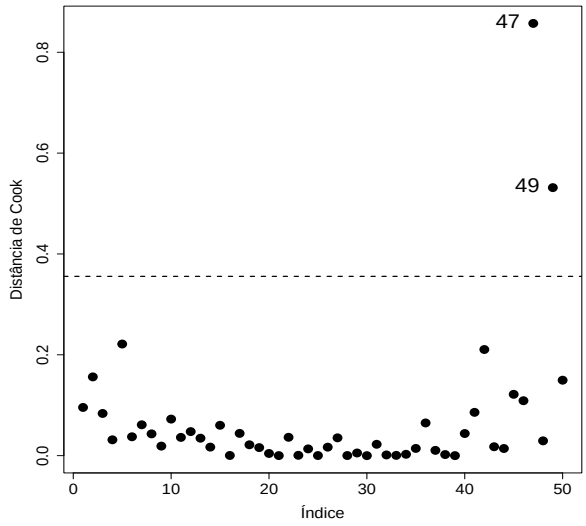

Figura 3.4: Gráfico da distância de Cook aproximada referente ao modelo gama ajustado ao dados sobre desempenho de turbinas de avião.

parâmetro de precisão é dada por $\hat{\phi} = 5,804\ (1,129)$, indicando que as distribuições dos tempos até a perda da velocidade não devem ser muito assimétricas. Pode-se tentar avaliar através de um teste apropriado se os indícios observados pelas estimativas individuais das médias são verificados conjuntamente. As hipóteses apropriadas são dadas por $\mathrm{H}_0 : \beta_4 = \beta_3 = 0$ contra $\mathrm{H}_1 : \beta_4 \neq 0$ e/ou $\beta_3 \neq 0$, que equivalem a testar o agrupamento dos tipos I, III e IV. Como $\hat{\phi}$ é relativamente alto pode-se aplicar a estatística F dada na Seção 1.7. Assim, sob H_0 obtém-se $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 9,091$ para 47 graus de liberdade e sob a hipótese alternativa $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 8,861$ para 45 graus de liberdade. A estatística F fica então dada por

$$\begin{aligned} \mathrm{F} &= \frac{(9,091 - 8,861)/2}{8,861/45} \\ &= 0,584, \end{aligned}$$

que leva a $\mathrm{P} = 0,562$, ou seja, pela não rejeição de H_0. As novas estimativas

são dadas por $\hat{\mu} = 9,71\ (0,81)$, $\hat{\beta}_2 = -3,66\ (1,19)$ e $\hat{\beta}_5 = 5,00\ (2,27)$. Obtém-se $\hat{\phi} = 5,66\ (1,10)$ e $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = 51,47$ para 47 graus de liberdade com $\mathrm{P} = 0,30$.

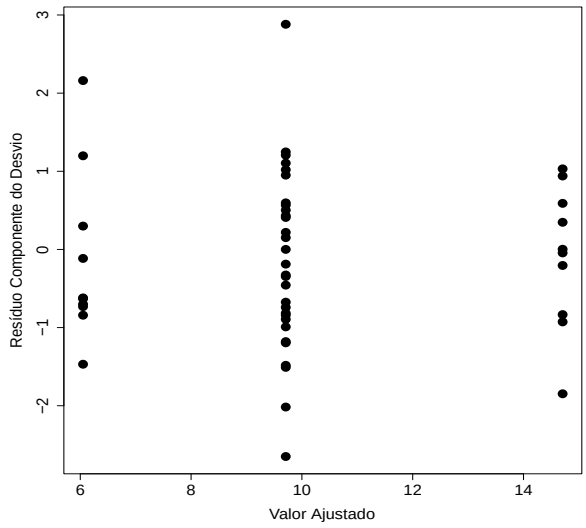

Figura 3.5: Resíduo componente do desvio contra o valor ajustado referente ao modelo gama ajustado aos dados sobre desempenho de turbinas de avião.

Na Figura 3.4 tem-se o gráfico de índices da distância de Cook aproximada. Nota-se um forte destaque para a observação #49 seguida da observação #47 que correspondem, respectivamente, aos valores 25,46 e 12,75 para o tempo até a perda da velocidade de um dos motores de tipo IV e tipo II. O valor 25,46, como é mostrado na Tabela 3.1, destoa dos demais tempos. A eliminação dessa observação aumenta a significância marginal de β_4, embora esse efeito continue não significativo a 10%. Não há mudanças inferenciais nos demais resultados.

O gráfico do resíduo componente do desvio contra o valor ajustado (Fi-

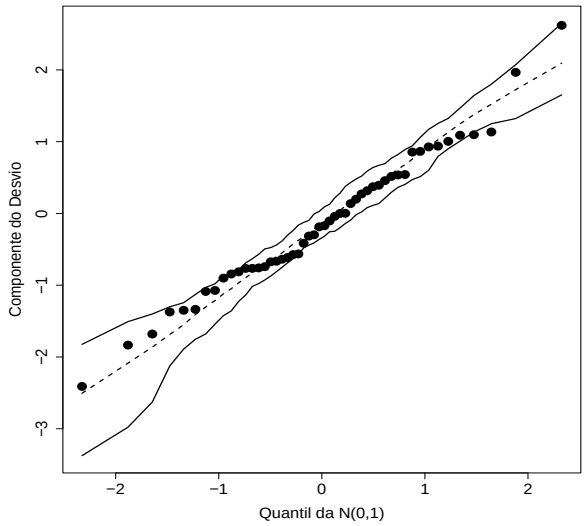

Figura 3.6: Gráfico normal de probabilidades referente ao modelo gama ajustado aos dados sobre desempenho de turbinas de avião.

gura 3.5) indica que a variabilidade foi controlada, ou seja, é adequado supor homogeneidade do coeficiente de variação nos 5 grupos. Já o gráfico normal de probabilidades com envelope para o resíduo componente do desvio é apresentado na Figura 3.6 e pode-se notar que não há indícios de afastamentos importantes da suposição de distribuição gama para os tempos até a perda da velocidade dos motores. Portanto, pode-se concluir neste exemplo que não há diferença significativa entre os tipos I, III e IV, enquanto os tipos II e V aparecem de forma significativa com o menor e maior tempo médio até a perda da velocidade, respectivamente.

3.4.2 Espinhel de fundo

O espinhel de fundo é definido como um método de pesca passivo, sendo utilizado em todo o mundo em operações de pesca de diferentes magnitudes,

da pesca artesanal a modernas pescarias mecanizadas. É adequado para capturar peixes com distribuição dispersa ou com baixa densidade, além de ser possível utilizá-lo em áreas irregulares ou em grandes profundidades. É um dos métodos que mais satisfazem às premissas da pesca responsável, com alta seletividade de espécies e comprimentos, alta qualidade do pescado, consumo de energia baixo e pouco impacto sobre o fundo oceânico. No arquivo **pesca.txt** estão parte dos dados de um estudo sobre a atividade das frotas pesqueiras de espinhel de fundo baseadas em Santos e Ubatuba no litoral paulista (vide Paula e Oshiro, 2001). A espécie de peixe considerada é o peixe-batata pela sua importância comercial e ampla distribuição espacial. Uma amostra de $n = 156$ embarcações foi analisada no período de 1995 a 1999 sendo 39 da frota de Ubatuba e 117 da frota de Santos.

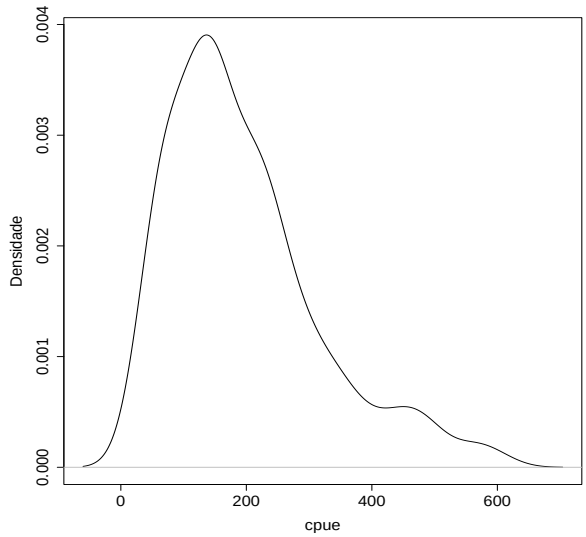

Figura 3.7: Densidade aproximada da cpue para todas as embarcações.

As variáveis consideradas para cada embarcação são as seguintes: frota

(Santos ou Ubatuba), ano (95 a 99), trimestre (1 ao 4), latitude (sul)[1] (de 23,25^o a 28,25^o), longitude (oeste)[2] (de 41,25^o a 50,75^o), dias de pesca, captura (quantidade de peixes batata capturados, em kg) e cpue (captura por unidade de esforço, kg/dias de pesca). Um dos objetivos desse estudo é tentar explicar a cpue méida pelas variáveis frota, ano, trimestre, latitude e longitude. Estudos similares realizados em outros países verificaram que é bastante razoável supor que a cpue tem distribuição assimétrica à direita, como é o caso da distribuição gama (vide, por exemplo, Goni et al.,1999).

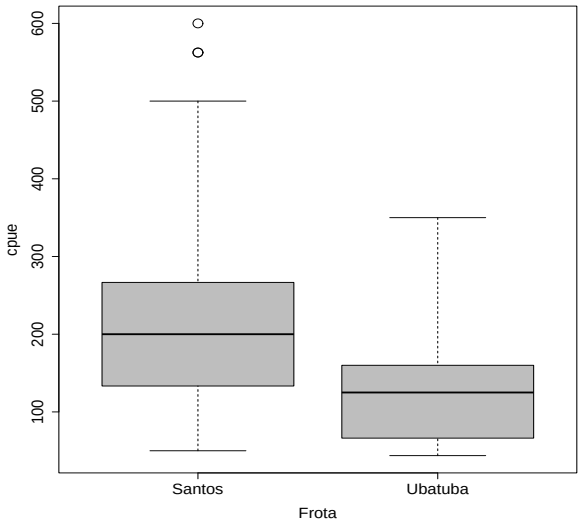

Figura 3.8: `Boxplots` da cpue segundo a frota.

Para ler o arquivo **pesca.txt** no `R` deve-se fazer o seguinte:

```
pesca = read.table("pesca.txt", header=TRUE)
frota = factor(frota)
ano = factor(ano)
```

[1]distância ao Equador medida ao longo do meridiano de Greenwich
[2]distância ao meridiano de Greenwich medida ao longo do Equador

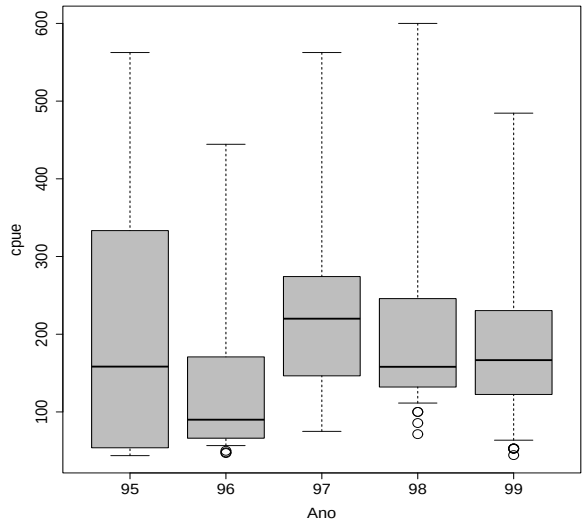

Figura 3.9: `Boxplots` da cpue segundo o ano.

`trimestre = factor(trimestre).`

Tabela 3.2
Medidas resumo para a distribuição da cpue segundo a frota e o ano.

Frota	Estatística	95	96	97	98	99
	Média	229,37	193,19	262,67	210,29	197,22
Santos	D.Padrão	148,07	132,55	153,60	122,95	103,45
	C. Variação	64,55%	68,61%	58,48%	58,44%	52,45 %
	n	19	8	17	27	46
	Média	47,08	96,09	210,56	174,43	140,85
Ubatuba	D. Padrão	4,73	59,19	77,51	99,16	71,59
	C. Variação	10,05%	61,60 %	36,81%	56,85%	50,83%
	n	3	12	6	5	13

Antes de propor um modelo para tentar explicar a cpue média pelas variáveis explicativas, será apresentada uma análise descritiva dos dados. Na Figura 3.7 tem-se a distribuição da cpue para todas as embarcações e pode-se

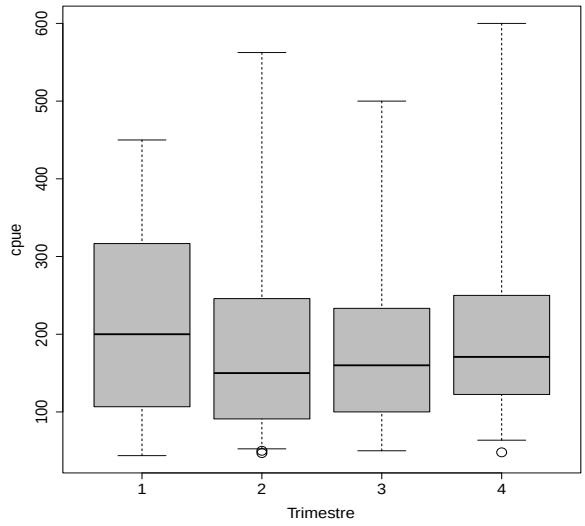

Figura 3.10: `Boxplots` da cpue segundo o trimestre.

notar uma assimetria acentuada à direita, confirmando constatações de estudos anteriores. Já nas Figuras 3.8, 3.9 e 3.10 são apresentados os `boxplots robustos` da cpue segundo os fatores frota, ano e trimestre, respectivamente. Nota-se uma superioridade da frota de Santos em relação à frota de Ubatuba, porém poucas diferenças entre os níveis dos fatores ano e trimestre, embora o ano de 97 tenha uma mediana um pouco superior aos demais anos.

Pela Figura 3.11 nota-se que a frota de Santos prefere latitudes e longitudes maiores do que a frota de Ubatuba. Pelos diagramas de dispersão entre cpue e latitude e cpue e longitude, apresentados na Figura 3.12, há indícios de um ligeiro crescimento da cpue com a latitude, porém não está bem definida a tendência da cpue com a longitude.

Na Tabela 3.2 são apresentadas as médias, desvios padrão e coeficientes de variação amostrais para as frotas de Santos e Ubatuba referentes ao período 95-99. Nota-se que o coeficiente de variação é mais homogêneo na frota de

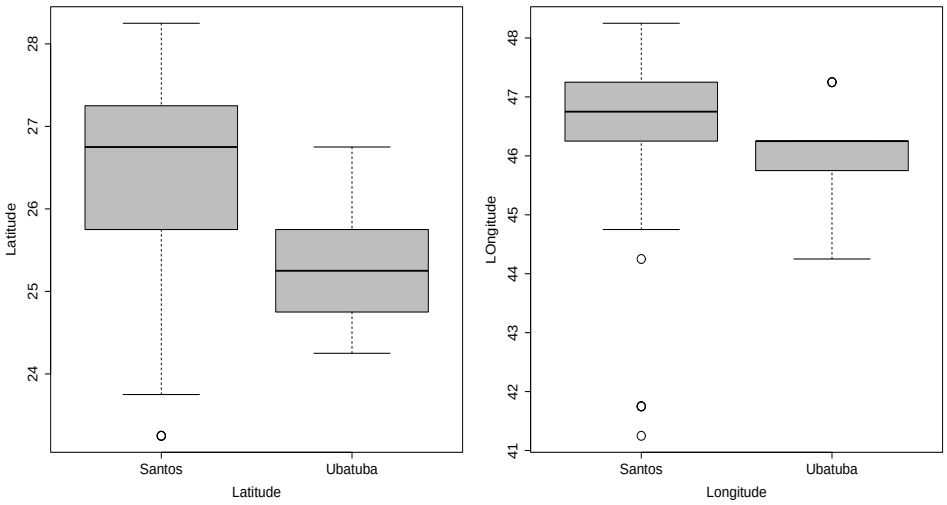

Figura 3.11: `Boxplots` da latitude e longitude segunda a frota.

Santos e em geral, exceto para os anos de 95 e 97 para a frota de Ubatuba, fica entre 50% e 70%. Porém deve-se levar em conta que para a frota de Ubatuba as amostras são pequenas. Como todas essas análises são marginais, somente através de um modelo apropriado é que será possível conhecer o efeito de cada variável explicativa na presença das demais na variação da cpue média. Será então assumido inicialmente um modelo de regressão com resposta gama modelando-se a média com coeficiente de variação constante.

Definindo então $Y_{ijk\ell}$ como sendo a cpue observada para a i-ésima embarcação da j-ésima frota, (Santos, $j = 1$; Ubatuba $j = 2$), no k-ésimo ano e ℓ-ésimo trimestre ($k, \ell = 1, 2, 3, 4$), supor que $Y_{ijk\ell} \overset{\text{ind}}{\sim} \text{G}(\mu_{ijk\ell}, \phi)$ com parte sistemática dada por

$$\log(\mu_{ijk\ell}) = \alpha + \beta_j + \gamma_k + \theta_\ell + \delta_1 \text{Latitude}_{ijk\ell} + \delta_2 \text{Longitude}_{ijk\ell}, \qquad (3.2)$$

em que β_j, γ_k e θ_ℓ denotam, respectivamente, os efeitos da j-ésima frota, k-ésimo ano e ℓ-ésimo trimestre. Como está sendo assumindo parametrização

casela de referência tem-se as restrições $\beta_1 = 0$, $\gamma_1 = 0$ e $\theta_1 = 0$. Latitude$_{ijk\ell}$ e longitude$_{ijk\ell}$ denotam, respectivamente, a latitude e longitude da i-ésima embarcação da frota j no k-ésimo ano e trimestre ℓ.

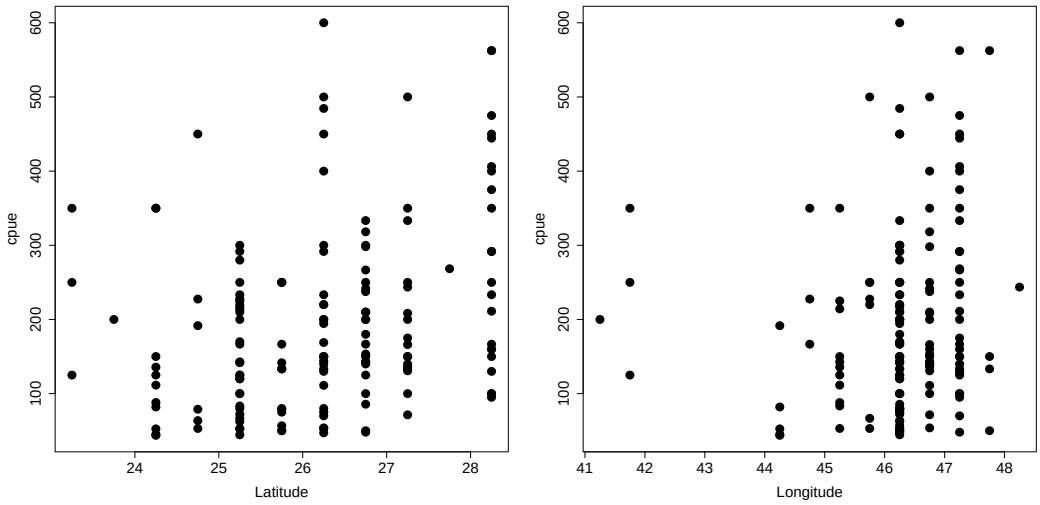

Figura 3.12: Diagramas de dispersão da cpue contra latitude e contra longitude.

Ajustando aos dados o modelo gama com parte sistemática dada por (3.2) e aplicando o método de Akaike (vide Seção 2.9) retira-se o fator trimestre, permanecendo no modelo os fatores frota e ano além das variáveis quantitativas latitude e longitude. Para ajustar o modelo e selecionar as variáveis explicativas deve-se aplicar os seguintes comandos:

```
attach(pesca)
fit1.pesca = glm(cpue ~ frota + ano + trimestre + latitude +
longitude, family=Gamma(link=log))
summary(fit1.pesca)
require(MASS)
```

```
stepAIC(fit1.pesca).
```

Tabela 3.3
Estimativas dos parâmetros referentes ao modelo gama ajustado aos dados sobre espinhel de fundo.

Efeito	Estimativa	E/E.Padrão
Constante	6,898	3,00
Latitude	0,204	2,81
Longitude	-0,150	-1,97
Frota-Ubatuba	-1,359	-3,68
Ano96	-0,064	-0,26
Ano97	0,141	0,74
Ano98	-0,043	-0,25
Ano99	-0,009	-0,06
FrotaUb*Ano96	0,806	1,77
FrotaUb*Ano97	1,452	3,20
FrotaUb*Ano98	1,502	3,32
FrotaUb*Ano99	1,112	2,76
ϕ	3,67	9,17

O procedimento `stepAIC` assume que o parâmetro ϕ é constante, ou seja, não muda de um modelo para o outro. Como isso, em geral, não é satisfeito deve-se aplicar algum procedimento alternativo a fim de confirmar o modelo escolhido pelo método `AIC`. Então foi aplicado o mesmo procedimento através da estatística da razão de versossimilhanças, confirmando-se a retirada do fator trimestre.

O teste da razão de verossimilhanças para incluir a interação entre os dois fatores que permaneceram no modelo, frota e ano, foi de $\xi_{RV} = 14,26$ para 4 graus de liberdade, obtendo-se P $= 0,0065$. Portanto, a interação será incluída no modelo. As estimativas do modelo final que inclui os efeitos principais latitude, longitude, frota e ano além da interação entre ano e frota são apresentadas na Tabela 3.3. O desvio do modelo foi de $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = 162,66$

com 144 graus de liberdade e $\mathrm{P} = 0,14$, indicando um modelo bem ajustado. Tem-se ainda que $\mathrm{R}^2 = 1 - \frac{\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}})}{\mathrm{D}(\mathbf{y};\bar{\mathbf{y}})} = \frac{49{,}464}{59{,}362} = 0,1667$.

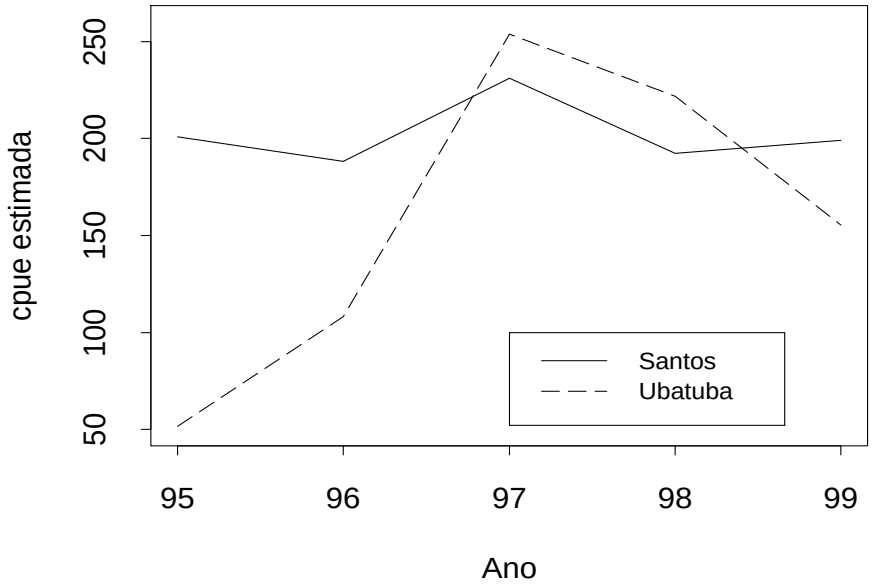

Figura 3.13: Estimativas da cpue média para as frotas de Santos e Ubatuba segundo o ano de operação fixando-se a latitude em 26^o e a longitude em 46^o através do modelo gama.

Nota-se que à medida que aumenta a latitude aumenta a cpue, ocorrendo tendência contrária à medida que aumenta a longitude. Logo, para latitudes altas e longitudes baixas (dentro dos limites amostrais), espera-se valores maiores para a captura por unidade de esforço. Com relação à frota e ao ano, como foi incluída interação entre esses fatores, a interpretação das estimativas deve ser feita com um pouco mais de cuidado. Para isso, é exibido na Figura 3.13 os valores esperados da cpue fixando latitude e longitude nos valores, respectivamente, 26^o e 46^o. Nota-se que até 96 os valores preditos para a frota de Ubatuba são bem menores do que os valores preditos para a frota de

Santos. Contudo, a partir de 97 as diferenças entre os valores preditos para as duas frotas diminuem. Os valores preditos para a frota de Santos variam pouco no período 95-99, diferentemente dos valores preditos para a frota de Ubatuba.

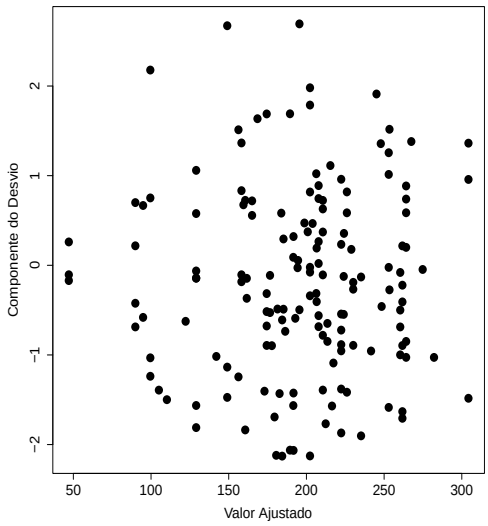

Figura 3.14: Gráfico do resíduo componente do desvio contra o valor ajustado referente ao modelo gama ajustado aos dados sobre espinhel de fundo.

Na Figura 3.14 tem-se o gráfico do resíduo componente do desvio contra o valor ajustado, indicando que a variabilidade foi controlada, ou seja, é razoável supor coeficiente de variação constante. No gráfico da distância de Cook aproximada (Figura 3.15) Três observações aparecem como possivelmente influentes, as embarcações #8, #17 e #52. A retirada de cada embarcação individualmente não muda a inferência, porém a retirada da observação #17 aumenta a significância da latitude e longitude. A embarcação #17 é da frota de Santos, obteve uma cpue de 450 (valor médio 195,5) numa latitude de 24,75^o (valor médio 26,22^o) e longitude de 46,25^o (valor médio

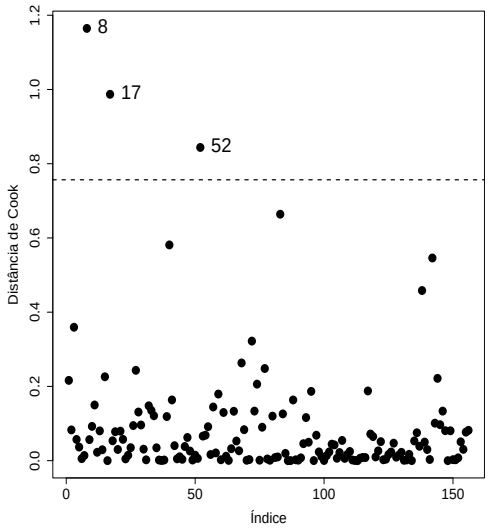

Figura 3.15: Gráfico da distância de Cook aproximada referente ao modelo gama ajustado aos dados sobre espinhel de fundo.

46,26^o) no ano de 99. Esperava-se para essa embarcação um valor menor para a cpue levando-se em conta os valores da latitude e longitude. Trata-se portanto de uma embarcação atípica. O gráfico normal de probabilidades com envelope gerado (Figura 3.16) não apresenta indícios fortes de que a distribuição gama seja inadequada para explicar a cpue.

3.4.3 Aplicação em seguros

A fim de ilustrar uma aplicação na área de seguros, considere parte dos dados descritos em de Jong e Heller (2008, pgs. 14-15) referentes aos valores pagos de seguros individuais (em dólares australianos) por danos com acidentes pessoais no período de julho de 1989 a junho de 1999. As análises serão restritas ao período de janeiro de 1998 a junho de 1999, um total de 769 seguros pagos. Além do valor pago ao segurado serão consideradas as se-

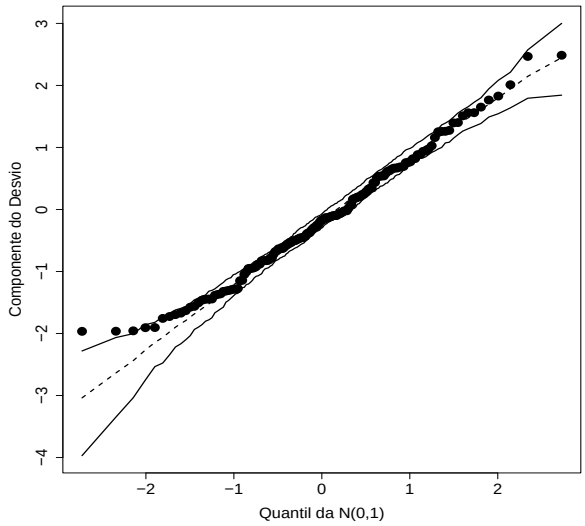

Figura 3.16: Gráfico normal de probabilidades referente ao modelo gama ajustado aos dados sobre espinhel de fundo.

guintes variáveis explicativas: `legrep`, representação legal (0: não, 1: sim) e `optime`, tempo operacional para pagamento do seguro. Essa última variável assume valores no intervalo (0, 100) e por exemplo um valor 23 significa que 23% dos seguros foram pagos antes do seguro em análise. Como está sendo considerado apenas parte dos dados (referentes aos últimos 18 meses), os valores de optime irão variar de 0,1 a 31,9. O subconjunto de dados analisado está descrito no arquivo **insurance.txt**.

Na Figura 3.17 tem-se o diagrama de dispersão entre o logaritmo do valor pago e o tempo operacional para os grupos sem representação legal e com representação legal. Nota-se para as apólices sem representação legal um crescimento aproximadamente quadrático do logaritmo do valor pago com o tempo operacional, contudo a variablidade parece ser maior para valores baixos do tempo operacional. Já para as apólices com representação legal

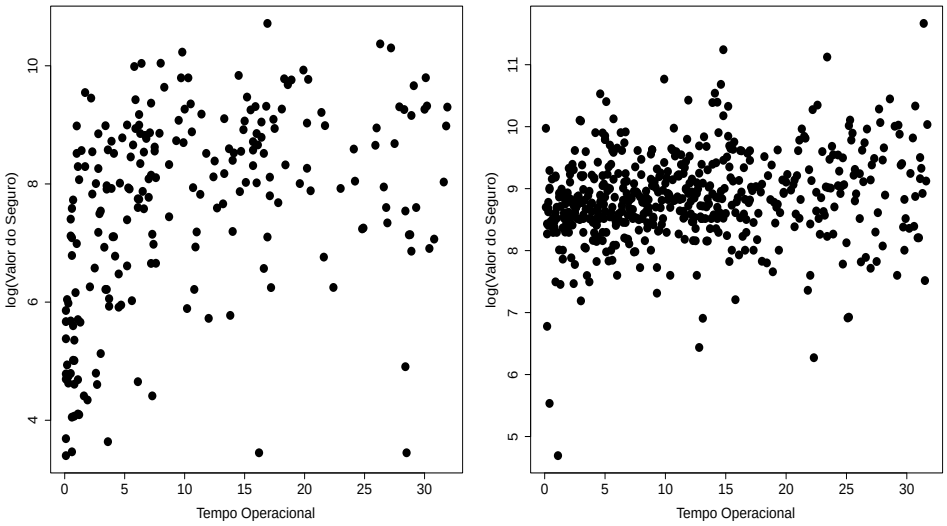

Figura 3.17: Diagrama de dispersão entre o valor pago de seguro e o tempo operacional para os grupos sem representação legal (esquerda) e com representação legal (direita).

nota-se que o logaritmo do valor pago cresce linearmente com o tempo operacional enquanto a variabilidade se mantém aproximadamente constante. Nota-se também que os valores pagos de seguro são em geral maiores para o grupo com representação legal.

Na Figura 3.18 tem-se a distribuição aproximada do valor pago de seguro para os dois grupos, sem representação legal e com representação legal. Em ambos os gráficos pode-se notar que a distribuição é fortemente assimétrica à direita, sugerindo distribuições gama ou normal inversa para explicar o valor pago de seguro.

Denote por Y_{ij} o valor pago de seguro para o j-ésimo indivíduo do i-ésimo grupo ($i = 0$, sem representação legal e $i = 1$ com representação legal) e $j = 1, \ldots, n_i$ sendo $n_0 = 227$ e $n_1 = 542$. Conforme sugerido pela Figura 3.18 será assumido inicialmente $Y_{ij} \overset{\text{ind}}{\sim} \text{G}(\mu_{ij}, \phi_i)$ com componentes

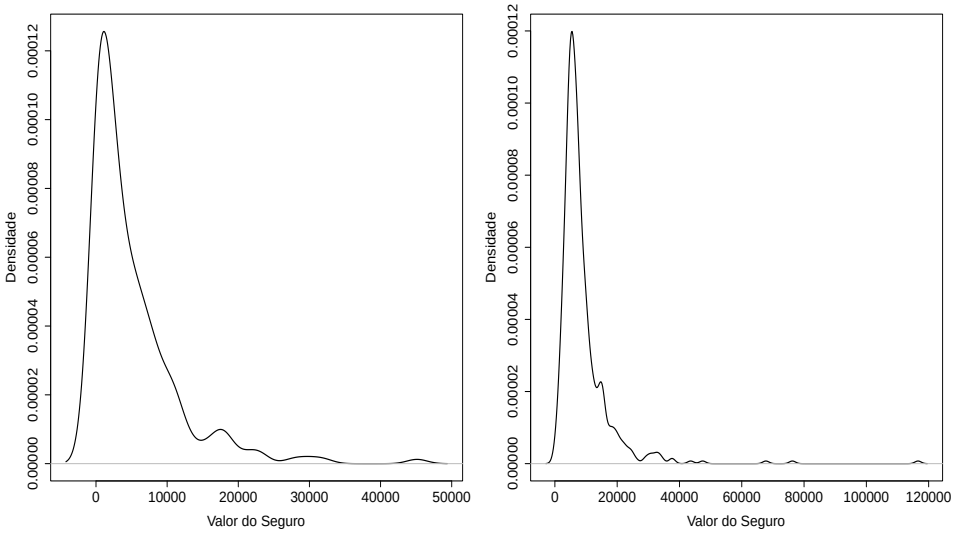

Figura 3.18: Distribuição do valor de seguro para os grupos sem representação legal (esquerda) e com representação legal (direita).

sistmáticos dados por

$$\begin{aligned} \log(\mu_{0j}) &= \alpha_0 + \beta_{10}\texttt{optime}_j + \beta_{20}\texttt{optime}_j^2 \text{ e} \\ \log(\mu_{1j}) &= \alpha_1 + \beta_{11}\texttt{optime}_j. \end{aligned}$$

Para ler os dados no R e ajustar o modelo deve-se aplicar os comandos

```
insurance = read.table("insurance.txt", header=TRUE)
attach(insurance)
fit0.insurance = glm(amount0 ~ optime0 + I(optime0^2),
family=Gamma(link=log))
summary(fit0.insurance)
fit1.insurance = glm(amount1 ~ optime1, family=Gamma(link=log))
summary(fit1.insurance)
require(MASS)
```

```
gamma.shape(fit0.insurance)
gamma.shape(fit1.insurance).
```

Tabela 3.4
Estimativas dos parâmetros referentes aos modelos com resposta gama ajustados aos dados sobre seguro.

Parâmetro	Estimativa	E/E.Padrão
α_0	7,223	44,13
β_{10}	0,204	6,72
β_{20}	-0,005	-5,08
ϕ_0	0,779	12,55
α_1	8,805	140,50
β_{11}	0,023	5,48
ϕ_1	2,225	17,66

As estimativas dos parâmetros dos modelos propostos, que foram ajustados separadamente, são descritas na Tabela 3.4. Nota-se pelas estimativas que as tendências observadas na Figuras 3.17 foram confirmadas de forma significativa. Contudo, pelos gráficos normais de probabilidade (Figura 3.19) nota-se indícios de afastamentos da distribuição gama para o valor pago de seguro, principalmente para o grupo com representação legal. Para o grupo sem representação legal nota-se que os menores valores do seguro foram superestimados pelo modelo.

Os desvios dos dois modelos foram, respectivamente, de $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = 0,779 \times 347,15 = 270,70$ com 224 graus de liberdade e $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = 2,225 \times 261,45 = 581,73$ com 540 graus de liberdade. Embora as estimativas de ϕ_0 e ϕ_1 sejam relativamente pequenas, há indícios pelos valores dos desvios de que os modelos não estão bem ajustados. Os coeficientes de determinação ficam, respectivamente, dados por $\mathrm{R}^2 = 1 - \frac{347.15}{419,59} = 0,173$ e $\mathrm{R}^2 = 1 - \frac{261,45}{283,83} = 0,079$, confirmando um ajuste mais adequado para o grupo sem representação legal.

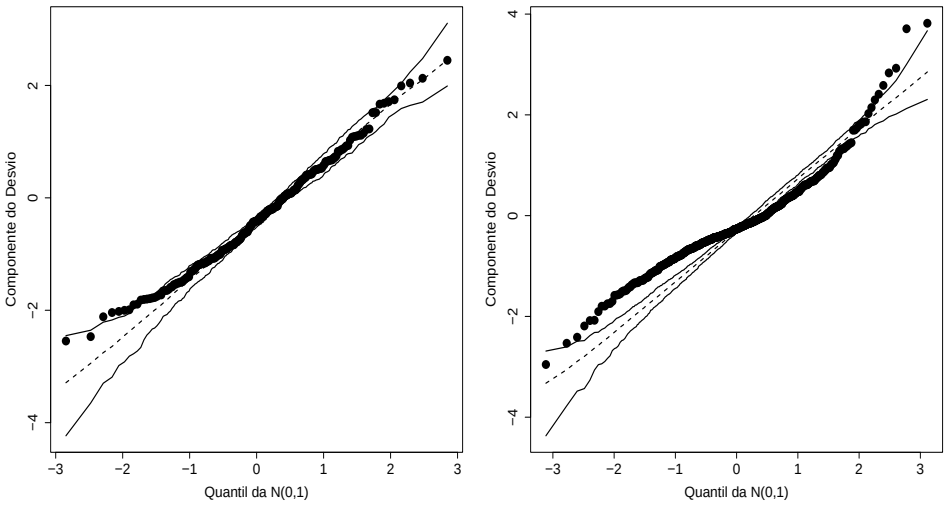

Figura 3.19: Gráfico normal de probabilidades para os modelos com resposta gama ajustados aos dados de seguro para os grupos sem representação legal (esquerda) e com representaçõa legal (direita).

Analisando os gráficos da distância de Cook e resíduo componente do desvio contra o valor ajustado (Figrua 3.20) apenas para o grupo sem representação legal, nota-se que não há indícios de observações aberrantes, contudo algumas observações aparecem como possivelmente influentes. Essas observações em geral correspondem a valores altos para o valor pago de seguro. A eliminação das observações destacadas não muda a inferência, todos os coeficientes continuam altamente significativos.

Para o grupo com representação legal a utilização de outras ligações ou mesmo outras distribuições são alternativas a fim de tentar melhorar a qualidade do ajuste. Paula et al. (2012) comparam ajustes de modelos com resposta gama com modelos com respostas Birnbaum-Saunders (BS) e Birnbaum-Saunders-t (BS-t) para explicar o valor pago de seguro para o grupo com representação legal, obtendo um ajuste satisfatório com o modelo

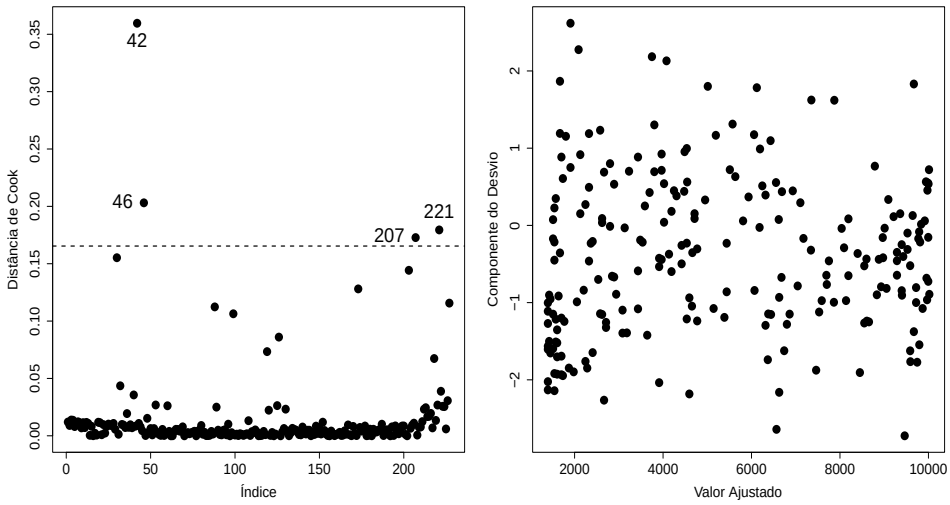

Figura 3.20: Gráficos de diagnóstico para o modelo com resposta gama ajustado aos dados de seguro para o grupo sem representação legal.

BS-t. Essa distribuição acomoda melhor as observações aberrantes que correspondem a valores altos de seguro. Recentemente, Cardozo et al. (2022) ajustaram os dados do valor de seguro pago para o grupo sem representação legal através de modelo log-linear com componente aditivo ao invés de componente quadrático e resposta gama generalizada, obtendo um ajuste mais adequado para explicar o valor pago de seguro..

3.5 Elasticidade

O modelo log-linear com resposta gama pode ser utilizado para a estimação da elasticidade entre a demanda de um produto e seu preço unitário. Como ilustração, supor que Y denota a demanda e X o preço unitário. É usual em Econometria (ver, por exemplo, Gujarati, 2006, Seção 6.4) assumir que

$$Y = \beta_1 x^{\beta_2} e^u, \tag{3.3}$$

em que u é um erro aleatório, em geral assumido $\mathrm{N}(0, \sigma^2)$. Isso implica em e^u seguir distribuição log-normal de média $e^{\sigma^2/2}$ e variância $e^{\sigma^2}(e^{\sigma^2} - 1)$. Em vez de uma log-normal pode-se assumir, alternativamente, outra distribuição com resposta positiva. Por exemplo, $e^u \sim \mathrm{G}(1, \phi)$. Assim, tem-se que a média de Y dado x fica dada por

$$\mu(x) = \beta_1 x^{\beta_2},$$

ou seja $\log(\mu(x)) = \log(\beta_1) + \beta_2 \log(x)$, um MLG com resposta gama e ligação logarítmica.

Para entender a elasticidade entre a demanda e o preço do produto, supor que o preço aumente $r \times 100\%$ de modo que o novo preço seja dado por $x_N = (1 + r)x$, para $0 < r < 1$. O novo valor esperado para a demanda fica dado por

$$\mu(x_N) = \beta_1 x_N^{\beta_2}$$

e a razão entre as demandas médias assume a forma

$$\begin{aligned} \frac{\mu(x_N)}{\mu(x)} &= e^{\beta_2 \log(1+r)} \\ &= (1 + r)^{\beta_2}. \end{aligned}$$

Para r pequeno tem-se a aproximação

$$\frac{\mu(x_N)}{\mu(x)} \cong (1 + r\beta_2),$$

de modo que se o preço aumentar 1% $(r = 0,01)$ a demanda aumenta β_2%, ou seja,

$$\frac{\mu(x_N)}{\mu(x)} = \left(1 + \frac{\beta_2}{100}\right).$$

O parâmetro β_2 é conhecido como elasticidade entre a demanda e o preço do produto.

3.5.1 Modelo de Cobb-Douglas

O modelo (3.3) pode ser estendido para duas ou mais variáveis explicativas as quais poderão representar outros tipos de preço ou mesmo algum tipo de insumo. Em particular, o modelo de Cobb-Douglas (ver, por exemplo, Gujarati, Exemplo 7.3) considera a seguinte equação para explicar a demanda de um produto pelos insumos de mão de obra e capital:

$$Y = \beta_1 x_2^{\beta_2} x_3^{\beta_3} e^u, \tag{3.4}$$

em que Y denota a demanda, x_2 o valor do insumo de mão de obra, x_3 o valor do insumo de capital e u o erro aleatório. Para x_3 fixado (x_2 fixado) o parâmetro β_2 (β_3) mede a elasticidade parcial entre a demanda e o insumo de mão de obra (capital). A soma $\beta_2 + \beta_3$ mede os retornos de escala, ou seja, se $\beta_2 + \beta_3 = 1$ significa que os retornos são proporcionais, dobrando o uso de insumos a demanda esperada aumenta duas vezes, triplicando os insumos há aumento de três vezes para a demanda esperada, e assim por diante. Se $\beta_2 + \beta_3 < 1$ os retornos de escala serão menores, dobrando os insumos espera-se demanda menor do que o dobro, e se $\beta_2 + \beta_3 > 1$ os retornos de escala serão maiores, dobrando os insumos espera-se que a demanda aumente mais que duas vezes.

Para mostrarmos esses resultados suponha que os novos insumos de mão de obra e de capital sejam dados por $x_{1N} = rx_1$ e $x_{2N} = rx_2$, ou seja, aumentam r vezes. Assim, a nova demanda esperada será dada por

$$\begin{aligned} \mu(x_{1N}, x_{2N}) &= \beta_1 (rx_2)^{\beta_2} (rx_3)^{\beta_3} \\ &= r^{(\beta_2+\beta_3)} \beta_1 x_2^{\beta_2} x_3^{\beta_3} \\ &= r^{\beta_2+\beta_3} \mu(x_1, x_2), \end{aligned}$$

em que $\mu(x_1, x_2)$ é a demanda esperada inicial. Logo, se $\beta_2 + \beta_3 = 1$ então $\mu(x_{1N}, x_{2N}) = r\mu(x_1, x_2)$, ou seja, a demanda esperada aumenta r vezes. Por

outro lado, se $\beta_2 + \beta_3 < 1$ tem-se que $\mu(x_{1N}, x_{2N}) < r\mu(x_1, x_2)$, ou seja, a demanda esperada aumenta menos que r vezes e se $\beta_2 + \beta_3 > 1$ tem-se que a demanda esperada aumenta mais que r vezes, $\mu(x_{1N}, x_{2N}) > r\mu(x_1, x_2)$.

Obviamente que existem várias distribuições candidatas para explicar e^u, sendo as distribuições gama e normal inversa as candidatas naturais na classe dos MLGs. Pode-se também assumir que $\log(u)$ tenha distribuição normal. Contudo, somente através de uma análise de diagnóstico é que pode-se avaliar a adequação de cada distribuição.

3.5.2 Aplicação

Como ilustração conside um experimento aleatorizado descrito em Griffiths et al.(1993, Seção 11.8.1c) em que a produtividade de milho (libras/acre) é estudada segundo várias combinações de nitrogênio e fosfato (40, 80, 120, 160, 200, 240, 280 e 320 libras/acre). Os dados estão descritos no arquivo **milho.txt**. Na Figura 3.21 tem-se os diagramas de dispersão entre a produtividade de milho e as quantidades de nitrogênio e fosfato, respectivamente, e pode-se notar nessas figuras há indícios de uma tendência crescente da produtividade com o aumento dos insumos. Nota-se também um aumento da variabilidade com o aumento das quantidades de nitrogênio e fostato, sugerindo que a suposição de distribuição gama ou normal inversa para $\log(u)$ no modelo de Cobb-Douglas pode levar a um ajuste adequado. Denote por Y_i a produtividade de milho dada a combinação (x_{1i}, x_{2i}) de nitrogênio e fosfato correspondente à i-ésima condição experimental e supor que $Y_i \overset{\text{ind}}{\sim} \text{G}(\mu_i, \phi)$ com parte sistemática dada por $\log(\mu_i) = \alpha + \beta_1 \log(x_{1i}) + \beta_2 \log(x_{2i})$, para $i = 1, \ldots, 30$. A leitura dos dados em R e os comandos para o ajuste do modelo gama log-linear são dados abaixo

```
milho = read.table(‘‘milho.txt’’, header=TRUE)
```

```
summary(milho)
attach(milho)
fit.milho = glm(produtividade ~ log(nitrogenio) + log(fostato),
family Gamma(link=log))
summary(fit.milho).
```

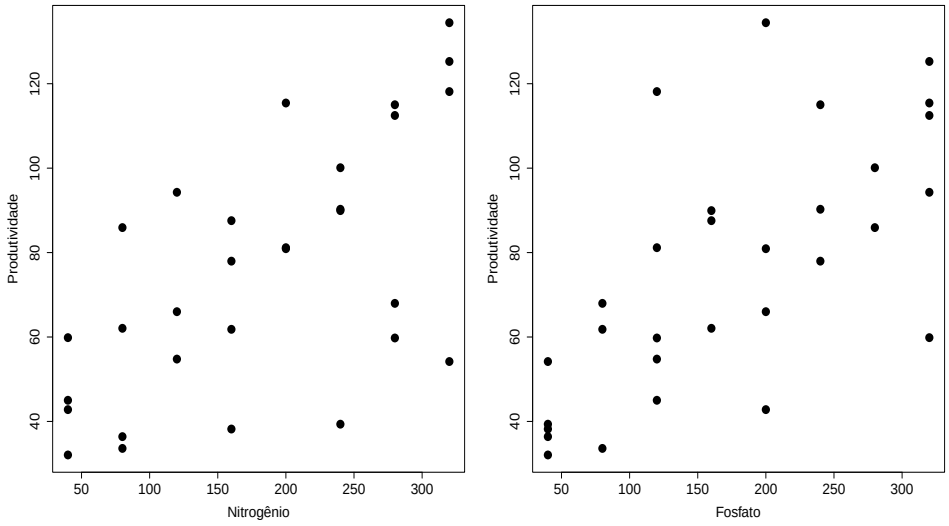

Figura 3.21: Diagramas de dispersão entre a quantidade de nitrogênio e a produtividade de milho (esquerda) e a quantidade de fosfato e a produtividade de milho (direita).

As estimativas são descritas na Tabela 3.5 e como pode-se notar os coeficientes são altamente significativos, confirmando as tendências observadas na Figura 3.21. Na Figura 3.22 tem-se os gráficos do resíduo componente do desvio contra o valor ajustado, indicando que a variabilidade está controlada, e da distância de Cook em que duas observações aparecem como possivelmente influentes. A eliminação de cada observação individualmente não altera de forma substancial os coeficentes estimados nem muda a inferência, ambos continuam altamente significativos. Porém, o intercepto fica

significativo a 5% com a eliminação da observação #28, indicando que essa observação pode estar mascarando o efeito do intercepto. A estimativa da precisão (relativamente alta) indica que um modelo com erros log-normal também poderia levar a um ajuste adequado. Já o gráfico normal de probabilidades (Figura 3.23) indica que a suposição de erros gama leva a um ajuste adequado não havendo observações aberrantes. A principal diferença em assumir erros gama ao invés de erros log-normal é a possibilidade de maior controle da variabilidade.

Tabela 3.5
Estimativas dos parâmetros referentes ao modelo de Cobb-Douglas ajustado ao dados sobre produtividade de milho.

Parâmetro	Estimativa	E/E.Padrão
α	0,469	1,67
β_1	0,350	8,30
β_2	0,410	10,07
ϕ	46,59	11,99

A fim de verificar como ocorrem os retornos de produtividade de milho com as aplicações de fosfato e nitrogênio será obtida a estimativa intervalar para $\beta_1 + \beta_2$. Deve-se obter inicialmente

$$\begin{aligned} \hat{\text{Var}}(\hat{\beta}_1 + \hat{\beta}_2) &= \text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_2) + 2\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) \\ &= 0,001776 + 0,001656 - 2 * 0,000145 \\ &= 0,003142. \end{aligned}$$

Essas quantidades são obtidas através do comando

```
vcov(fit.milho).
```

Assim uma estimativa intervalar de coeficiente de confiança de 95% fica dada por $[0,35 + 0,41 \pm 1,96 * \sqrt{0,003142}] = [0,65; 0,87]$ que não cobre o valor

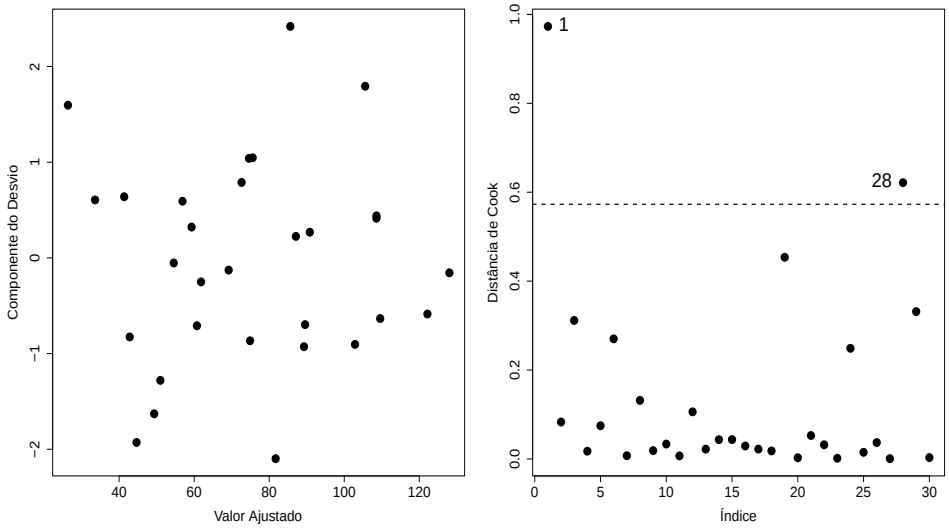

Figura 3.22: Gráfico do resíduo componente do desvio contra o valor ajustado (esquerda) e distância de Cook (direita) referentes ao modelo de Cobb-Douglas ajustado aos dados do experimento sobre produtividade de milho.

1,0. Portanto, dobrando as aplicações de insumos de nitrogênio e fosfato os retornos esperados de produtividade devem aumentar menos do que duas vezes.

3.6 Distribuição normal inversa

Supor que Y é uma variável aleatória com distribuição normal inversa de média μ e parâmetro de dispersão ϕ^{-1}. Denota-se $Y \sim \text{NI}(\mu, \phi)$, cuja função densidade de probabilidade é expressa na forma

$$\begin{aligned} f(y;\mu,\phi) &= \sqrt{\frac{\phi}{2\pi y^3}}\exp\left\{-\frac{\phi(y-\mu)^2}{2\mu^2 y}\right\} \\ &= \exp\left[\phi\left\{-\frac{y}{2\mu^2}+\frac{1}{\mu}\right\}-\frac{1}{2}\left\{\log(2\pi y^3/\phi)+\frac{\phi}{y}\right\}\right], \end{aligned}$$

em que $y > 0$, $\mu > 0$ e $\phi > 0$.

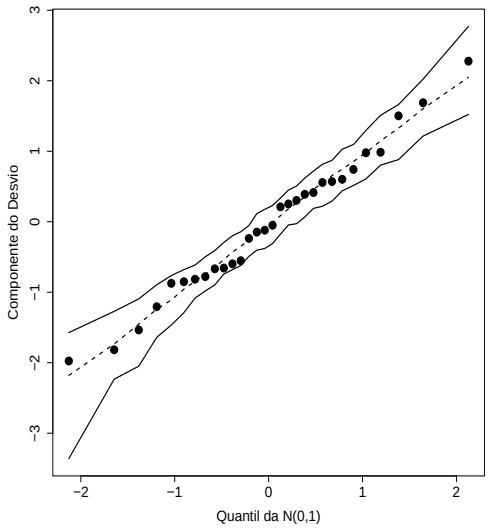

Figura 3.23: Gráfico normal de probabilidades referente ao modelo de Cobb-Douglas ajustado aos dados do experimento sobre produtividade de milho.

Na Figura 3.24 tem-se a densidade da distribuição normal inversa variando o parâmetro de precisão para μ fixado. Nota-se que para valores pequenos do parâmetro de precisão a distribuição normal inversa é fortemente assimétrica à direita, contudo à medida que ϕ aumenta a distribuição normal inversa fica mais simétrica em torno da média. Pode-se mostrar que à medida que ϕ aumenta Y se aproxima de uma distribuição normal de média μ e variância $\mu^3\phi^{-1}$. Logo, similarmente à distribuição gama, a normal inversa torna-se atrativa para o estudo de variáveis aleatórias assimétricas e também simétricas em que a variância depende de forma cúbica da média. Uma discussão sobre as suposições teóricas para a construção da distribuição normal inversa pode ser encontrada, por exemplo, em Leiva et al.(2009, Cap. 2).

A função de sobrevivência da distribuição normal inversa de média μ e parâmetro de dispersão ϕ^{-1} (ver, por exemplo, Collett, 2003, pp. 198-199) é

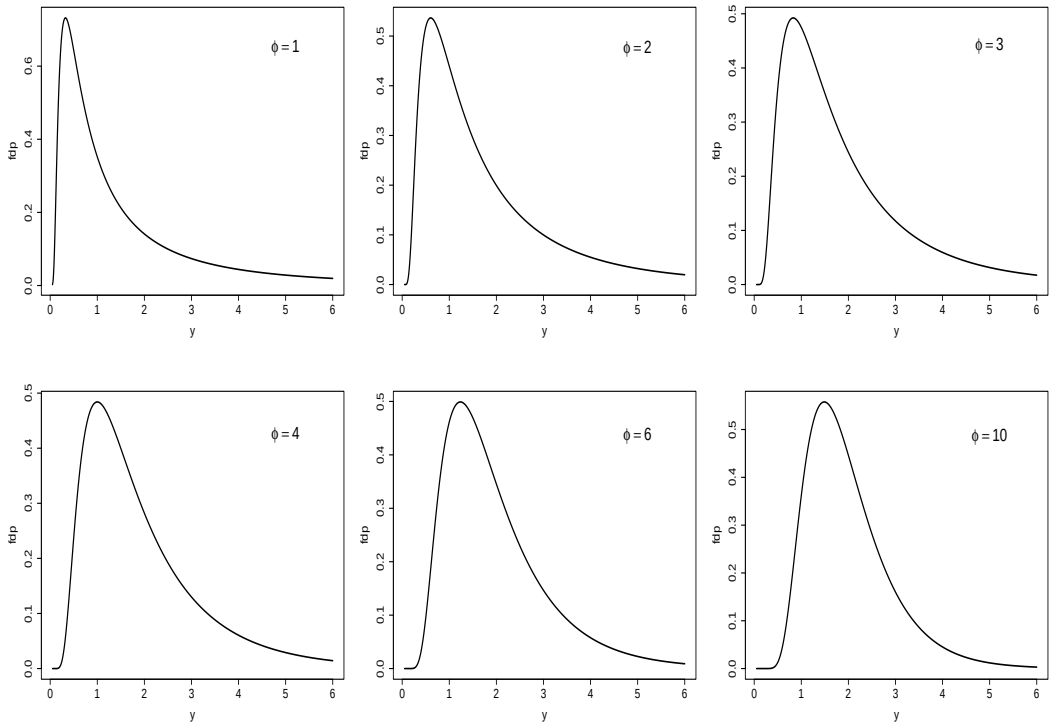

Figura 3.24: Densidades da distribuição normal inversa para alguns valores do parâmetro de dispersão e supondo $\mu = 2$.

dada por

$$S(t) = \Phi\{(1 - t\mu^{-1})\sqrt{\phi t^{-1}} - \exp(2\phi/\mu)\Phi\{-(1 + t\mu^{-1})\sqrt{\phi t^{-1}}\}.$$

A função de risco fica expressa na forma $h(t) = f(t)/S(t)$ em que $f(y)$ denota a função densidade da NI(μ, ϕ).

3.7 Modelos com resposta normal inversa

Sejam $Y_1, \ldots, Y_n$ variáveis aleatórias independentes tais que $Y_i \sim \text{NI}(\mu_i, \phi)$. Esta sendo assumido que essas variáveis possuem médias diferentes e mesma dispersão ϕ^{-1}. Ademais, supor que $g(\mu_i) = \eta_i$ em que $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ com $\mathbf{x}_i =$

$(x_{i1}, \ldots, x_{ip})^\top$ contendo valores de variáveis explicativas e $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ sendo o vetor de parâmetros de interesse. As ligações mais usadas no caso normal inversa são identidade ($\mu_i = \eta_i$), logarítmica ($\log\mu_i = \eta_i$) e recíproca quadrática ($\mu_i = \eta_i^{-2}$), esta última sendo a ligação canônica.

3.7.1 Qualidade do ajuste

Como foi visto na Seção 2.4 o desvio de um modelo com resposta normal inversa é dado por $\mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \phi \mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})$ em que

$$\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2/(y_i \hat{\mu}_i^2), \tag{3.5}$$

com $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$, $\hat{\eta}_i = \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}$ e $y_i > 0$. Como ϕ é desconhecido devemos estimá-lo, por exemplo através de máxima verossimilhança, cuja solução é dada por $\hat{\phi} = n/\mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$. Supondo que o modelo postulado está correto tem-se, para ϕ grande, que o desvio $\mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ segue distribuição qui-quadrado com $(n-p)$ graus de liberdade. Assim, valores altos para o desvio podem indicar inadequação do modelo ou falta de ajuste.

3.7.2 Técnicas de diagnóstico

O resíduo componente do desvio padronizado para os modelos com resposta normal inversa assumem a forma

$$t_{D_i} = \sqrt{\frac{2\hat{\phi}}{1 - \hat{h}_{ii}}} \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i \sqrt{y_i}},$$

em que $y_i > 0$ e h_{ii} é o i-ésimo elemento da diagonal principal da matriz $\mathbf{H} = \mathbf{W}^{\frac{1}{2}} \mathbf{X} (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{\frac{1}{2}}$ com $\omega_i = (d\mu_i/d\eta_i)^2/\mu_i^3$. Na expressão para t_{D_i} no caso da distribuição normal inversa o sinal do resíduo é o mesmo de $(y_i - \hat{\mu}_i)$. Estudos de simulação indicam que o resíduo t_{D_i} se aproxima da distribuição normal, particularmente para ϕ grande.

Similarmente aos modelos com resposta gama pode-se obter uma expressão aproximada para a distância de Cook quando a i-ésima observação é excluída. Essa expressão fica dada por

$$\mathrm{LD}_i = \frac{\hat{\phi}\hat{h}_{ii}}{(1-\hat{h}_{ii})^2}\frac{(y_i-\hat{\mu}_i)^2}{\hat{\mu}_i^2}.$$

Aqui também são sugeridos gráficos de t_{D_i} e $\hat{h}_{ii}$ contra os valores ajustados $\hat{\mu}_i$ e gráficos de índices de LD_i.

3.8 Aplicação

Considerar parte dos dados de um experimento desenvolvido no Departamento de Nutrição da Faculdade de Saúde Pública da USP em que 5 formas diferentes de um novo tipo de *snack*, com baixo teor de gordura saturada e de ácidos graxos, foram comparados ao longo de 20 semanas. Neste novo produto a gordura vegetal hidrogenada, responsável pela fixação do aroma do produto, foi substituída, totalmente ou parcialmente, por óleo de canola. As formas são as seguintes: A (22% de gordura, 0% de óleo de canola), B (0% de gordura, 22% de óleo de canola), C (17% de gordura, 5% de óleo de canola), D (11% de gordura, 11% de óleo de canola) e E (5% de gordura, 17% de óleo de canola). O experimento foi conduzido de modo que nas semanas pares 15 embalagens de cada um dos produtos A, B, C, D e E fossem analisadas em laboratório e observadas diversas variáveis (ver Paula et al., 2004). Em particular, será inicialmente estudado o comportamento da textura dos produtos através da força necessária para o cisalhamento. Os dados referentes a esta variável estão disponíveis no arquivo **snack.txt**.

Para ler o arquivo **snack.txt** no `R` deve-se fazer o seguinte:

```
snack = read.table("snack.txt", header=TRUE)
grupo = factor(grupo)
```

```
summary(snack)
attach(snacks).
```

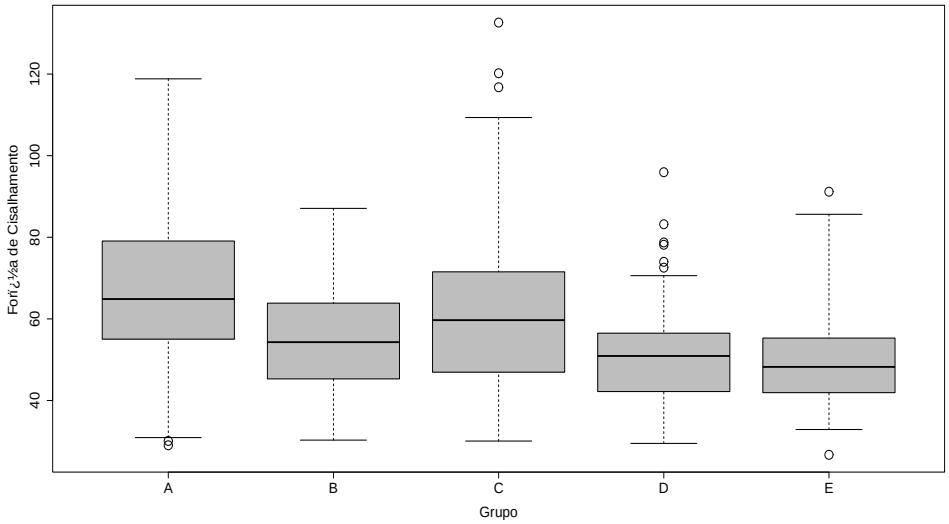

Figura 3.25: `Boxplots` da força de cisalhamento segundo o grupo e para todas as semanas.

Nota-se pela Figura 3.25, em que são apresentados os `boxplots robustos` da força de cisalhamento segundo o grupo e para todas as semanas, que os grupos A e C possuem os maiores valores, enquanto o grupo B tem valores intermediários e os grupos D e E têm os menores valores. Essa tendência pode ser observada pelos valores medianos da força de cisalhamento de cada grupo. Observa-se também que, exceto para o grupo B, todos os grupos apresentam valores discrepantes em geral destoando como valores altos em relação aos demais do mesmo grupo. Nota-se ainda uma assimetria à direita na distribuição da força de cisalhamento para todos os grupos. Esssas tendências são confirmadas pela tabela dada a seguir em que são apresentadas as médias, desvio padrão e coeficiente de variação para a força de cisalhamento para cada grupo.

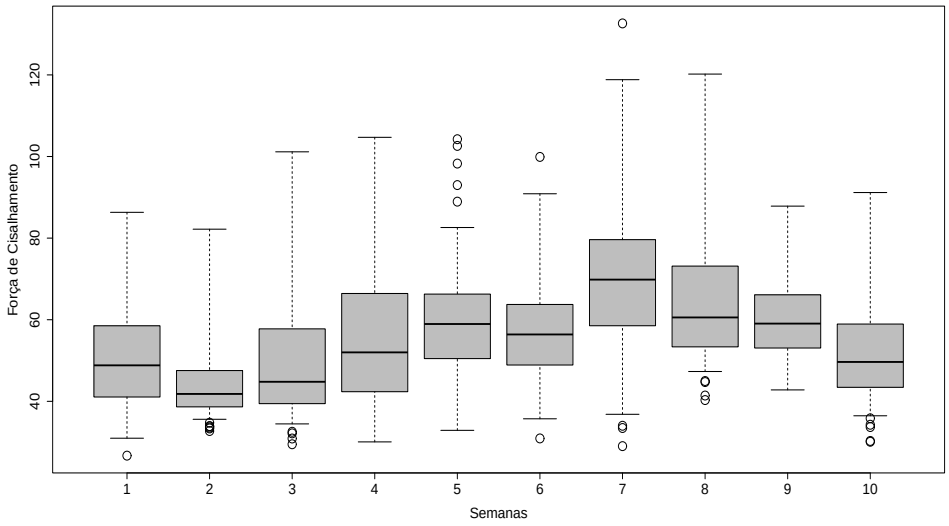

Figura 3.26: `Boxplots` da força de cisalhamento segundo a semana e para todos os grupos.

Estatística	Grupo A	Grupo B	Grupo C	Grupo D	Grupo E
Média	66,201	55,294	61,632	51,027	50,257
D.Padrão	18,707	13,143	19,601	10,960	11,402
C. Variação	28,20%	23,80%	31,80%	21,50%	22,70%

Já na Figura 3.26, em que são apresentados os `boxplots robustos` para todos os grupos ao longo das 20 semanas, uma tendência crescente é observada até a $14^{\underline{a}}$ semana seguida de um decrescimento até a última semana. Verifica-se também, para cada semana, que a distribuição da força de cisalhamento mostra-se assimétrica à direita sugerindo uma distribuição gama ou normal inversa. Esssas tendências são confirmadas pelo gráfico de perfis para a força de cisalhamento (vide Figura 3.27) e pela tabela dada a seguir em que são apresentadas as médias, desvio padrão e coeficiente de variação para a força de cisalhamento para cada semana.

Estatística	Semana 2	Semana 4	Semana 6	Semana 8	Semana 10
Média	50,95	44,66	50,08	55,57	60,15
D.Padrão	13,12	9,76	15,97	16,28	14,72
C. Variação	25,80%	21,90%	31,90%	29,30%	24,50%
Estatística	Semana 12	Semana 14	Semana 16	Semana 18	Semana 20
Média	57,84	71,57	65,18	60,37	52,45
D.Padrão	13,61	20,17	16,95	10,25	12,58
C. Variação	23,50%	28,20%	26,00%	17,00%	24,00%

Assim, denote por Y_{ijk} a força de cisalhamento referente à k-ésima réplica do i-ésimo grupo na j-ésima semana, para $k = 1, \ldots, 15$, $j = 2, 4, 6, \ldots, 20$ e i =1(A),2(B),3(C),4(D) e E(5). A fim de comparar as duas distribuições assimétricas supor que $Y_{ijk} \overset{\text{ind}}{\sim} \text{G}(\mu_{ij}, \phi)$ e $Y_{ijk} \overset{\text{ind}}{\sim} \text{NI}(\mu_{ij}, \phi)$, respectivamente, com parte sistemática dada por

$$\mu_{ij} = \alpha + \beta_i + \gamma_1 \text{semana}_j + \gamma_2 \text{semana}_j^2, \tag{3.6}$$

em que $\beta_1 = 0$. Portanto, α é o efeito da forma A, controlando pela semana, e $\alpha + \beta_i$ (i=2,3,4,5) são os efeitos das demais formas B, C, D e E, respectivamente. Está sendo assumida a mesma tendência para os cinco tipos de *snack*. Alternativamente, poderia ser incluída interação entre grupo e semana, possibilitando o ajuste de tendências separadas para cada grupo.

Para ajustar o modelo (3.6) com resposta normal inversa sem interação deve-se usar os comandos

```
s1 = semana
s2 = s1*s1
fit1.snack = glm(textura ~ grupo + s1 + s2,
family=inverse.gaussian(link=identity))
summary(fit1.snack).
```

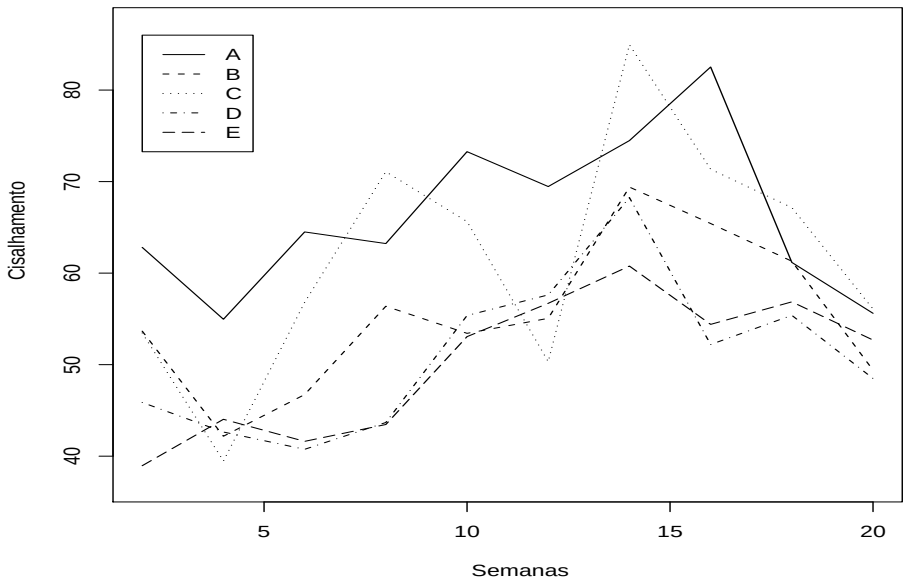

Figura 3.27: Perfis da força de cisalhamento segundo as semanas e os grupos.

Abaixo seguem os comandos para o ajuste com interação

`fit2.snack = glm(textura ~ grupo + s1 + s2 + s1*grupo +s2*grupo, family=inverse.gaussian(link=identity))`

`summary(fit2.snack).`

Contudo a interação entre grupo e semana não é significativa. Este é um exemplo em que há uma ligeira superioridade da distribuição normal inversa em relação à distribuição gama. Embora a função de variância da normal inversa seja cúbica enquanto para a gama tem-se função de variância quadrática, nem sempre é possível diferenciar de forma clara os dois ajustes. Nota-se pela Figura 3.28 que o gráfico de resíduos de Pearson contra os valores ajustados apresenta uma tendência sistemática crescente sob o modelo gama, que é amenizada sob o modelo com erros normal inversa. Os dois modelos

ajustam-se muito bem aos dados como pode-se notar pelo valor do desvio do modelo gama $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 756,87$ (753 g.l.) com P=0,35 e pelo gráfico normal de probabilidades para o modelo com resposta normal inversa apresentado na Figura 3.29.

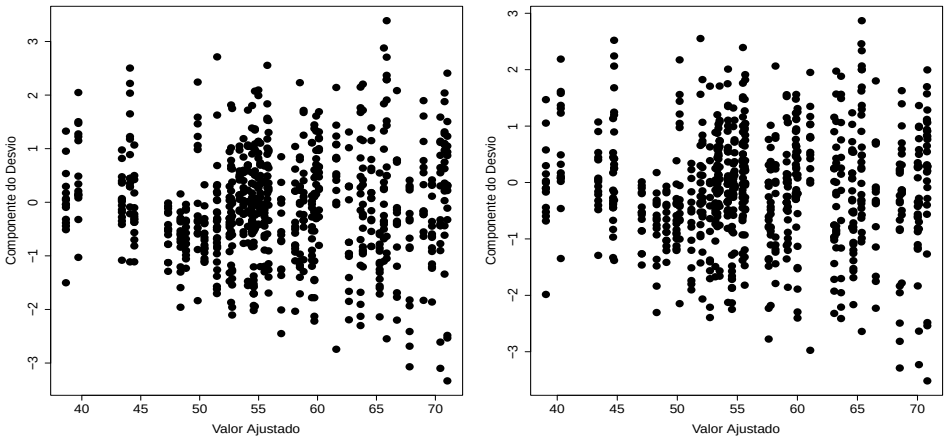

Figura 3.28: Gráficos do resíduo de Pearson contra o valor ajustado referentes aos modelos gama (esquerda) e normal inversa (direita) ajustados aos dados sobre *snacks* .

Tabela 3.6
Estimativas dos parâmetros referentes ao modelo com resposta normal inversa ajustado aos dados sobre snacks.

Efeito	Estimativa	E/E.Padrão
Constante	50,564	26,32
Grupo B	-10,916	-6,41
Grupo C	-5,459	-3,03
Grupo D	-15,357	-9,42
Grupo E	-16,596	-10,30
Semana	2,727	8,18
Semana^2	-0,091	-5,90
ϕ	1005	-

Na Tabela 3.6 são apresentadas as estimativas sob o modelo com resposta normal inversa. Todos os efeitos são altamente significativos, em particular o efeito de semana na forma quadrática. Controlando esse efeito, a maior força média de cisalhamento ocorre com o produto sob a forma A (ausência de óleo de canola) e a menor força média de cisalhamento ocorre com as formas D e E, confirmando-se as tendências observadas na Figura 3.22.

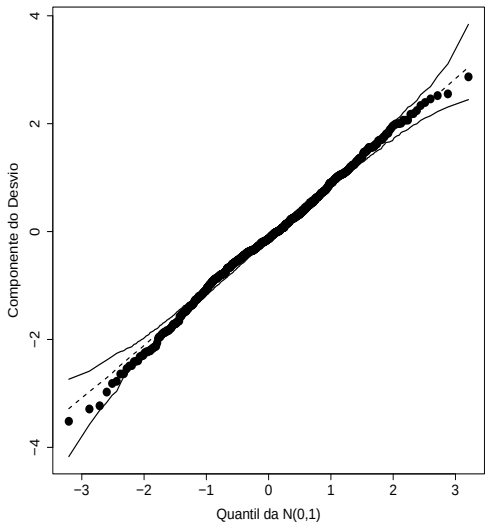

Figura 3.29: Gráfico normal de probabilidades referente ao modelo com resposta normal inversa ajustado aos dados sobre *snacks*.

Na Figura 3.30 tem-se os valores preditos para os 5 grupos ao longo das 20 semanas. A estimativa do parâmetro de precisão indica que a distribuição da força de cisalhamento em cada grupo, fixando o tempo, é aproximadamente normal. Contudo, a variância depende da média. A forma cúbica para a variância mostrou-se ligeiramente superior à forma quadrática. Outras formas para ajustar a variância podem ser testadas, como por exemplo, através de modelos de quase-verossimilhança que serão discutidos no Capítulo 5. O

paralelismo entre as curvas apresentadas na Figura 3.30 é devido à não inclusão de interação entre semana e grupo. Alternativamente, poderia ser incluída uma função para cada grupo, ou então, o efeito semana poderia ser controlado através de funções aditivas.

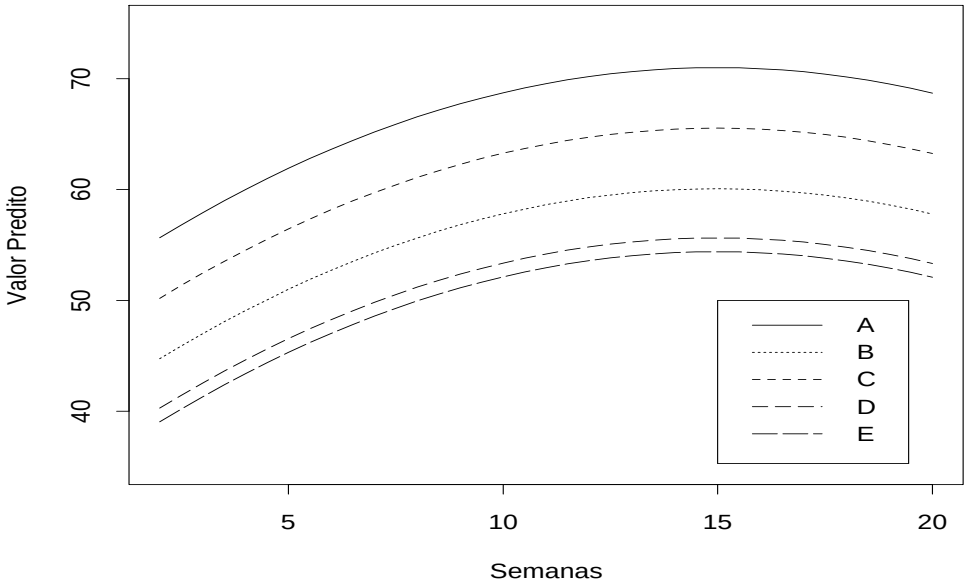

Figura 3.30: Valores preditos para a força média de cisalhamento para as 5 formas de *snacks* através do modelo com resposta normal inversa.

Algumas observações foram detectadas como possivelmente influentes (vide Figura 3.31): #2 (2ª semana, grupoB), #8 (2ª semana, grupo B), #10 (2ª semana, grupo B), #311 (2ª semana, grupo C), #405 (14ª semana, grupo C) #465 (2ª semana, grupo D) e #744 (última semana, grupo E). Embora os valores preditos para a força de cisalhamento dessas amostras estejam abaixo da média, os valores observados são em geral altos quando comparados com os valores dos grupos e das semanas correspondentes. Também o fato de 5 dessas observações terem ocorrido logo na segunda semana pode ser um indício de alguma dificuldade inicial com o experimento. A eliminação

dessas 7 observações do total de 744 observações leva a algumas variações desproporcioanis. Por exemplo, as estimativas dos efeitos dos grupos B e C diminuem, respectivamente, 9,1% e 14%. Todavia, não ocorrem mudanças inferenciais entre os efetios dos grupos B, C, D e E com relação ao grupo A.

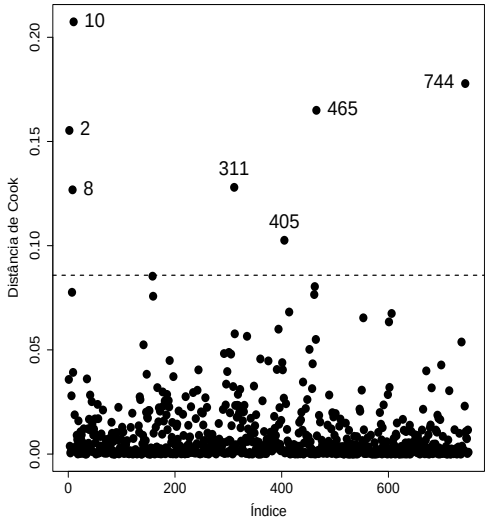

Figura 3.31: Gráfico da distância de Cook referente ao modelo com resposta normal inversa ajustado aos dados sobre *snacks*.

3.9 Modelagem simultânea da média e da dispersão

Eventualmente o coeficiente de variação pode não ser constante variando com as observações. Smyth (1989) introduziu os modelos lineares generalizados duplos com modelagem conjunta da média e do parâmetro de precisão (ou dispersão) e desenvolveu um processo de estimação baseado no método de máxima verossimilhança que será descrito a seguir. Contudo, outros métodos alternativos de estimação, tais como máxima verossimilhança restrita, foram

propostos mais recentemente com o intuito de reduzir o viés das estimativas de máxima verossimilhança, particularmente dos coeficientes do componente de dispersão. Uma discussão a respeito desses métodos pode ser encontrada em Smyth e Verbyla (1999).

A fim de formalizar os MLGs duplos supor que $Y_1, \ldots, Y_n$ são variáveis aleatórias independentes com função densidade ou função de probabilidades expressa na forma

$$f(y; \theta_i, \phi_i) = \exp[\phi_i\{y\theta_i - b(\theta_i)\} + c(y, \phi_i)],$$

em que $c(y, \phi_i) = d(\phi_i) + \phi_i a(y) + u(y)$. Essa decomposição, como visto na Seção 2.7.2, vale somente para as distribuições normal, normal inversa e gama da família exponencial. Além disso, supor que

$$g(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} \;\text{ e }\; h(\phi_i) = \lambda_i = \mathbf{z}_i^\top \boldsymbol{\gamma},$$

em que $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ e $\mathbf{z}_i = (z_{i1}, \ldots, z_{iq})^\top$ contêm valores de variáveis explicativas e $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ e $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$ são os parâmetros a serem estimados.

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$, então o logaritmo da função de verossimilhança fica dado por

$$\begin{aligned} \mathrm{L}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} [\phi_i\{y_i\theta_i - b(\theta_i)\} + d(\phi_i) + \phi_i a(y_i) + u(y_i)] \\ &= \sum_{i=1}^{n} \{\phi_i t_i + d(\phi_i) + u(y_i)\}, \end{aligned} \tag{3.7}$$

em que $t_i = y_i\theta_i - b(\theta_i) + a(y_i)$. Portanto, se θ_i for fixado a expressão (3.7) coincide com o logaritmo da função de verossimilhança de um modelo da família exponencial com respostas independentes $T_1, \ldots, T_n$ (valores observados $t_1, \ldots, t_n$), parâmetros canônicos $\phi_1, \ldots, \phi_n$ e parâmetro de dispersão

igual a 1. Pelas propriedades da família exponencial segue que

$$\mu_{T_i} = \mathrm{E}(T_i) = -d'(\phi_i) \;\; \text{e} \;\; \mathrm{Var}(T_i) = -d''(\phi_i).$$

Essas quantidades são descritas na Tabela 3.7 para as distribuições normal, normal inversa e gama. Os resultados acima podem ser obtidos, alternativamente, aplicando-se condições usuais de regularidade no logaritmo da verossimilhança dado em (3.7).

Tabela 3.7
Derivação de algumas quantidades para distribuições da família exponencial.

	Normal	Normal inversa	Gama
t_i	$y_i\mu_i - \frac{1}{2}(\mu_i^2 + y_i^2)$	$-\{y_i/2\mu_i^2 - \mu_i^{-1} + (2y_i)^{-1}\}$	$\log(y_i/\mu_i) - y_i/\mu_i$
$d(\phi)$	$\frac{1}{2}\log\phi$	$\frac{1}{2}\log\phi$	$\phi\log\phi - \log\Gamma(\phi)$
$d'(\phi)$	$(2\phi)^{-1}$	$(2\phi)^{-1}$	$(1 + \log\phi) - \psi(\phi)$
$d''(\phi)$	$-(2\phi^2)^{-1}$	$-(2\phi^2)^{-1}$	$\phi^{-1} - \psi'(\phi)$

Conforme observado por Smyth (1989) para as distribuições normal e normal inversa chamando $D_i = -2T_i$ $(i = 1, \ldots, n)$ segue que

$$\mathrm{E}(D_i) = \phi_i^{-1} \;\; \text{e} \;\; \mathrm{Var}(D_i) = \frac{\mathrm{E}^2(D_i)}{\nu},$$

em que $\nu = \frac{1}{2}$. Portanto, a expressão (3.7) pode ser interpretada para os modelos com resposta normal e normal inversa como um MLG de respostas independentes $D_1, \ldots, D_n$ com distribuição gama de médias $\phi_1^{-1}, \ldots, \phi_n^{-1}$, respectivamente, e parâmetro de dispersão $\nu^{-1} = 2$. Assim, para θ_i fixado, os parâmetros da dispersão podem ser estimados alternativamente através de um MLG com respostas independentes gama, função de ligação $h(\cdot)$ e parâmetro de dispersão igual a 2.

3.9.1 Estimação

A função escore e a matriz de informação de Fisher para $\boldsymbol{\beta}$ podem ser obtidas facilmente seguindo os passos da Seção 2.5.1. Assim, obtém-se

$$\begin{aligned} \mathbf{U}_\beta &= \mathbf{X}^\top \boldsymbol{\Phi} \mathbf{W}^{\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{y} - \boldsymbol{\mu}) \text{ e} \\ \mathbf{K}_{\beta\beta} &= \mathbf{X}^\top \boldsymbol{\Phi} \mathbf{W} \mathbf{X}, \end{aligned}$$

em que $\mathbf{X}$ é uma matriz $n \times p$ de linhas $\mathbf{x}_i^\top$ $(i = 1, \ldots, n)$, $\mathbf{W} = \text{diag}\{\omega_1, \ldots, \omega_n\}$ com pesos $\omega_i = (d\mu_i/d\eta_i)^2/V_i$, $\mathbf{V} = \text{diag}\{V_1, \ldots, V_n\}$, $\boldsymbol{\Phi} = \text{diag}\{\phi_1, \ldots, \phi_n\}$, $\mathbf{y} = (y_1, \ldots, y_n)^\top$ e $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$.

Para obter a função escore para o parâmetro $\boldsymbol{\gamma}$, será calculado inicialmente a derivada

$$\begin{aligned} \partial \text{L}(\boldsymbol{\theta})/\partial \gamma_j &= \sum_{i=1}^{n} \left\{ \frac{d\phi_i}{d\lambda_i} \frac{\partial \lambda_i}{\partial \gamma_j} t_i + d'(\phi_i) \frac{d\phi_i}{d\lambda_i} \frac{\partial \lambda_i}{\partial \gamma_j} \right\} \\ &= \sum_{i=1}^{n} \left\{ \frac{1}{h'(\phi_i)} z_{ij} t_i + d'(\phi_i) \frac{1}{h'(\phi_i)} z_{ij} \right\} \\ &= \sum_{i=1}^{n} \frac{z_{ij}}{h'(\phi_i)} \{t_i + d'(\phi_i)\}, \end{aligned}$$

em que $h'(\phi_i) = d\lambda_i/d\phi_i$. Portanto, em forma matricial obtém-se

$$\mathbf{U}_\gamma = \mathbf{Z}^\top \mathbf{H}_\gamma^{-1} (\mathbf{t} - \boldsymbol{\mu}_T),$$

em que $\mathbf{H}_\gamma = \text{diag}\{h'(\phi_1), \ldots, h'(\phi_n)\}$, $\mathbf{t} = (t_1, \ldots, t_n)^\top$ e $\boldsymbol{\mu}_T = (\text{E}(T_1), \ldots, \text{E}(T_n))^\top$ $= (-d'(\phi_1), \ldots, -d'(\phi_n))^\top$.

Para obter a matriz de informação de Fisher para o parâmetro $\boldsymbol{\gamma}$ é preciso das derivadas

$$\begin{aligned} \partial^2 \text{L}(\boldsymbol{\theta})/\partial \gamma_j \partial \gamma_\ell &= -\sum_{i=1}^{n} \frac{z_{ij}}{\{h'(\phi_i)\}^2} \left[d''(\phi_i) h(\phi_i) \frac{d\phi_i}{d\lambda_i} z_{i\ell} - h''(\phi_i) \{t_i + d'(\phi_i)\} \frac{d\phi_i}{d\lambda_i} z_{i\ell} \right] \\ &= -\sum_{i=1'}^{n} \frac{z_{ij} z_{i\ell}}{\{h'(\phi_i)\}^2} \left[d''(\phi_i) - \frac{h''(\phi_i)}{h'(\phi_i)} \{t_i + d'(\phi_i)\} \right], \end{aligned}$$

cujos valores esperados ficam dados por

$$\mathrm{E}\left\{-\frac{\partial^2 \mathrm{L}(\boldsymbol{\theta})}{\partial\gamma_j\partial\gamma_\ell}\right\} = -\sum_{i=1}^{n}\frac{d''(\phi_i)}{\{h'(\phi_i)\}^2}z_{ij}z_{i\ell}.$$

Logo, em forma matricial tem-se que

$$\mathbf{K}_{\gamma\gamma} = \mathbf{Z}^\top\mathbf{P}\mathbf{Z},$$

em que $\mathbf{P} = \mathbf{V}_\gamma\mathbf{H}_\gamma^{-2}$, $\mathbf{V}_\gamma = \mathrm{diag}\{-d''(\phi_1),\ldots,-d''(\phi_n)\}$. Devido à ortogonalidade entre os parâmetros θ_i e ϕ_i, segue diretamente a ortogonalidade entre $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$. Assim, a matriz de informação de Fisher para $\boldsymbol{\theta}$ é bloco diagonal $\mathbf{K}_{\theta\theta} = \mathrm{diag}\{\mathbf{K}_{\beta\beta}, \mathbf{K}_{\gamma\gamma}\}$.

Similarmente aos MLGs pode-se desenvolver um processo iterativo escore de Fisher para encontrar as estimativas de máxima verossimilhança $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$. Após algumas manipulações algébricas chega-se ao proceso iterativo

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top\boldsymbol{\Phi}^{(m)}\mathbf{W}^{(m)}\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\Phi}^{(m)}\mathbf{W}^{(m)}\mathbf{y}^{*(m)} \quad \text{e} \tag{3.8}$$

$$\boldsymbol{\gamma}^{(m+1)} = (\mathbf{Z}^\top\mathbf{P}^{(m)}\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{P}^{(m)}\mathbf{z}^{*(m)}, \tag{3.9}$$

em que $\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}^{-\frac{1}{2}}\mathbf{V}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})$, $\mathbf{z}^* = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{V}_\gamma^{-1}\mathbf{H}_\gamma(\mathbf{t}-\boldsymbol{\mu}_T)$ e $m = 0, 1, 2, \ldots$. Conforme mencionado por Smyth (1989) o processo iterativo (3.8)-(3.9) pode ser resolvido alternando-se as duas equações até a convergência. Pode-se iniciar o processo iterativo (3.8) com as estimativas do MLG com ϕ_i comum a todas as observações.

Sob as condições de regularidade apresentadas na Seção 2.6 segue para n grande que $\hat{\boldsymbol{\beta}} \sim \mathrm{N}_p(\boldsymbol{\beta}, \mathbf{K}_{\beta\beta}^{-1})$ e $\hat{\boldsymbol{\gamma}} \sim \mathrm{N}_q(\boldsymbol{\gamma}, \mathbf{K}_{\gamma\gamma}^{-1})$, respectivamente. Além disso, devido à ortogonalidade entre $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ segue a independência assintótica entre $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$.

3.9.2 Métodos de diagnóstico

Nesta subseção será apresentada a derivação de alguns procedimentos de diagnóstico para a classe dos MLGs duplos (vide Paula, 2013).

Resíduos

Na classe dos MLGs duplos pode-se definir desvios para a média e para a precisão, respectivamente. O desvio para a média assume a mesma expressão da classe dos MLGs em que somente a média é ajustada, com ϕ_i no lugar de ϕ. Denota-se esse desvio por $\mathrm{D}_1^*(\mathbf{y};\hat{\boldsymbol{\mu}},\boldsymbol{\phi}) = \sum_{i=1}^n d_1^{*2}(y_i;\hat{\mu}_i,\phi_i)$, em que $d_1^{*2}(y_i;\hat{\mu}_i,\phi_i) = 2\phi_i[y_i(\tilde{\theta}_i - \hat{\theta}_i) + \{b(\hat{\theta}_i) - b(\tilde{\theta}_i)\}]$. Para ϕ_i grande $\forall i$ o desvio $\mathrm{D}_1^*(\mathbf{y};\hat{\boldsymbol{\mu}},\boldsymbol{\phi})$ pode ser comparado com os quantis da distribuição qui-quadrado com $(n-p)$ graus de liberdade. Para o modelo normal heteroscedástico o desvio para a média fica dado por $\mathrm{D}_1^*(\mathbf{y};\hat{\boldsymbol{\mu}},\boldsymbol{\phi}) = \sum_{i=1}^n \sigma_i^{-2}(y_i - \hat{y}_i)^2$. Na prática deve-se substituir ϕ_i por $\hat{\phi}_i = h^{-1}(\hat{\lambda}_i) = \mathbf{z}_i^\top\hat{\boldsymbol{\gamma}}$.

O resíduo Studentizado, no modelo normal heteroscedástico, assume a forma

$$t_i^* = \frac{y_i - \hat{y}_i}{\hat{\sigma}_i\sqrt{1-\hat{h}_{ii}}},$$

em que $\hat{h}_{ii} = \hat{\sigma}_i^2\mathbf{x}_i^\top(\mathbf{X}^\top\hat{\boldsymbol{\Phi}}\mathbf{X})^{-1}\mathbf{x}_i$ com $\boldsymbol{\Phi} = \mathrm{diag}\{\sigma_1^{-2},\ldots,\sigma_n^{-2}\}$. Para os demais MLGs duplos o resíduo componente do desvio para a média fica dado por

$$t_{D_{1i}} = \frac{d_1^*(y_i;\hat{\mu}_i,\hat{\phi}_i)}{\sqrt{1-\hat{h}_{ii}}},$$

em que $d_1^*(y_i;\hat{\mu}_i,\hat{\phi}_i) = \pm\sqrt{d_1^{*2}(y_i;\hat{\mu}_i,\hat{\phi}_i)}$, o sinal continua sendo o mesmo de $(y_i - \hat{\mu}_i)$ e $\hat{h}_{ii}$ é o i-ésimo elemento da diagonal principal da matriz

$$\hat{\mathbf{H}} = \hat{\boldsymbol{\Phi}}^{\frac{1}{2}}\hat{\mathbf{W}}^{\frac{1}{2}}\mathbf{X}(\mathbf{X}^\top\hat{\boldsymbol{\Phi}}\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^\top\hat{\boldsymbol{\Phi}}^{\frac{1}{2}}\hat{\mathbf{W}}^{\frac{1}{2}},$$

ou seja,

$$\hat{h}_{ii} = \hat{\phi}_i \hat{\omega}_i \mathbf{x}_i^\top (\mathbf{X}^\top \hat{\mathbf{\Phi}} \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i.$$

Sugere-se o gráfico normal de probabilidades para $t_{D_{1i}}$ e o gráfico de $t_{D_{1i}}$ contra os valores ajustados.

Denote por $\mathrm{D}_2^*(\mathbf{y}; \hat{\boldsymbol{\phi}}, \boldsymbol{\mu}) = \sum_{i=1}^n d_2^{*2}(y_i; \hat{\phi}_i, \mu_i)$ o desvio para a precisão, em que $d_2^{*2}(y_i; \hat{\phi}_i, \mu_i) = 2[t_i(\tilde{\phi}_i - \hat{\phi}) + \{d(\tilde{\phi}_i) - d(\hat{\phi}_i)\}]$, $\tilde{\phi}_i$ é solução para ϕ_i sob o modelo saturado sendo dada por $d'(\tilde{\phi}_i) = -t_i$. Para os modelos com resposta normal e normal inversa tem-se que $\tilde{\phi}_i = -(2t_i)^{-1}$. Já para modelos com resposta gama $\tilde{\phi}_i$ é a solução da equação $\{\psi(\tilde{\phi}_i) - \log\tilde{\phi}_i + 1\} = t_i$. Aqui também para ϕ_i grande $\forall i$ o desvio $\mathrm{D}_2^*(\mathbf{y}; \hat{\boldsymbol{\phi}}, \boldsymbol{\mu})$ pode ser comparado com os quantis da distribuição qui-quadrado com $(n - q)$ graus de liberdade.

O resíduo componente do desvio para a precisão fica dado por

$$t_{D_{2i}} = \frac{d_2^*(y_i; \hat{\phi}_i, \hat{\mu}_i)}{\sqrt{1 - \hat{r}_{ii}}},$$

em que $d_2^*(y_i; \hat{\phi}_i, \hat{\mu}_i) = \pm\sqrt{d_2^{*2}(y_i; \hat{\phi}_i, \hat{\mu}_i)}$, o sinal sendo o mesmo de $\{\hat{t}_i + d'(\hat{\phi}_i)\}$ e $\hat{r}_{ii}$ é o i-ésimo elemento da diagonal principal da matriz

$$\hat{\mathbf{R}} = \hat{\mathbf{P}}^{\frac{1}{2}} \mathbf{Z} (\mathbf{Z}^\top \hat{\mathbf{P}} \mathbf{Z})^{-1} \mathbf{Z}^\top \hat{\mathbf{P}}^{\frac{1}{2}},$$

ou seja,

$$\hat{r}_{ii} = \hat{p}_i \mathbf{z}_i^\top (\mathbf{Z}^\top \hat{\mathbf{P}} \mathbf{Z})^{-1} \mathbf{z}_i.$$

Note que $p_i = -d''(\phi_i)\{h'(\phi_i)\}^{-2}$. Por exemplo, para ligação logarítmica tem-se que $h(\phi_i) = \log\phi_i$ então $h'(\phi_i) = \phi_i^{-1}$ e portanto $p_i = -\phi_i^2 d''(\phi_i)$. Assim, para os modelos com resposta normal e normal inversa segue que $p_i = \phi_i^2(2\phi_i^2)^{-1} = \frac{1}{2}$ e para os modelos com resposta gama $p_i = \phi_i\{\phi_i\psi'(\phi_i) - 1\}$.

Sugere-se o gráfico normal de probabilidades para $t_{D_{2i}}$ e o gráfico de $t_{D_{2i}}$ contra os valores ajustados.

Influência

Para avaliar a sensibilidade das estimativas dos parâmetros que modelam a média pode-se usar a medida de influência LD_i definida na Seção 2.8.3 com $\hat{\phi}_i$ no lugar de $\hat{\phi}$, que será definida por

$$\mathrm{LD}_i^{\beta} = \left\{ \frac{\hat{h}_{ii}}{1-\hat{h}_{ii}} \right\} t_{S_i}^2,$$

em que

$$t_{S_i} = \frac{\sqrt{\hat{\phi}_i}(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i(1-\hat{h}_{ii})}}.$$

Gráficos de índices de LD_i^{β} e $\hat{h}_{ii}$ contra os valores ajustados são recomendados.

Para avaliar a sensibilidade da estimativa $\hat{\boldsymbol{\gamma}}$ quando a i-ésima observação é deletada será utilizada uma aproximação de um passo, que é obtida de forma similar à aproximação de uma passo $\hat{\boldsymbol{\beta}}_{(i)}$ descrita na Seção 2.8.3, dada por

$$\hat{\boldsymbol{\gamma}}_{(i)} = \hat{\boldsymbol{\gamma}} - \frac{(\mathbf{Z}^\top \hat{\mathbf{P}} \mathbf{Z})^{-1} \mathbf{z}_i \{ t_i + d'(\hat{\phi}_i) \}}{h'(\hat{\phi}_i)(1-\hat{r}_{ii})}, \tag{3.10}$$

em que $\hat{r}_{ii}$ é o i-ésimo elemento da diagonal principal da matriz $\mathbf{R}$.

Uma medida para avaliar a influência nas estimativas dos parâmetros da precisão fica dada por

$$\begin{aligned} \mathrm{LD}_i^{\gamma} &= (\hat{\boldsymbol{\gamma}}_{(i)} - \hat{\boldsymbol{\gamma}})^\top (\mathbf{Z}^\top \hat{\mathbf{P}} \mathbf{Z})(\hat{\boldsymbol{\gamma}}_{(i)} - \hat{\boldsymbol{\gamma}}) \\ &= \left\{ \frac{\hat{r}_{ii}}{1-\hat{r}_{ii}} \right\} t_{T_i}^2, \end{aligned}$$

em que

$$\begin{aligned} t_{T_i} &= \frac{t_i + d'(\hat{\phi}_i)}{h'(\hat{\phi}_i)\sqrt{\hat{p}_i(1-\hat{r}_{ii})}} \\ &= \frac{t_i + d'(\hat{\phi}_i)}{\sqrt{-d''(\hat{\phi}_i)(1-\hat{r}_{ii})}}. \end{aligned}$$

Gráficos de índices de LD_i^{γ} e $\hat{r}_{ii}$ contra os valores ajustados são recomendados.

Para os modelos com resposta normal e com resposta normal inversa o resíduo t_{T_i} assume a forma

$$t_{T_i} = \frac{t_i + (2\hat{\phi}_i)^{-1}}{(\sqrt{2}\hat{\phi}_i)^{-1}\sqrt{1-\hat{r}_{ii}}},$$

e para modelos com resposta gama tem-se que

$$t_{T_i} = \frac{t_i + \{1 + \log\hat{\phi}_i - \psi(\hat{\phi}_i)\}}{\sqrt{\{\psi'(\hat{\phi}_i) - \hat{\phi}_i^{-1}\}(1-\hat{r}_{ii})}}.$$

Verbyla (1993) apresenta uma aproximação de uma passo para $\hat{\boldsymbol{\gamma}}_{(i)}$ para o caso normal usando um esquema de perturbação específico para modelos normais heteroscedásticos. Para obter a aproximação apresentada em (3.10) usa-se a ponderação de casos usual para MLGs. Estudos sobre a qualidade da aproximação apresentada em (3.10) ainda não foram desenvolvidos.

3.9.3 Aplicação

Pela análise descritiva apresentada na Seção 3.8.1 sobre o comportamento da força de cisalhamento dos cinco tipos de *snack* ao longo das 20 semanas e também pelo gráfico de perfis para a força de cisalhamento (Figura 3.32) nota-se que o coeficiente de variação não parece ser constante. Assim, a modelagem dupla da média e da precisão pode levar a um ajuste mais satisfatório para o modelo com resposta gama. Dessa forma supor que $Y_{ijk} \overset{\text{ind}}{\sim} \mathrm{G}(\mu_{ij}, \phi_{ij})$, em que Y_{ijk} denota a força de cisalhamento referente à k-ésima réplica do i-ésimo grupo na j-ésima semana, para $k = 1, \ldots, 15$, $j = 2, 4, 6, \ldots, 20$ e i =1(A),2(B),3(C),4(D) e E(5), com parte sistemática dada por

$$\begin{aligned} \mu_{ij} &= \beta_0 + \beta_i + \beta_6 \mathrm{semana}_j + \beta_7 \mathrm{semana}_j^2 \quad \text{e} \\ \log(\phi_{ij}) &= \gamma_0 + \gamma_i + \gamma_6 \mathrm{semana}_j + \gamma_7 \mathrm{semana}_j^2, \end{aligned}$$

em que $\beta_1 = 0$ e $\gamma_1 = 0$. Portanto β_0 e γ_0 são os efeitos da forma A, controlando-se pela semana, na média e na precisão, respectivamente, enquanto $\beta_0 + \beta_i$ e $\gamma_0 + \gamma_i$ são os efeitos das demais formas B, C, D e E na média e precisão, respectivamente.

Tabela 3.8
Estimativas dos parâmetros referentes ao MLG duplo com resposta gama ajustado aos dados sobre snacks.

	Média		Dispersão	
Efeito	Estimativa	E/E.Padrão	Estimativa	E/E.Padrão
Constante	36,990	11,53	1,560	7,27
Grupo B	-10,783	-6,40	0,477	2,95
Grupo C	-3,487	-1,98	0,050	0,31
Grupo D	-14,829	-9,18	0,815	5,05
Grupo E	-15,198	-9,54	0,817	5,06
Semana	5,198	9,88	0,155	3,91
Semana^2	-0,189	-8,88	-0,005	-2,99

O MLG duplo pode ser ajustado no `R` através dos seguintes comandos:

```
require(dglm)
fit3.snack = dglm(cisalhamento ~ grupo + s1 + s2,
~ grupo + s1 + s2, family=Gamma(link=identity))
summary(fit3.snack).
```

Note que a biblioteca `dglm` faz o ajuste de $\log(\phi_i^{-1})$, ou seja da dispersão, sendo necessário fazer as adaptações nos modelos com resposta gama e normal inversa para obter $\log(\phi_i)$, ajuste da precisão. Em particular no caso de modelos normais heteroscedásticos tem-se diretamente o ajuste de $\log(\sigma_i^2)$, em que σ_i^2 é a variância.

Na Tabela 3.8 são apresentadas as estimativas com os respectivos erros padrão dos parâmetros da média e da dispersão. Pode-se notar pelas estimativas dos parâmetros da média as mesmas tendências observadas na Figura

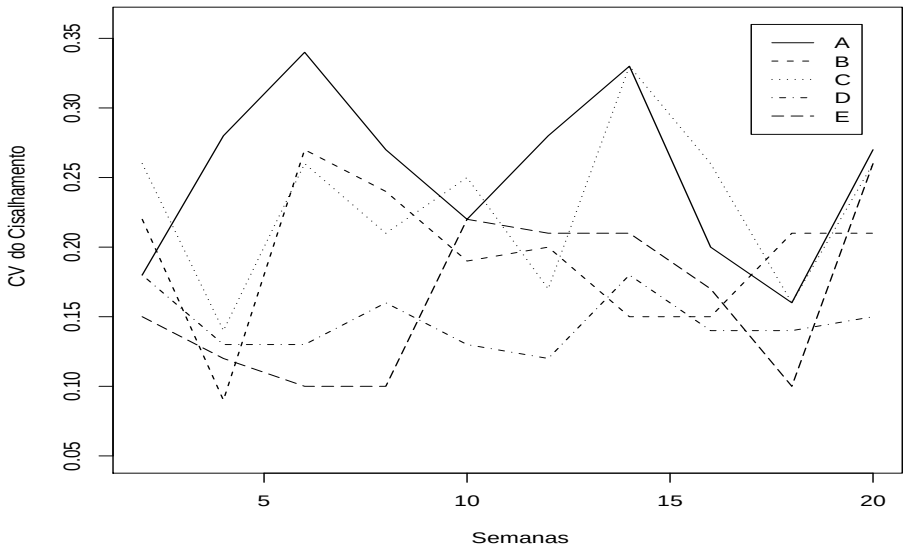

Figura 3.32: Perfis do CV da força de cisalhamento segundo as semanas e os grupos.

3.30 para o modelo com resposta normal inversa. O grupo A tem a maior média para a força de cisalhamento enquanto os grupos D e E têm as menores médias. Com relação às estimativas dos parâmetros da dispersão nota-se que a variabilidade (no sentido do coeficiente de variação) depende do tempo de forma quadrática e que os grupos A e C apresentam maior variabilidade enquanto os grupos D e E apresentam as menores variabilidades.

Nota-se ainda que os mesmos efeitos que são significativos para os parâmetros da média são também significativos para os parâmetros da dispersão. Apenas três observações, #430, #595 e #744, aparecem como possivelmente influentes nos parâmetros da média e da dispersão, como pode ser observado pelas Figuras 3.33 e 3.34. A eliminação desses pontos não muda a inferência. Pelos gráficos normais de probabilidades para o resíduo componente do des-

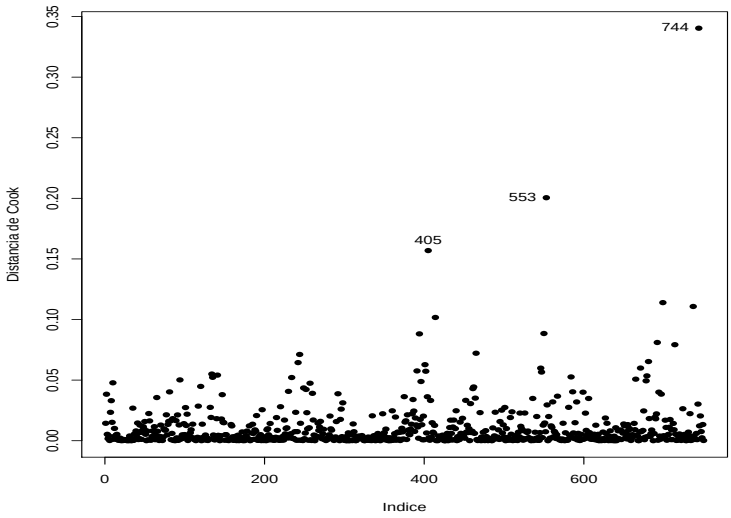

Figura 3.33: Gráfico da distância de Cook para as estimativas dos coeficientes da média referente ao MLG duplo com resposta gama ajustado aos dados sobre *snacks*.

vio para a média e para a dispersão apresentados nas Figuras 3.35 e 3.36, respectivamente, não há indícios de inadequação do MLG duplo.

3.10 Exercícios

1. Seja $Y \sim \mathrm{G}(\mu, \phi)$ e considere a variável aleatória $\log(Y)$. Use a condição de regularidade $\mathrm{E}(\mathrm{U}_\phi) = 0$ para mostrar que $\mathrm{E}\{\log(Y)\} = \log(\mu) - \log(\phi) + \psi(\phi)$, em que $\mathrm{U}_\phi = \partial \mathrm{L}(\mu, \phi)/\partial\phi$.

2. Seja $Y \sim \mathrm{NI}(\mu, \phi)$ e considere a variável aleatória Y^{-1}. Use a condição de regularidade $\mathrm{E}(\mathrm{U}_\phi) = 0$ para mostrar que $\mathrm{E}(Y^{-1}) = \mu^{-1} + \phi^{-1}$, em que $\mathrm{U}_\phi = \partial \mathrm{L}(\mu, \phi)/\partial\phi$.

3. Mostre que o desvio da distribuição gama para o caso i.i.d., ou seja

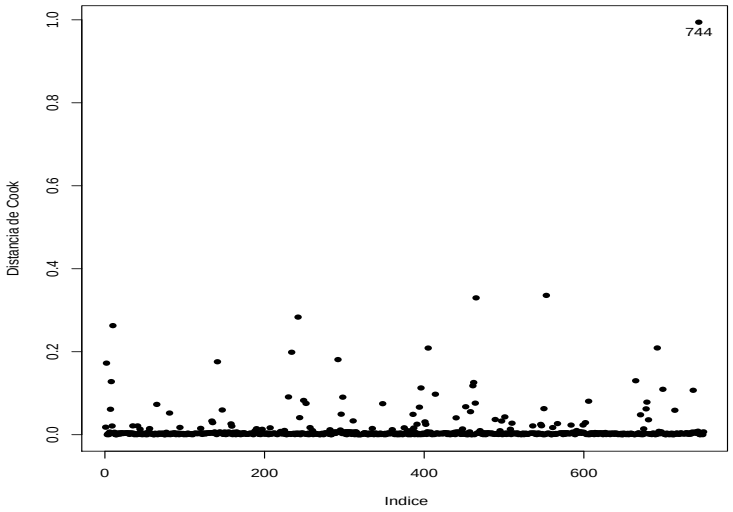

Figura 3.34: Gráfico da distância de Cook para as estimativas dos coeficientes da dispersão referente ao MLG duplo com resposta gama ajustado aos dados sobre *snacks*.

$Y_i \overset{\text{iid}}{\sim} \text{G}(\mu, \phi)$, é dado por $\text{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2n\phi\log(\bar{y}/\tilde{y})$, em que $\tilde{y}$ é a média geométrica das observações, isto é $\tilde{y} = (\Pi_{i=1}^n y_i)^{1/n}$.

4. Sejam $Y_i \sim \text{FE}(\mu_1, \phi_1)$, $i = 1, \ldots, m$, e $Y_i \sim \text{FE}(\mu_2, \phi_2)$, $i = m+1, \ldots, n$, variáveis aleatórias mutuamente independentes. Encontre a estimativa comum de máxima verossimilhança para ϕ_1 e ϕ_2 sob a hipótese $\text{H}_0 : \phi_1 = \phi_2$. Particularize para os casos gama e normal inversa.

5. Supor $Y_i \overset{\text{ind}}{\sim} \text{N}(\mu, \sigma_i^2)$ com $\log(\sigma_i^2) = \alpha + \gamma z_i$, para $i = 1, \ldots, n$. Como fica a matriz modelo $\mathbf{Z}$? Obtenha a estatística do teste da razão de verossimilhanças para testar $\text{H}_0 : \gamma = 0$ contra $\text{H}_1 : \gamma \neq 0$? Qual a distribuição nula assintótica da estatística do teste? Obtenha inicialmente as estimativas para (μ, σ_i^2) sob as hipóteses H_0 e $\text{H}_0 \cup \text{H}_1$.

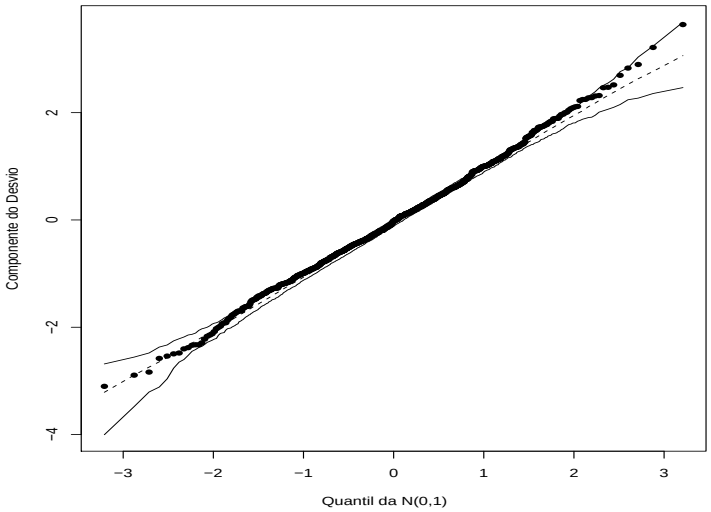

Figura 3.35: Gráfico normal de probabilidades para o resíduo componente do desvio para a média referente ao MLG duplo com resposta gama ajustado aos dados sobre *snacks*.

6. Supor $Y_{ij} \overset{\text{ind}}{\sim} \text{NI}(\mu, \phi_i)$ para $i = 1, 2$ e $j = 1, \ldots, r$ com $\sqrt{\phi_1} = \lambda_1 = \alpha$ e $\sqrt{\phi_2} = \lambda_2 = \alpha + \Delta$. Inicialmente obter as matrizes $\mathbf{Z}$ e $\mathbf{P}$. Em seguida obter as variâncias e covariância assintóticas $\text{Var}(\widehat{\alpha})$, $\text{Var}(\widehat{\Delta})$ e $\text{Cov}(\widehat{\alpha}, \widehat{\Delta})$ deixando em função dos componentes de $\mathbf{P}$. Obter $\widehat{\alpha}$ e $\widehat{\Delta}$ (use a propriedade de invariãncia). Mostre que a estatística do teste de Wald para testar $\text{H}_0 : \Delta = 0$ contra $\text{H}_1 : \Delta \neq 0$ pode ser expressa na forma

$$\xi_W = 2r \frac{\{\sqrt{\widehat{\phi}_2} - \sqrt{\widehat{\phi}_1}\}^2}{\widehat{\phi}_1 + \widehat{\phi}_2}.$$

Mostre que $\widehat{\mu} = (\widehat{\phi}_1 \bar{y}_1 + \widehat{\phi}_2 \bar{y}_2)/(\widehat{\phi}_1 + \widehat{\phi}_2)$. Qual a distribuição nula assintótica da estatística do teste?

7. Na tabela abaixo (Lawless, 2003) são apresentados os resultados de um

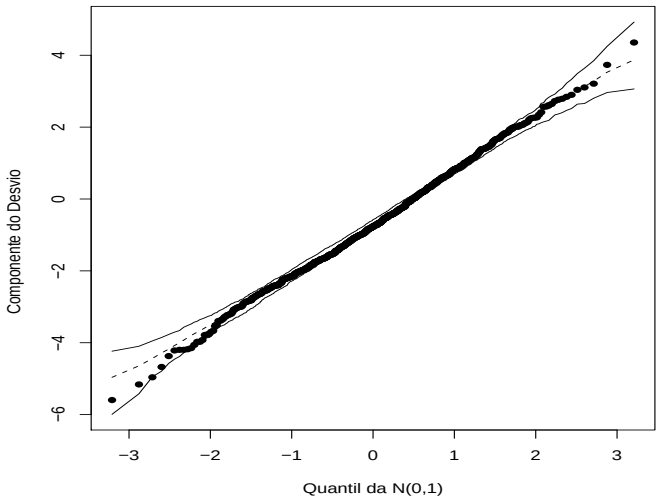

Figura 3.36: Gráfico normal de probabilidades para o resíduo componente do desvio para a dispersão referente ao MLG duplo com resposta gama ajustado aos dados sobre *snacks*.

experimento em que a resistência (em horas) de um determinado tipo de vidro foi avaliada segundo quatro níveis de voltagem (em kilovolts) e duas temperaturas (em graus Celsus). Esses dados estão também disponíveis no arquivo **vidros.txt**. Na primeira coluna do arquivo tem-se o tempo de resistência, na segunda coluna a voltagem(1: 200kV, 2: 250kV, 3: 300kV e 4: 350kV) e na terceira coluna a temperatura (1: 170^{o}C e 2: 180^{o}C). Seja Y_{ijk} o tempo de resistência da k-ésima amostra de vidro submetida à i-ésima temperatura e à j-ésima voltagem.

Faça inicialmente uma análise descritiva dos dados, por exemplo apresentando os perfis médios da resistência segundo a voltagem para os dois níveis de temperatura. Cacule também para cada casela algumas medidas descritivas tais como média, desvio padrão e coeficiente de variação. Comente.

	Voltagem(kV)			
Temperatura (^{o}C)	200	250	300	350
170	439	572	315	258
	904	690	315	258
	1092	904	439	347
	1105	1090	628	588
180	959	216	241	241
	1065	315	315	241
	1065	455	332	435
	1087	473	380	455

O interesse principal desse estudo é comparar as resistências médias, denotadas por μ_{ij}, $i = 1, 2$ e $j = 2, 3, 4$. É usual neste tipo de estudo assumir respostas com alguma distribuição assimétrica. Assim, supor que $Y_{ijk} \overset{\text{ind}}{\sim} \text{G}(\mu_{ij}, \phi)$. Considere inicialmente uma reparametrização tipo casela de referência sem interação, em que $\mu_{11} = \alpha$, $\mu_{1j} = \alpha + \beta_j$, $\mu_{21} = \alpha + \gamma$ e $\mu_{2j} = \alpha + \gamma + \beta_j$ $j = 2, 3, 4$.

Verifique se é possível incluir a interação entre voltagem e temperatura. Procure responder com o modelo final de que forma os níveis de voltagem e temperatura afetam o tempo médio de resistência dos vidros. Apresente, por exemplo, os perfis médios ajustados e interprete a estimativa de dispersão. Faça também uma análise de diagnóstico.

8. Supor $Y_i \overset{\text{iid}}{\sim} \text{NI}(\mu, \phi)$, para $i = 1, \ldots, n$. Mostre que a estatística do teste da razão de verossimilhanças para testar $\text{H}_0 : \phi = 1$ contra $\text{H}_1 : \phi \neq 1$ pode ser expressa na forma

$$\xi_{RV} = n(\hat{\phi}^{-1} - 1) + n \log(\hat{\phi}),$$

e mostre que $\hat{\phi} = n/\text{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})$ é a estimativa de máxima verossimilhança de ϕ. Qual a distribuição nula assintótica da estatística do teste?

9. Sejam $Y_1, \ldots, Y_n$ variáveis aleatórias independentes tais que $Y_i \sim \mathrm{G}(\mu_i, \phi)$ com parte sistemática dada por $\log(\mu_i) = \beta_0 + \beta_1(x_i - \bar{x})$. Responda aos itens abaixo:

   (a) como fica a matriz de informação de Fisher para $\boldsymbol{\theta} = (\beta_0, \beta_1, \phi)^\top$ e a variância assintótica de $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\phi}$?

   (b) Como fica o teste de escore para testar $\mathrm{H}_0 : \phi = 1$ contra $\mathrm{H}_1 : \phi \neq 1$?

   (c) Mostre que a estatística do teste de escore para testar as hipóteses $\mathrm{H}_0 : \beta_0 = 1, \beta_1 = 0$ contra $\mathrm{H}_1 : \beta_0 \neq 1$ ou $\beta_1 \neq 0$ pode ser expressa na forma

$$\xi_{SR} = \frac{\hat{\phi}^0}{e^2}\left[n(\bar{y} - e)^2 + \frac{\{\sum_{i=1}^n (x_i - \bar{x})(y_i - e)\}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right].$$

   Qual a distribuição nula assintótica de ξ_{SR}?

10. Supor $Y_1, \ldots, Y_n$ variáveis aleatórias independentes tais que $Y_i \sim \mathrm{FE}(\mu, \phi_i)$ com $\log(\phi_i) = \alpha + \gamma z_i$. Responda às seguintes questões:

   (i) como fica a matriz modelo $\mathbf{Z}$?

   (ii) Calcule a variância assintótica de $\hat{\gamma}$.

   (iii) Como fica a estatística de escore para testar $\mathrm{H}_0 : \gamma = 0$ contra $\mathrm{H}_1 : \gamma \neq 0$? Qual a distribuição nula assintótica da estatística do teste?

11. Supor $Y_i \overset{\text{ind}}{\sim} \mathrm{N}(\mu, \sigma_i)$, em que $\log(\sigma_i) = \gamma_0 + \gamma_1 z_i$, em que σ_i denota o desvio padrão de Y_i, para $i = 1, \ldots, n$. Obter $\mathbf{U}_\gamma$ e $\mathbf{K}_{\gamma\gamma}$. Como fica a estimação de μ e $\boldsymbol{\gamma}$? Obtenha a estatística do teste da razão de verossimilhanças para testar $\mathrm{H}_0 : \gamma_1 = 0$ contra $\mathrm{H}_1 : \gamma_1 \neq 0$. Qual a distribuição nula assintótica da estatística do teste?

12. Supor $Y_i \overset{\text{iid}}{\sim} \text{NI}(\mu, \phi)$, para $i = 1, \ldots, n$, em que $\gamma = \log(\phi)$. Obter a estimativa de máxima verossimilhança $\hat{\gamma}$ (dado $\hat{\phi}$) e $\text{K}_{\gamma\gamma}$. Como fica a estatística do teste de Wald para testar $\text{H}_0 : \gamma = 0$ contra $\text{H}_1 : \gamma \neq 0$?

13. Sejam $Y_1, \ldots, Y_n$ variáveis aleatórias independentes tais que $Y_i \sim \text{NI}(\mu_i, \phi)$ com $\mu_i^{-1} = \alpha$. Encontre $\hat{\alpha}$ e $\text{Var}(\hat{\alpha})$. Como fica a estatística de Wald para testar $\text{H}_0 : \alpha = 1$ contra $\text{H}_1 : \alpha \neq 1$? Qual a distribuição nula assintótica da estatística do teste?

14. Supor $Y_1, \ldots, Y_n$ variáveis aleatórias independentes tais que $Y_i \sim \text{NI}(\mu_i, \phi)$ e $\sqrt{\mu_i} = \eta_i^{-1}$ com $\eta_i = \alpha + \beta(x_i - \bar{x})$, em que $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$, $i = 1, \ldots, n$. Responda às seguintes questões:

    (i) como fica a matriz modelo $\mathbf{X}$?

    (ii) Calcule as variâncias assintóticas $\text{Var}(\hat{\alpha})$ e $\text{Var}(\hat{\beta})$. Calcule $\text{Cov}(\hat{\alpha}, \hat{\beta})$ e comente.

    (iii) Como fica a estatística de Wald para testar $\text{H}_0 : \beta = 0$ contra $\text{H}_1 : \beta \neq 0$? Qual a distribuição nula assintótica da estatística do teste?

15. Supor $Y_i \overset{\text{ind}}{\sim} \text{G}(\mu_i, \sigma_i)$, em que $\log(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ e $\log(\sigma_i) = \lambda_i = \mathbf{z}_i^\top \boldsymbol{\gamma}$, em que σ_i denota o coeficiente de variação de Y_i, para $i = 1, \ldots, n$. Obter $\mathbf{U}_\beta$, $\mathbf{U}_\gamma$, $\mathbf{K}_{\beta\beta}$ e $\mathbf{K}_{\gamma\gamma}$ e desenvolva um processo iterativo duplo para obter as estimativas de máxima verossimilhança $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$.

16. Sejam $Y_i, i = 1, \ldots, n$, variáveis aleatórias independentes com distribuição gama de média μ_i e parâmetro de precisão ϕ. Mostre que a estatística da razão de verossimilhanças para testar $\text{H}_0 : \phi = 1$ contra $\text{H}_1 : \phi \neq 1$ vale

$$\xi_{RV} = 2n[\log(\hat{\phi}) - \log\Gamma(\hat{\phi}) - (\hat{\phi} - 1)\{1 - \psi(\hat{\phi})\}],$$

em que $\Gamma(\phi)$ é a função gama e $\psi(\phi)$ é a função digama (Cordeiro et al., 1994). Use o resultado $\log(\hat{\phi}) - \psi(\hat{\phi}) = \bar{\mathrm{D}}/2$, em que $\bar{\mathrm{D}} = \sum_{i=1}^{n} \mathrm{D}(y_i; \hat{\mu}_i)/n$ denota o desvio médio do modelo correspondente.

17. Supor Y_{ij} variáveis aleatórias mutuamente independentes tais que $Y_{ij} \sim \mathrm{G}(\mu_i, \phi)$ para $i = 1, 2$ e $j = 1, \ldots, m$, sendo $\log(\mu_1) = \alpha - \beta$ e $\log(\mu_2) = \alpha + \beta$. (i) Obtenha a matrix modelo $\mathbf{X}$. (ii) Expresse em forma fechada as estimativas de máxima verossimilhança $\hat{\alpha}$ e $\hat{\beta}$. (iii) Calcule as variâncias assintóticas $\mathrm{Var}(\hat{\alpha})$ e $\mathrm{Var}(\hat{\beta})$ e mostre que $\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = 0$. (iv) Como fica o teste de escore para testar $\mathrm{H}_0 : \beta = 0$ contra $\mathrm{H}_1 : \beta \neq 0$? Qual a distribuição nula assintótica da estatística do teste?

18. No arquivo **energy.txt** estão descritos os dados referentes ao consumo de energia em 53 domicílios (Montgomery et al., 2001, pgs. 175-178) em que o total de energia consumido num determinado mês (em kilowatts-hora) é utilizado para explicar a demanda de energia na hora de pico. Faça inicialmente uma análise descritiva dos dados. Use um modelo de regressão normal linear (com erros homocedásticos) para explicar a demanda média no horário de pico através do consumo mensal de energia. Proponha, alternativamente, modelos com erros heteroscedásticos. Compare os ajustes e interprete os coeficientes estimados do modelo escolhido.

19. A fim de avaliar-se a qualidade de um determinado filme utilizado em máquinas fotográficas, o tempo de duração do filme (em horas) é relacionado com a densidade máxima do filme sob três condições experimentais conforme descrito na tabela abaixo (Myers et al., 2002, p. 192) e também no arquivo **dfilme.txt**.

Tempo	D_{max} (72^oC)	Tempo	D_{max} (82^oC)	Tempo	D_{max} (92^oC)
72	3,55	48	3,52	24	3,46
144	3,27	96	3,35	48	2,91
216	2,89	144	2,50	72	2,27
288	2,55	192	2,10	96	1,49
360	2,34	240	1,90	120	1,20
432	2,14	288	1,47	144	1,04
504	1,77	336	1,19	168	0,65

Faça um diagrama de dispersão entre o tempo de duração e a densidade máxima para cada condição experimental e proponha um modelo com resposta gama para ajustar esses dados. Aplique métodos de diagnóstico e interprete as estimativas do modelo selecionado.

20. O arquivo **claims.txt** contém uma amostra aleatória de 996 apólices de seguros de veículos extraídas do livro de de Jong e Heller (2008) referente ao período 2004-2005. As variáveis do arquivo estão na seguinte ordem : (i) `valorv` (valor do veículo em 10000 dolares australianos), (ii) `expos` (exposição do veículo), (iii) `nsinistros` (número de sinistros no período), (iv) `csinistros` (custo total dos sinistros em dolares australianos), (v) `tipov` (tipo do veículo em 11 categorias), (vi) `idadev` (idade do veículo em 4 categorias), (vii) `sexoc` (sexo do condutor principal), (viii) `areac` (área de residência do condutor principal) e (ix) `idadec` (idade do condutor principal em 6 categorias).

    Faça inicialmente uma análise descritiva dos dados e procure agrupar em um número menor de categorias algumas variáveis categóricas. Considere como variável resposta `cmsinistros = csinistros/nsinistros`. Aplique numa primeira etapa modelos com resposta gama e normal inversa com ϕ constante. Faça uma análise de diagnóstico. Numa se-

gunda etapa, se necessário, aplique modelos duplos com resposta gama e normal inversa. Faça também uma análise de diagnóstico. Para o modelo final selecionado interprete os coeficientes estimados.

21. O arquivo **fluxo.txt** contém parte dos dados de um experimento desenvolvimento em 2006 nas Faculdades de Medicina e de Filosofia, Letras e Ciências Humanas da USP e analisado no Centro de Estatística Aplicada do IME-USP (CEA0P16) para avaliar o fluxo da fala de falantes do Português Brasileiro segundo o gênero, idade e escolaridade. Uma amostra de 595 indivíduos residentes na cidade de São Paulo com idade entre 2 e 99 anos foi avaliada segundo a fala auto-expressiva. O indivíduo era apresentado a uma figura e orientado a discorrer sobre a mesma durante um tempo mínimo de 3 minutos e máximo de 6 minutos. Para crianças de 2 e 3 anos, as amostras foram obtidas com a colaboração dos pais.

    As variáveis estão descritas na seguinte ordem: (i) `idade`, idade em anos do indivíduo, (ii) `gênero`, gênero do indivíduo (1:feminino, 2:masculino), (iii) `interj`, número de interjeições durante o discurso, (iv) `fpm`, fluxo de palavras por minuto e (v) `fsm`, fluxo de sílabas por minuto. Para ler o arquivo no R use os comandos

    ```
    fluxo = read.table("fluxo.txt", header=TRUE)
    ```

    ```
    genero = factor(genero).
    ```

    Faça inicialmente uma análise descritiva dos dados, boxplots individuais e diagramas de dispersão de cada variável explicativa contra `fpm` (que será assumida como resposta). Depois proponha um modelo linear normal homocedástico e verifique a possibilidade de também modelar a variância.

Ajustar um modelo normal heterocedástico usando o `GAMLSS`. Note que neste caso é modelado o desvio padrão ao invés da variância, como ocorre na biblioteca `dglm`. Interpretar os gráficos gerados pelos comandos `plot`, `wp` e `term.plot`.

22. No arquivo **rent** do GAMLSS são descritas 9 variáveis observadas numa amostra aleatória de 1967 unidades habitacionais da cidade de Munich em 1993. Para fins de análise iremos considerar as seguintes variáveis: (i) `R` (valor mensal líquido do alugel em DM), (ii) `Fl` (área útil em m^2), (iii) `A` (ano da construção), (iv) `H` (variável binária referente à existência de aquecimento central, 0: sim, 1: não) e (v) `loc` (qualidade da localização do imóvel, 1: abaixo da média, 2: na média e 3: acima da média). O arquivo está disponibilizado diretamente no GAMLSS, no entanto é preciso informar que a variável `loc` é categórica através do comando

    ```
    loc=factor(loc).
    ```

    A variável explicativa `A` é considerada contínua. Fazer inicialmente uma análise descritiva dos dados, tais como densidade da variável resposta, boxplots e diagramas de dispersão entre as variáveis explicativas contínuas e a variável resposta. Procure selecionar um modelo gama duplo com ligação logarítmica para explicar o valor médio mensal do aluguel e o coeficiente de dispersão. Fazer uma análise de diagnóstico e interpretar os coeficientes estimados do moelo selecionado.

23. Considere o arquivo **BigMac2003** da biblioteca `alr4` do R, em que são descritas as seguintes variáveis de 69 cidades de diversos países:

    - `BigMac`: minutos de trabalho para comprar um Big Mac

- `Bread`: minutos de trabalho para comprar 1kg de pão
- `Rice`: minutos de trabalho para comprar 1kg de arroz
- `FoodIndex`: índice de preços de alimentos
- `Bus`: valor da passagem de ônibus (em USD)
- `Apt`: valor do aluguel (em USD) de um apartamento padrão de 3 dormitórios
- `TeachGI`: salário bruto anual (em 1000 USD) de um professor de ensino fundamental
- `TeachNI`: salário líquido anual (em 1000 USD) de um professor de ensino fundamental
- `TaxRate`: imposto pago (em porcentagem) por um professor de ensino fundamental
- `TeachHours`: carga horária semanal (em horas) de um professor de ensino fundamental.

Para disponibilizar e visualizar um resumo dos dados use na sequência os seguintes comandos do R:

```
require(alr4)

require(MASS)

attach(BigMac2003)

summary(BigMac2003).
```

O objetivo principal do estudo é relacionar a variável `BigMac` com as demais variáveis explicativas. Apresente a densidade da variável resposta, as correlações lineares amostrais bem como os diagramas de dispersão

(com tendência) entre a variável resposta e cada uma das variáveis explicativas. Comente. Padronize as variáveis explicativas. Por exemplo, para padronizar a variável explicativa `Bread` use o comando

`sBread = scale(Bread, center = TRUE, scale = TRUE).`

Ajustar inicialmente um modelo com resposta gama e ligação logarítmica no GAMLSS através do comando

`fit1.bigmac = gamlss(BigMac` $\sim$ `·, family=GA, data=BigMac2003).`

Através do procedimento `stepGAIC` fazer uma seleção das variáveis explicativas

`fit2.bigmac = stepGAIC(fit1.bigmac).`

Para o submodelo selecionado aplicar análises de resíduos através dos comandos `plot(fit2.bigmac)` e `wp(big.mac)`. Construir o gráfico da distância de Cook. Comente. Classifique as variáveis explicativas segundo o impacto na explicação da média da variável resposta. Apresente e comente o `term.plot(fit2.big.mac)`.

24. No arquivo **raia.txt** (Paula e Kumagaia, 2014) são descritas as seguintes variáveis observadas numa amostra de 186 descarregamentos pesqueiros na Bahia de todos os Santos (costa nordeste brasileira), no período de janeiro de 2012 a janeiro de 2013, referentes à captura da raia-branca através do método artesanal grozeira: (i) `periodo` (período da pesca, seco ou chuvoso), (ii) `local` (local da pesca, área1, área2, área3 e área4), (iii) `mare` (maré, quadratura ou sizígia), (iv) `vvento` (velocidade do vento, em m/s), (v) `tmax` (temperatura máxima, em ^{o}C), (vi) `tmin` (temperatura mínima, em ^{o}C), (vii) `ins` (insolação, em horas) e (viii) `cpue` (captura por unidade de esforço, em kg). As variáveis (iii) a (vii) foram observadas no local de pesca.

O objetivo principal do estudo é relacionar a cpue média com as demais variáveis explicativas. Para ler esse arquivo no R faça o seguinte:

`raia = read.table("raia.txt", header=TRUE).`

Para deixar o arquivo disponível use o comando

`attach(raia).`

Informar que as variáveis `local` e `maré` são categóricas

```
raia$local = factor(raia$local, levels=1:4, labels=c("área1",
"área2", "área3", "área4"))
```

```
raia$mare = factor(raia$mare, levels=1:2, labels=c("quadratura",
"sizígia")).
```

Faça inicialmente uma análise descritiva construindo boxplots robustos e diagramas de dispersão de cada variável explicativa contínua contra a variável resposta `cpue`. Calcule também as correlações lineares entre as variáveis. Comente.

Proponha um modelo gama com ligação logarítmica com todas as variáveis explicativas e ajuste no `R` usando o comando `glm`. Use o comando `stepAIC` para selecionar um submodelo. Tente incluir interações de $1^{\underline{a}}$ ordem ao nível de significância de 10%. Obtenha $\widehat{\phi}$ e o correspondente erro padrão estimado, além da função desvio e o respectivo valor-P. Construa o gráfico da distância de Cook e do envelope gerado com o resíduo componente do desvio. Verifique o impacto das observações atípicas e interprete os coeficientes do modelo final.

Finalmente, ajustar o modelo final pelo `GAMLSS`. Comente os gráficos de resíduos quantílicos gerados pelos comandos `plot` e `wp`.

25. A seguir é descrito um conjunto de dados em que pacientes com leucemia foram classificados segundo a ausência ou presença de uma característica morfológica nas células brancas (Feigl e Zelen, 1965).

AG Positivo		AG Negativo	
WBC	Tempo	WBC	Tempo
2300	65	4400	56
750	156	3000	65
4300	100	4000	17
2600	134	1500	7
6000	16	9000	16
10500	108	5300	22
10000	121	10000	3
17000	4	19000	4
5400	39	27000	2
7000	143	28000	3
9400	56	31000	8
32000	26	26000	4
35000	22	21000	3
100000	1	79000	30
100000	1	100000	4
52000	5	100000	43
100000	65		

Pacientes classificados de AG positivo foram aqueles com a presença da característica e pacientes classificados de AG negativo não apresentaram a característica. É apresentado também o tempo de sobrevivência do paciente (em semanas) após o diagnóstico da doença e o número de células brancas (WBC) no momento do diagnóstico. Esses dados estão descritos no arquivo **sobrev.txt**. Supondo que o tempo de sobrevivência após o diagnóstico segue uma distribuição gama, proponha um modelo para explicar o tempo médio de sobrevivência dados log(WBC)

e AG(=1 positivo, =0 negativo). Faça uma análise de diagnóstico com o modelo ajustado e interprete as estimativas.

Ajustar o modelo no `GAMLSS`. Interpretar os gráficos gerados pelos comandos `plot`, `wp` e `term.plot`.

Capítulo 4

Modelos para Dados Binários

4.1 Introdução

Neste capítulo serão apresentados modelos para a análise de dados com resposta binária, isto é, resposta que admite apenas dois resultados. Comumente é chamado de sucesso o resultado mais importante da resposta ou aquele que pretende-se relacionar com as demais variáveis de interesse. É comum encontrar situações práticas em que esse tipo de resposta aparece. Como ilustração, seguem alguns exemplos: (i) o resultado do diagnóstico de um exame de laboratório, positivo ou negativo; (ii) o resultado da inspeção de uma peça recém fabricada, defeituosa ou não defeituosa; (iii) a opinião de um eleitor a respeito da implantação do voto distrital, favorável ou outra opinião; (iv) o resultado de um teste de aptidão aplicado a um estudante, aprovado ou reprovado; (v) classificação de um cliente de uma instituição financeira com relação a um empréstimo para financiamento imobiliário, adimplente ou inadimplente; (vi) o resultado de uma promoção de uma rede de lojas enviando para cada cliente um cupom com desconto, cupom utilizado ou cupom não utilizado num determinado período, etc. Há também situações em que apenas duas possibilidades são consideradas de interesse para uma variável contínua, valo-

res menores do que um valor de referência v_0 e valores maiores ou iguais a v_0. Nesses casos, pode-se considerar uma nova variável binária para essas duas possibilidades. Por exemplo, numa determinada prova de conhecimentos v_0 pode ser a nota mínima para ser aprovado no exame, ou o valor mínimo para um exame de laboratório ser considerado alterado. Assim, variáveis binárias podem surgir naturalmente num experimento ou serem criadas dependendo do interesse do estudo.

Inicialmente, uma resenha dos principais métodos clássicos para a análise de tabelas de contingência do tipo 2×2 será apresentada neste capítulo. Em seguida, será descrito o modelo de regressão logística para a análise de tabelas de contingência 2×2. Também serão discutidos procedimentos para a seleção de variáveis em modelos logísticos, métodos de diagnóstico, alguns tipos de modelos de dose-resposta, sobredispersão e regressão logística condicional.

4.2 Métodos clássicos: uma única tabela 2×2

Métodos clássicos em tabelas de contingência 2×2 são datados da década de 1950. Os primeiros trabalhos foram motivados pelo interesse na inferência de certos parâmetros com grande aplicabilidade na área biomédica, especialmente em Epidemiologia, tais como risco relativo e razão de chances. Vários trabalhos foram publicados durante as décadas de 1950 e 1960 e até hoje as técnicas desenvolvidas têm sido utilizadas, particularmente na análise descritiva dos dados, antes de um tratamento mais sofisticado através de modelagem estatística de regressão. Nesta seção será apresentada uma resenha das principais técnicas segundo o ponto de vista inferencial clássico. Embora a metodologia apresentada possa ser aplicada em qualquer área do conhecimento, será dado ênfase para a área biomédica em que tem ocorrido um número maior de aplicações.

4.2.1 Risco relativo

Supor que os indivíduos de uma determinada população sejam classificados segundo um fator com dois níveis, A e B, e a presença ou ausência de uma certa doença, denotados por D e $\bar{D}$, respectivamente. As proporções populacionais ficam, nesse caso, descritas conforme a tabela abaixo.

	Fator	
Doença	A	B
D	P_1	P_3
$\bar{D}$	P_2	P_4

Portanto, pode-se definir as seguintes quantidades:

$P_1/(P_1+P_2)$: proporção de indivíduos classificados como doentes no grupo A e

$P_3/(P_3+P_4)$: proporção de indivíduos classificados como doentes no grupo B.

A razão entre as duas proporções acima foi denominada por Cornfield (1951) como sendo o risco relativo de doença entre os níveis A e B, ou seja

$$\text{RR} = \frac{P_1/(P_1+P_2)}{P_3/(P_3+P_4)} = \frac{P_1(P_3+P_4)}{P_3(P_1+P_2)}. \tag{4.1}$$

Cornfield (1951) também notou que se a doença for rara ($P_1 << P_2$ e $P_3 << P_4$) a quantidade (4.1) assume a forma simplificada

$$\psi = \frac{P_1 P_4}{P_3 P_2}, \tag{4.2}$$

a qual denominou *odds ratio*, que será denominada razão de chances. Muitas vezes é comum ψ ser chamado de risco relativo, embora isso somente seja válido quando P_1 e P_3 forem muito pequenos. A grande vantagem do uso de ψ

é a facilidade inferencial tanto na abordagem tradicional como na abordagem através de regressão.

Como em geral a porcentagem de indivíduos doentes é muito menor do que a porcentagem de não doentes, é bastante razoável num estudo cujo objetivo é avaliar a associação entre algum fator particular e uma certa doença, que a quantidade de doentes na amostra seja a maior possível. Assim, a amostragem retrospectiva, em que os indivíduos são escolhidos separadamente nos estratos D e $\bar{D}$, pode ser mais conveniente do que os demais procedimentos amostrais. Um cuidado, entretanto, deve-se ter nesses estudos. É importante que os doentes (casos) sejam comparáveis aos não doentes (controles) segundo outros fatores (fatores potenciais de confundimento), possivelmente associados com a doença. Nos estudos prospectivos, em que a amostragem é feita nos estratos A e B, esse tipo de problema pode ser controlado, embora em geral seja necessário um longo período até a obtenção de um número suficiente de doentes para uma análise estatística mais representativa.

As inferências para os estudos retrospectivos e prospectivos são idênticas, assim será descrito apenas o caso retrospectivo. Assim, assume-se que no estrato D são amostrados n_1 indivíduos e que no estrado $\bar{D}$ são amostrados n_2 indivíduos. O número observado de indivíduos com presença de A nos estratos D e $\bar{D}$ será denotado por y_1 e y_2, respectivamente. Os dados resultantes dessa amostragem podem ser resumidos conforme a tabela abaixo.

	Fator		
Doença	A	B	Total
D	y_1	$n_1 - y_1$	n_1
$\bar{D}$	y_2	$n_2 - y_2$	n_2

Esse tipo de abordagem pode ser estendida para quaisquer situações práticas em que pretende-se comparar dois estratos de uma determinada

população segundo a ocorrência de algum evento de interesse. Por exemplo, A poderia denotar os condutores do sexo masculino com apólice de seguro de automóvel de uma seguradora, enquanto B denotaria os condutores do sexo feminino da mesma seguradora. O evento D poderia ser a utilização da apólice para cobrir alguma sinistralidade num determinado período. Assim, pode-se estimar a razão de chances entre condutores do sexo masculino e condutores do sexo feminino de utilização da apólice para cobrir sinistralidade. Como o evento D neste caso não deve ser raro, risco relativo e razão de chances devem ser quantidades diferentes. A seguir será discutida a abordagem clássica para analisar a tabela acima.

4.2.2 Modelo probabilístico não condicional

Denota-se por Y_1 e Y_2 o número de indivíduos com presença de A nos estratos D e $\bar{D}$, respectivamente. Será também assumido que essas variáveis são binomiais independentes, isto é $Y_1 \sim \mathrm{B}(n_1, \pi_1)$ e $Y_2 \sim \mathrm{B}(n_2, \pi_2)$, respectivamente. Logo, a função de probabilidade conjunta de (Y_1, Y_2) fica dada por

$$f(y_1, y_2; \pi_1, \pi_2) = \binom{n_1}{y_1}\binom{n_2}{y_2}\pi_1^{y_1}\pi_2^{y_2}(1-\pi_1)^{n_1-y_1}(1-\pi_2)^{n_2-y_2}. \qquad (4.3)$$

Seguindo a notação da seção anterior, tem-se que $\pi_1 = P_1/(P_1+P_3)$, $1-\pi_1 = P_3/(P_1+P_3)$, $\pi_2 = P_2/(P_2+P_4)$ e $1-\pi_2 = P_4/(P_2+P_4)$. Assim, mostra-se que

$$\psi = \frac{P_1P_4}{P_3P_2} = \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)},$$

e consequentemente que $\pi_1 = \pi_2\psi/\{\pi_2\psi + 1 - \pi_2\}$. A expressão (4.3) pode então ser expressa apenas em função de (ψ, π_2),

$$\begin{aligned} f(y_1, y_2; \psi, \pi_2) &= \exp\left[\log\left\{\binom{n_1}{y_1}\binom{n_2}{y_2}\right\} + y_1\log(\psi) + (y_1+y_2)\log\left(\frac{\pi_2}{1-\pi_2}\right)\right] \times \\ &\quad \times \frac{(1-\pi_2)^n}{(\psi\pi_2 + 1 - \pi_2)^{n_1}}, \end{aligned} \qquad (4.4)$$

em que $n = n_1 + n_2$. O logaritmo da função de verossimilhança fica portanto dado por

$$\begin{aligned} \mathrm{L}(\psi, \pi_2) &= \log\left\{\binom{n_1}{y_1}\binom{n_2}{y_2}\right\} + y_1\log(\psi) + (y_1+y_2)\log\left(\frac{\pi_2}{1-\pi_2}\right) + \\ &\quad + n\log(1-\pi_2) + n_1\log(\psi\pi_2 + 1 - \pi_2). \end{aligned}$$

Pode-se mostrar que a maximização de $\mathrm{L}(\psi, \pi_2)$ leva às estimativas de máxima verossimilhança $\tilde{\pi}_2 = \frac{y_2}{n_2}$ e $\tilde{\psi} = \frac{y_1(n_2-y_2)}{y_2(n_1-y_1)}$, respectivamente.

A variância assintótica de $\tilde{\psi}$ é expressa na forma

$$\mathrm{Var}_A(\tilde{\psi}) = \psi^2\left\{\frac{1}{n_1\pi_1(1-\pi_1)} + \frac{1}{n_2\pi_2(1-\pi_2)}\right\}.$$

Esse resultado pode ser mostrado utilizando teoria assintótica relacionada com os estimadores de máxima verossimilhança de π_2 e ψ, sendo as correspondentes variâncias assintóticas obtidas através da inversa da matriz de informação de Fisher para (ψ, π_2). Outra possibilidade para mostrar esse resultado é através da aplicação do método delta, em que obtém-se a variância assintótica de $\tilde{\psi}$ através das variâncias e covariância assintóticas de $\tilde{\pi}_1$ e $\tilde{\pi}_2$. Assim, para n_1 e n_2 grandes, tem-se que

$$\mathrm{Var}_A(\tilde{\psi}) = \left[\frac{\partial\psi}{\partial\boldsymbol{\pi}}\right]^\top \mathrm{Var}_A(\tilde{\boldsymbol{\pi}}) \left[\frac{\partial\psi}{\partial\boldsymbol{\pi}}\right],$$

em que $[\partial\psi/\partial\boldsymbol{\pi}] = [\partial\psi/\partial\pi_1, \partial\psi/\partial\pi_2]^\top$ e $\mathrm{Var}_A(\tilde{\boldsymbol{\pi}}) = \mathrm{diag}\{\frac{\pi_1(1-\pi_1)}{n_1}, \frac{\pi_2(1-\pi_2)}{n_2}\}$ com $\boldsymbol{\pi} = (\pi_1, \pi_2)^\top$.

Alguns autores preferem trabalhar com $\log(\psi)$, uma vez que a aproximação assintótica para a distribuição normal é mais rápida para $\log(\tilde{\psi})$ do que para $\tilde{\psi}$. Assim, pode-se mostrar sob condições gerais de regularidade, que a estimativa não condicional $\log(\tilde{\psi})$ segue para grandes amostras distribuição normal de média $\log(\psi)$ e variância assintótica $\mathrm{Var}_A\{\log(\tilde{\psi})\} =$

$\{1/n_1\pi_1(1-\pi_1) + 1/n_2\pi_2(1-\pi_2)\}$. Esse resultado também pode ser facilmente mostrado através do método delta aplicado à transformação $\log(\psi)$, em que

$$\text{Var}_A\{\log(\tilde{\psi})\} = \{d\log(\psi)/d\psi\}^2\text{Var}_A(\tilde{\psi})$$

com $d\log(\psi)/d\psi = 1/\psi^2$.

Em virtude de $\text{E}(\tilde{\psi}) = \infty$, que impossibilita qualquer tipo de inferência para pequenas amostras, testes exatos usando um modelo condicional tem sido preferido. Esses testes serão discutidos na próxima seção.

4.2.3 Modelo probabilístico condicional

Devido aos problemas inferenciais com o modelo não condicional para pequenas amostras, a utilização de um modelo condicional, cuja construção será discutida a seguir, tem sido a solução encontrada sob o ponto de vista clássico para fazer inferências a respeito de ψ.

Assim, aplicando o teorema da fatorização para a função de probabilidade (4.4), mostra-se que o conjunto de estatísticas $(Y_1, Y_1 + Y_2)$ é suficiente minimal para o vetor de parâmetros $[\log\psi, \log\{\pi_2/(1-\pi_2)\}]$. Logo, a distribuição de (Y_1, Y_2) condicionada a $Y_1 + Y_2 = m$, deverá resultar numa função de probabilidade que depende apenas do parâmetro de interese ψ. Essa distribuição resultante (ver Cornfield, 1956) tem sido largamente utilizada em pequenas amostras. Alguns autores questionam, entretanto, o procedimento adotado, uma vez que a estatística $Y_1 + Y_2$ não é ancilar para ψ; isto é, contém informações a respeito do parâmetro ψ (ver discussão, por exemplo, em Lehnman e Casella, 2011).

O condicionamento de (Y_1, Y_2) em $Y_1 + Y_2 = m$ produz o modelo caracterizado pela família de distribuições hipergeométricas não centrais, cuja função

de probabilidade é definida por

$$f(y_1|m;\psi) = \frac{\binom{n_1}{y_1}\binom{n_2}{m-y_1}\psi^{y_1}}{\sum_t \binom{n_1}{t}\binom{n_2}{m-t}\psi^t}, \tag{4.5}$$

em que $0 < \psi < \infty$ e t varia de $\max(0, m - n_2)$ a $\min(n_1, m)$. Em particular, quando $\psi = 1$, a expressão (4.5) fica reduzida à conhecida distribuição hipergeométrica central, com função de probabilidade dada por

$$f(y_1|m;\psi = 1) = \frac{\binom{n_1}{y_1}\binom{n_2}{m-y_1}}{\binom{n_1+n_2}{m}}.$$

A média e a variância de $Y_1|m$ são, respectivamente, dadas por

$$\mathrm{E}(1) = \mathrm{E}(Y_1|m;\psi = 1) = \frac{mn_1}{n}$$

e

$$\mathrm{V}(1) = \mathrm{Var}(Y_1|m;\psi = 1) = \frac{n_1 n_2 (n-m)m}{n^2(n-1)}.$$

Para o modelo condicional (4.5) o logaritmo da função de verossimilhança fica expresso na forma

$$\mathrm{L}(\psi) = \log\left\{\binom{n_1}{y_1}\binom{n_2}{y_2}\right\} + y_1\log(\psi) - \log\left\{\sum_t \binom{n_1}{t}\binom{n_2}{m-t}\psi^t\right\}.$$

Denote por $\hat{\psi}$ a estimativa de máxima verossimilhança condicional. Essa estimativa pode ser expressa como a solução positiva da equação $y_1 = \mathrm{E}(Y_1|m;\hat{\psi})$. Tem-se que o momento de ordem r da distribuição condicional, $\mathrm{E}(Y_1^r|m;\psi)$ é dado por $\mathrm{E}(Y_1^r|m;\psi) = P_r(\psi)/P_0(\psi)$, em que

$$P_r(\psi) = \sum_t t^r \binom{n_1}{t}\binom{n_2}{m-t}\psi^t, \quad r = 1, 2, \ldots$$

e $P_0(\psi) = \sum_t \binom{n_1}{t}\binom{n_2}{m-t}\psi^t$. Assim, a equação de máxima verossimilhança para obter $\hat{\psi}$ fica reescrita na forma

$$y_1 - \frac{P_1(\hat{\psi})}{P_0(\hat{\psi})} = 0. \tag{4.6}$$

Com o aumento de n_1, n_2, m e $n - m$, fica impraticável obter $\hat{\psi}$ através de (4.6), uma vez que essa equação contém polinômios em $\hat{\psi}$ de grau bastante elevado. Uma saída, nesses casos, é resolver (4.6) através de métodos numéricos que não requerem a extração das raízes do polinômio $P_1(\psi)P_0^{-1}(\psi)$ (ver McCullagh e Nelder, 1989, p. 256 ; Silva, 1992).

Para ilustrar a obtenção de $\hat{\psi}$, considere a tabela abaixo.

	A	B	Total
D	1	3	4
$\bar{D}$	1	2	3

Tem-se, nesse caso, que $n_1 = 4, n_2 = 3$ e $m = 2$. A função de probabilidade da distribuição condicional fica então dada por

$$f(y_1|m;\psi) = \binom{4}{y_1}\binom{3}{2-y_1}\psi^{y_1} / \sum_t \binom{4}{t}\binom{3}{2-t}\psi^t,$$

em que o somatório varia no intervalo $0 \leq t \leq 2$. Isso resulta nas probabilidades condicionais

$$\begin{aligned} f(0|m;\psi) &= 3/\{3 + 12\psi + 6\psi^2\} \\ f(1|m;\psi) &= 12\psi/\{3 + 12\psi + 6\psi^2\} \text{ e} \\ f(2|m;\psi) &= 6\psi^2/\{3 + 12\psi + 6\psi^2\}. \end{aligned}$$

A equação $\mathrm{E}(Y_1|m;\hat{\psi}) = y_1$ fica então dada por

$$12\hat{\psi} + 12\hat{\psi}^2 = 3 + 12\hat{\psi} + 6\hat{\psi}^2,$$

que é equivalente a $6\hat{\psi}^2 = 3$ ou $\hat{\psi} = 0,707$.

Similarmente ao estimador não condicional, pode-se mostrar para grandes amostras que $\hat{\psi}$ segue distribuição normal de média ψ e variância assintótica $\mathrm{Var}_A(\hat{\psi}) = \mathrm{V}_A^{-1}(\psi)$, em que

$$\mathrm{V}_A^{-1}(\psi) = \left[\frac{1}{\mathrm{E}_A(\psi)} + \frac{1}{n_1 - \mathrm{E}_A(\psi)} + \frac{1}{m - \mathrm{E}_A(\psi)} + \frac{1}{n_2 - m + \mathrm{E}_A(\psi)}\right],$$

e $\mathrm{E}_A(\psi)$ sai da equação

$$\frac{\mathrm{E}_A(\psi)\{n_2 - m + \mathrm{E}_A(\psi)\}}{\{n_1 - \mathrm{E}_A(\psi)\}\{m - \mathrm{E}_A(\psi)\}} = \psi, \tag{4.7}$$

que para ψ fixo resulta numa equação quadrática em $\mathrm{E}_A(\psi)$. Mostra-se, para $\psi \neq 1$, que a única raiz de (4.7) que satisfaz $\max(0, m - n_2) \leq \mathrm{E}_A(\psi) \leq \min(n_1, m)$ é dada por

$$\mathrm{E}_A(\psi) = ||r| - s|,$$

em que $r = \frac{1}{2}[n/(\psi - 1) + m + n_1]$ e $s = [r^2 - mn_1\psi/(\psi - 1)]^{\frac{1}{2}}$.

Quando $\psi = 1$, a expressão (4.7) não resulta numa forma quadrática em $\mathrm{E}_A(\psi)$. Verifica-se facilmente, nesse caso, que

$$\mathrm{E}_A(1) = \frac{mn_1}{n}$$

e

$$\mathrm{V}_A(1) = \frac{n_1 n_2 m(n - m)}{n^3}.$$

Pode-se notar que a média e a variância assintótica de $\hat{\psi}$, quando $\psi = 1$, coincidem praticamente com a média e a variância da distribuição condicional dada em (4.5).

4.2.4 Teste de hipóteses

Testes exatos

Uma vez conhecida a distribuição condicional que depende apenas do parâmetro de interesse ψ, pode-se desenvolver testes exatos para pequenas amostras.

Um caso de interesse seria testar $\mathrm{H}_0 : \psi = \psi_0$ contra $\mathrm{H}_1 : \psi < \psi_0$, em que ψ_0 é um valor conhecido. O nível descritivo (valor-P) do teste, isto é, a probabilidade sob H_0 de obtenção de valores tão ou mais desfavoráveis a H_0 (no sentido de H_1) é definido por

$$\mathrm{P}_I = \sum_{t \leq y_1} f(t|m; \psi_0),$$

em que o somatório vai de $\max(0, m - n_2)$ até y_1. Analogamente, para testar $\mathrm{H}_0 : \psi = \psi_0$ contra $\mathrm{H}_1 : \psi > \psi_0$, tem-se que $\mathrm{P}_S = \sum_{t \geq y_1} f(t|m; \psi_0)$. Nesse caso, o somatório vai de y_1 até $\min(n_1, m)$. Para o teste bilateral, $\mathrm{H}_0 : \psi = \psi_0$ contra $\mathrm{H}_1 \neq \psi_0$, o nível descritivo é definido por $\mathrm{P}= 2\min\{\mathrm{P}_I, \mathrm{P}_S\}$.

Em particular, quando $\psi_0 = 1$, está sendo testada a não existência de associação entre o fator e a doença, sendo o teste resultante conhecido como teste exato de Fisher (ver, por exemplo, Everitt, 1977). Nesse caso, o nível descritivo é obtido computando as probabilidades da distribuição hipergeométrica central.

Pode-se também utilizar o modelo condicional (4.5) para a estimação intervalar de ψ. Os respectivos limites de confiança serão baseados em P_I e P_S e denotados por $\hat{\psi}_I$ e $\hat{\psi}_S$, respectivamente. Como ilustração, supor que o interesse é construir um intervalo de confiança de coeficiente $(1 - \alpha)$ para ψ. Os limites $\hat{\psi}_I$ e $\hat{\psi}_S$ ficam então, invertendo a região crítica do teste $\mathrm{H}_0 : \psi = \psi_0$ contra $\mathrm{H}_1 : \psi \neq \psi_0$, determinados pelas equações

$$\frac{\alpha}{2} = \sum_{t \leq y_1} f(t|m; \hat{\psi}_S) \quad \text{e} \quad \frac{\alpha}{2} = \sum_{t \geq y_1} f(t|m; \hat{\psi}_I),$$

que são polinômios de grau elevado em $\hat{\psi}_S$ e $\hat{\psi}_I$ à medida que os tamanhos amostrais crescem, o que praticamente inviabiliza a solução das equações. Nesses casos, uma alternativa é trabalhar com intervalos assintóticos.

Voltando à tabela da seção anterior, supor que o interesse é testar $\mathrm{H}_0 : \psi = 1$ contra $\mathrm{H}_1 : \psi \neq 1$. Tem-se então os níveis descritivos $\mathrm{P}_I = f(0|m; \psi = 1) + f(1|m; \psi = 1) = 15/21$ e $\mathrm{P}_S = f(1|m; \psi = 1) + f(2|m; \psi = 1) = 18/21$ o que leva a P$= 1,0$. Por outro lado, os limites $\hat{\psi}_I$ e $\hat{\psi}_S$ ficam dados por

$$\frac{\alpha}{2} = \sum_{t=0}^{1} f(t|m; \hat{\psi}_S) \quad \text{e} \quad \frac{\alpha}{2} = \sum_{t=1}^{2} f(t|m; \hat{\psi}_I)$$

que é equivalente, supondo $\alpha = 0,20$, a

$$0,10 = f(0|m; \hat{\psi}_S) + f(1|m; \hat{\psi}_S) \quad \text{e} \quad 0,10 = f(1|m; \hat{\psi}_I) + f(2|m; \hat{\psi}_I),$$

que levam às equações

$$0,10 = \frac{4\hat{\psi}_I + 2\hat{\psi}_I^2}{1 + 4\hat{\psi}_I + 2\hat{\psi}_I^2} \quad (\hat{\psi}_I = 0,0274)$$

e

$$0,10 = \frac{1 + 4\hat{\psi}_S}{1 + 4\hat{\psi}_S + 2\hat{\psi}_S^2} \quad (\hat{\psi}_S = 18,25).$$

Testes assintóticos

Para grandes amostras, n_1, n_2, m e $n - m$ grandes, a distribuição condicional (4.5) se aproxima de uma distribuição normal de média $\mathrm{E}_A(\psi)$ e variância $\mathrm{V}_A(\psi)$ (ver Hannan e Harkness, 1963). Esse fato tem sido utilizado para o desenvolvimento de testes assintóticos para testar $\mathrm{H}_0 : \psi = \psi_0$ contra $\mathrm{H}_1 : \psi \neq \psi_0$ ($\mathrm{H}_1 : \psi > \psi_0$ ou $\mathrm{H}_1 : \psi < \psi_0$). No caso de $\mathrm{H}_1 : \psi \neq \psi_0$, utiliza-se a estatística qui-quadrado dada abaixo

$$X^2 = \frac{\{y_1 - \mathrm{E}_A(\psi_0)\}^2}{\mathrm{V}_A(\psi_0)}, \tag{4.8}$$

que sob H_0 segue assintoticamente distribuição qui-quadrado com 1 grau de liberdade. Para $\mathrm{H}_1 : \psi < \psi_0$ e $\mathrm{H}_1 : \psi > \psi_0$, o nível descritivo é dado por

$$\mathrm{P}_I = Pr\left\{ Z \leq \frac{y_1 - \mathrm{E}_A(\psi_0)}{\sqrt{\mathrm{V}_A(\psi_0)}} \right\}$$

e

$$\mathrm{P}_S = Pr\left\{ Z \geq \frac{y_1 - \mathrm{E}_A(\psi_0)}{\sqrt{\mathrm{V}_A(\psi_0)}} \right\},$$

respectivamente, em que Z segue distribuição $\mathrm{N}(0,1)$. Em particular, quando $\psi_0 = 1$, a estatística qui-quadrado (4.8) fica reduzida à forma conhecida

$$X^2 = \frac{\left\{ y_1 - \frac{mn_1}{n} \right\}^2}{n_1 n_2 m(n-m)/n^3}. \tag{4.9}$$

Um intervalo assintótico de confiança para ψ pode ser obtido utilizando a distribuição assintótica de $\log(\tilde{\psi})$. Os limites desse intervalo são dados por

$$\tilde{\psi}_I = \exp[\log(\tilde{\psi}) - z_{(1-\alpha/2)} \sqrt{\hat{\mathrm{V}}\mathrm{ar}_\mathrm{A}\{\log(\tilde{\psi})\}}]$$

e

$$\tilde{\psi}_S = \exp[\log(\tilde{\psi}) + z_{(1-\alpha/2)} \sqrt{\hat{\mathrm{V}}\mathrm{ar}_\mathrm{A}\{\log(\tilde{\psi})\}}],$$

em que $z_{(1-\alpha/2)}$ denota o quantil $(1-\alpha/2)$ da distribuição normal padrão e

$$\hat{\mathrm{V}}\mathrm{ar}_\mathrm{A}\{\log(\tilde{\psi})\} = \left[\frac{1}{y_1} + \frac{1}{n_1 - y_1} + \frac{1}{y_2} + \frac{1}{n_2 - y_2} \right].$$

Esses limites podem ser expressos em uma outra forma, levando-se em conta a estatística qui-quadrado para testar $\mathrm{H}_0 : \psi = 1$ contra $\mathrm{H}_1 : \psi \neq 1$. Essa estatística é denotada por

$$X^2 = \frac{\{\log(\tilde{\psi})\}^2}{\hat{\mathrm{V}}\mathrm{ar}_\mathrm{A}\{\log(\tilde{\psi})\}}, \tag{4.10}$$

que segue, para grandes amostras, distribuição qui-quadrado com 1 grau de liberdade. Assim, os limites ficam reexpressos nas formas

$$\tilde{\psi}_I = \tilde{\psi}^{(1-z_{(1-\alpha/2)}/X)}$$

e

$$\tilde{\psi}_S = \tilde{\psi}^{(1+z_{(1-\alpha/2)}/X)}.$$

Alguns autores (ver Breslow e Day, 1980, p. 135) têm constatado que para $n_1 = n_2$ a probabilidade de cobertura do intervalo $(\tilde{\psi}_I, \tilde{\psi}_S)$ é em geral menor do que o valor nominal utilizado. Por outro lado, quando n_1 e n_2 são muito diferentes, essa probabilidade de cobertura é superestimada. Uma sugestão, nesses casos, é utilizar o valor de X obtido do teste condicional (4.9) em vez do valor obtido do teste não condicional (4.10).

4.3 Métodos clássicos: k tabelas 2×2

Muitas vezes tem-se interesse em controlar a associação entre dois fatores binários através de um terceiro fator, comumente chamado de fator de confundimento. O principal objetivo com esse tipo de estratificação é eliminar ou pelo menos reduzir a influência desses fatores na associação de interesse. Uma maneira mais eficiente de controlar fatores de confundimento é através da regressão logística, que será discutida na Seção 4.6. Nesta seção, será considerado apenas um fator de confundimento com k níveis, que são amostrados n_i indivíduos no i-ésimo estrato (n_{1i} casos e n_{2i} controles) e que os mesmos são classificados conforme a tabela 2×2 abaixo.

	Fator		
Doença	A	B	Total
D	y_{1i}	$n_{1i} - y_{1i}$	n_{1i}
$\bar{D}$	y_{2i}	$n_{2i} - y_{2i}$	n_{2i}

Seguindo a mesma notação das seções anteriores tem-se que as estimativas não condicional e condicional de ψ_i são, respectivamente, tais que

$$\tilde{\psi}_i = \frac{y_{1i}(n_{2i} - y_{2i})}{y_{2i}(n_{1i} - y_{1i})} \quad \text{e} \quad y_{1i} - \frac{P_{1i}(\hat{\psi}_i)}{P_{0i}(\hat{\psi}_i)} = 0.$$

As propriedades assintóticas de $\tilde{\psi}_i$ e $\hat{\psi}_i$ são as mesmas de $\tilde{\psi}$ e $\hat{\psi}$ da Seção 4.2, bem como as formas dos testes de hipóteses e da estimação intervalar.

4.3.1 Estimação da razão de chances comum

Um teste de interesse quando há k tabelas de contingência 2×2 é verificar a ausência de interação entre os estratos, isto é, verificar se a associação entre o fator e a doença não muda de um estrato para o outro. Isso é equivalente a verificar se as razões de chances são homogêneas, ou seja, testar as hipóteses

$$\begin{aligned} &\mathrm{H}_0: \quad \psi_1 = \cdots = \psi_k \\ &\mathrm{H}_1: \quad \text{pelo menos dois valores diferentes.} \end{aligned}$$

Há várias propostas de estimativas para a razão de chances comum. As estimativas de máxima verossimilhança não condicional e condicional serão denotadas por $\tilde{\psi}$ e $\hat{\psi}$, respectivamente. A primeira estimativa pode ser obtida facilmente através do ajuste de uma regressão logística, enquanto que a segunda é mais complexa do ponto de vista computacional e será omitida.

Duas estimativas não iterativas foram propostas por Mantel e Haenszel (1959) e Wolf (1955), as quais serão denotadas por $\hat{\psi}_{MH}$ e $\hat{\psi}_W$, respectivamente. A estimativa de Mantel-Haenszel é definida por

$$\hat{\psi}_{MH} = \frac{\sum_{i=1}^{k} y_{1i}(n_{2i} - y_{2i})/n_i}{\sum_{i=1}^{k} y_{2i}(n_{1i} - y_{1i})/n_i},$$

e pode também ser expressa como uma média ponderada de estimativas não

condicionais

$$\hat{\psi}_{MH} = \frac{\sum_{i=1}^{k} v_i \tilde{\psi}_i}{\sum_{i=1}^{k} v_i},$$

em que $v_i = y_{2i}(n_{1i} - y_{1i})/n_i$. O estimador de Mantel-Haenszel é consistente e assintoticamente normal com variância assintótica dada por

$$\mathrm{Var_A}(\hat{\psi}_{MH}) = \psi^2 \sum_{i=1}^{k} a_i \omega_i^{-1} / (\sum_{i=1}^{k} a_i)^2,$$

em que $\omega_i = \{n_{1i}\pi_{1i}(1 - \pi_{1i})\}^{-1} + \{n_{2i}\pi_{2i}(1 - \pi_{2i})\}^{-1}$ e $a_i = n_{1i}n_{2i}(1 - \pi_{1i})\pi_{2i}/n_i$. A estimativa de Wolf é dada por

$$\hat{\psi}_W = \exp\left\{ \frac{\sum_{i=1}^{k} \tilde{\omega}_i^{-1} \log(\tilde{\psi}_i)}{\sum_{i=1}^{k} \tilde{\omega}_i^{-1}} \right\},$$

em que $\tilde{\omega}_i = \{1/y_{1i} + 1/(n_{1i} - y_{1i}) + 1/y_{2i} + 1/(n_{2i} - y_{2i})\}$. Esse estimador é também consistente e assintoticamente normal com variância dada por

$$\mathrm{Var_A}(\hat{\psi}_W) = \psi^2 \omega^{-1},$$

em que $\omega = \omega_1^{-1} + \cdots + \omega_k^{-1}$. Como $\log(\hat{\psi}_W)$ converge mais rapidamente para a distribuição normal do que $\hat{\psi}_W$, uma estimativa intervalar de coeficiente de coefiança $(1 - \alpha)$ para ψ comum fica dada por

$$\tilde{\psi}_I = \exp[\log(\hat{\psi}_W) - z_{(1-\alpha/2)} \sqrt{\hat{\mathrm{V}}\mathrm{ar_A}\{\log(\hat{\psi}_W)\}}]$$

e

$$\tilde{\psi}_S = \exp[\log(\hat{\psi}_W) + z_{(1-\alpha/2)} \sqrt{\hat{\mathrm{V}}\mathrm{ar_A}\{\log(\hat{\psi}_W)\}}],$$

em que $z_{(1-\alpha/2)}$ denota o quantil $(1 - \alpha/2)$ da distribuição normal padrão e $\hat{\mathrm{V}}\mathrm{ar_A}\{\log(\hat{\psi}_W)\} = 1/\sum_{i=1}^{k} \tilde{\omega}_i^{-1}$. Similarmente pode-se encontrar estimativas assintóticas intervalares para ψ comum utilizando o estimador de Mantel-Haenszel.

4.3.2 Testes de homogeneidade

Supor que o interesse é testar as hipóteses H_0 e H_1 definidas na seção anterior. A estatística da razão de verossimilhanças que assume o produto de $2k$ binomiais independentes é a mais utilizada nesse caso. Do ponto de vista de análise preliminar dos dados, duas estatísticas têm sido sugeridas. A primeira delas (vide Hosmer et al., 2013), é definida abaixo

$$X^2_{HL} = \sum_{i=1}^{k} \tilde{\omega}_i^{-1} \{\log(\tilde{\psi}_i) - \log(\hat{\psi}_W)\}^2,$$

que segue, sob H_0 e assintoticamente (para n_{1i} e n_{2i} grandes, $\forall i$), distribuição qui-quadrado com $k-1$ graus de liberdade. A outra estatística, definida em Breslow e Day (1980, p. 42), é baseada no modelo condicional, sendo expressa na forma

$$X^2_{BD} = \sum_{i=1}^{k} \frac{\{y_{1i} - \mathrm{E}_{A_i}(\hat{\psi}_{MH})\}^2}{\mathrm{V}_{A_i}(\hat{\psi}_{MH})},$$

que também segue, sob H_0 e para grandes amostras, distribuição qui-quadrado com $k-1$ graus de liberdade. A estatística do teste é avaliada na estimativa não iterativa de Mantel-Haenszel ao invés da estimativa condicional $\hat{\psi}$.

Quando a hipótese nula não é rejeitada, um teste imediato é verificar a não existência de associação entre o fator e a doença, mantendo apenas o efeito da estratificação. Esse teste, conhecido como teste de Mantel-Haenszel (1959), utiliza a seguinte estatística:

$$X^2_{MH} = \frac{\{\sum_{i=1}^{k} y_{1i} - \sum_{i=1}^{k} \mathrm{E}_{A_i}(1)\}^2}{\sum_{i=1}^{k} \mathrm{V}_{A_i}(1)},$$

que, sob $H_0 : \psi = 1$, segue para grandes amostras (n_i grande $\forall i$ ou para k grande) distribuição qui-quadrado com 1 grau de liberdade. Similarmente ao caso de uma única tabela 2×2, um intervalo assintótico de confiança para

ψ com coeficiente de confiança $(1-\alpha)$ fica dado por

$$(\hat{\psi}_I, \hat{\psi}_S) = \hat{\psi}_{MH}^{(1\pm z_{(1-\alpha/2)}/X_{MH})},$$

em que $X_{MH} = \sqrt{X^2_{MH}}$. Para melhorar a aproximação para a distribuição normal, é usal aplicar correção de continuidade no teste de Mantel-Haenszel.

4.4 Métodos clássicos: tabelas $2 \times k$

A dicotomização de um fator com mais de 2 níveis, a fim de deixar mais simples o estudo da associação entre esse fator e uma determinada doença, pode omitir informações relevantes acerca da associação de cada um dos níveis agrupados e a doença em estudo. Assim, sempre que possível, deve-se manter para as análises o maior número possível de níveis do fator. Uma tabela resultante, nesse caso, é dada abaixo.

	Fator				
Doença	Nível 1	Nível 2	$\cdots$	Nível k	Total
D	y_{11}	y_{12}	$\cdots$	$n_1 - \sum_{i=1}^{k-1} y_{1i}$	n_1
$\bar{D}$	y_{21}	y_{22}	$\cdots$	$n_2 - \sum_{i=1}^{k-1} y_{2i}$	n_2

Analogamente ao caso de uma única tabela 2×2, assume-se que são amostrados n_1 elementos do estrato D e n_2 elementos do estrato $\bar{D}$ e que $(Y_{i1}, \ldots, Y_{ik})^\top$ segue distribuição multinomial de parâmetros $(\pi_{i1}, \ldots, \pi_{ik})^\top$, com $\pi_{ik} = 1 - \sum_{j=1}^{k-1} \pi_{ij}$, $i = 1, 2$. Comumente, para analisar as associações entre os níveis do fator e a doença, define-se um nível do fator como referência, que formará com os demais as razões de chances. Escolhendo o nível 1 como referência, as razões de chances ficam dadas por

$$\psi_1 = 1 \ \text{ e } \ \psi_j = \frac{\pi_{1j}\pi_{21}}{\pi_{2j}\pi_{11}}, \ j = 2, \ldots, k,$$

em que ψ_j é a razão de chances entre o nível j e o nível 1 do fator. As análises inferenciais através do uso do modelo multinomial são tratadas em textos correntes de análise de dados categorizados (ver, por exemplo, Agresti, 1990). Aqui, o estudo será restrito ao modelo condicional, que é obtido após o condicionamento de $(Y_{i1}, \ldots, Y_{ik})^\top$, $i = 1, 2$, nas estatísticas suficientes minimais $Y_{1j} + Y_{2j} = m_j$, $j = 1, \cdots, k$. O modelo resultante é caracterizado pela distribuição hipergeométrica multivariada não central que depende apenas dos parâmetros de interesse $\psi_1, \ldots, \psi_k$ (ver McCullagh e Nelder, 1989, p. 261). Em particular, a hipótese de ausência de associação completa entre os níveis do fator e a doença é definida por $\mathrm{H}_0 : \psi_j = 1, \forall j$, que será avaliada através da distribuição hipergeométrica central k-dimensional, cuja função de probabilidade é o produto de k distribuições hipergeométricas centrais

$$f(\mathbf{y}_1|\mathbf{m}; \boldsymbol{\psi} = \mathbf{1}) = \prod_{j=1}^{k} \frac{\binom{n_{1j}}{y_{1j}}\binom{n_{2j}}{m_j - y_{1j}}}{\binom{n_{1j}+n_{2j}}{m_j}}, \tag{4.11}$$

em que $\mathbf{y}_1 = (y_{11}, \ldots, y_{1k})^\top$, $\mathbf{m} = (m_1, \ldots, m_k)^\top$ e $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_k)^\top$. A média, variância e covariância correspondentes à distribuição (4.11) são, respectivamente, dadas por

$$\mathrm{E}_j(1) = \mathrm{E}(Y_{1j}|m_j; \psi = 1) = \frac{m_j n_1}{n},$$

$$\mathrm{V}_j(1) = \mathrm{Var}(Y_{1j}|m_j; \psi = 1) = \frac{n_1 n_2 (n - m_j) m_j}{n^2(n-1)}$$

e

$$\mathrm{C}_{j\ell} = \mathrm{Cov}(Y_{1j}, Y_{1\ell}|m_j, m_\ell; \psi = 1) = -\frac{m_j m_\ell n_1 n_2}{n^2(n-1)}, \ j \neq \ell,$$

em que $n = n_1 + n_2$. Um teste estatístico para H_0, que tem sido largamente utilizado para testar a homogeneidade de k proporções (Armitage, 1971), é

dado por

$$\begin{aligned} X_A^2 &= \frac{(n-1)}{n}\sum_{j=1}^{k}\{y_{1j}-\mathrm{E}_j(1)\}^2\left\{\frac{1}{\mathrm{E}_j(1)}+\frac{1}{m_j-\mathrm{E}_j(1)}\right\} \\ &= (n-1)\left\{\frac{1}{n_1}+\frac{1}{n_2}\right\}\sum_{j=1}^{k}\frac{\{y_{1j}-\mathrm{E}_j(1)\}^2}{m_j}, \end{aligned} \tag{4.12}$$

que segue, sob H_0 e para valores grandes de n_1, n_2 e $m_j, \forall j$, distribuição qui-quadrado com $k-1$ graus de liberdade. Entretanto, quando os níveis do fator são quantitativos ou qualitativos ordinais, pode ser mais informativo o uso de um teste para a tendência do risco da doença com o aumento dos níveis do fator. Como ilustração, supor que há k doses x_j, $j = 1, \ldots, k$, associadas aos k níveis do fator. Um teste apropriado é considerar a regressão dos desvios $\{y_{1j} - \mathrm{E}_j(1)\}$ sobre x_j (Armitage, 1955; Mantel, 1963). A estatística correspondente fica dada por

$$X_{HOM}^2 = \frac{n^2(n-1)[\sum_{j=1}^{k} x_j\{y_{1j}-\mathrm{E}_j(1)\}]^2}{n_1 n_2\{n\sum_{j=1}^{k} x_j^2 m_j - (\sum_{j=1}^{k} x_j m_j)^2\}}, \tag{4.13}$$

que segue, para grandes amostras e sob H_0, distribuição qui-quadrado com $k-1$ graus de liberdade.

Uma outra maneira de analisar a associação entre o fator e a doença é através da amostragem nos k níveis do fator de interesse. Nesse caso, a distribuição resultante é um produto de k binomiais independentes e a hipótese de ausência de associação entre o fator e a doença pode ser avaliada através do ajuste de uma regressão logística, que será discutida na Seção 4.6. Por outro lado, se também forem fixados os totais n_1 e n_2, a distribuição condicional resultante é uma hipergeométrica não central k-dimensional que sob H_0 fica reduzida a (4.11). Logo, as estatísticas dadas em (4.12) e (4.13) podem ser aplicadas, pelo menos numa análise preliminar dos dados, para avaliar a ausência de associação total entre o fator e a doença.

Generalizações de (4.12) e (4.13) para o caso de h estratos são dadas em Breslow e Day (1980, pgs. 148-149).

4.5 Aplicações

4.5.1 Associação entre fungicida e desenvolvimento de tumor

Como ilustração de análise de k tabelas do tipo 2×2, será analisado o conjunto de dados apresentado em Innes et al. (1969), referente a um estudo para avaliar o possível efeito cancerígeno do fungicida Avadex. No estudo, 403 camundongos são observados. Desses, 65 receberam o fungicida e foram acompanhados durante 85 semanas, verificando o desenvolvimento ou não de tumor cancerígeno neste período. Os demais animais não receberam o fungicida (grupo controle) e também foram acompanhados pelo mesmo período, verificando a ocorrência ou não de tumor no período. Dois fatores potenciais de confundimento, sexo e raça, foram considerados nas análises. Os dados do experimento são resumidos na Tabela 4.1.

Em virtude dos valores relativamente altos das marginais das quatro tabelas 2×2 formadas pela combinação dos fatores sexo e raça, será aplicada uma análise através do modelo não condicional. Tem-se então, na primeira coluna da Tabela 4.2, as estimativas pontuais das razões de chances de tumor maligno entre o grupo tratado e o grupo controle. Na segunda coluna tem-se as estimativas intervalares assintóticas de 95% para ψ. Nota-se que, embora todas as estimativas sinalizem para uma associação positiva, apenas o primeiro intervalo de confiança não cobre o valor $\psi = 1$, evidenciando associação apenas no primeiro estrato, ao nível de significância de 5%.

Tabela 4.1
Classificação dos camundongos conforme a raça (R1 ou R2), sexo, grupo e ocorrência ou não de tumor cancerígeno.

Estrato	Grupo	Com tumor	Sem tumor	Total
	Tratado	4	12	16
R1-Macho	Controle	5	74	79
	Total	9	86	95
	Tratado	2	14	16
R2-Macho	Controle	3	84	87
	Total	5	98	103
	Tratado	4	14	18
R1-Fêmea	Controle	10	80	90
	Total	14	94	108
	Tratado	1	14	15
R2-Fêmea	Controle	3	79	82
	Total	4	93	97

Tabela 4.2
Estimativas das razões de chances de tumor cancerígeno nos estratos de camundongos.

Estrato	Estimativa $\tilde{\psi}$	Intervalo assintótico
R1-Macho	4,93	[1,163 ; 21,094]
R2-Macho	4,00	[0,612 ; 26,102]
R1-Fêmea	2,29	[0,629 ; 8,306]
R2-Fêmea	1,88	[0,183 ; 19,395]

Para simplificar os cálculos, será considerado o estimador de Wolf a fim de obter a estimativa de ψ comum aos estratos. Tem-se então as seguintes estimativas:

$\log(\tilde{\psi}_i)$	$\tilde{\omega}_i$
1,600	0,5465
1,386	0,9160
0,827	0,4335
0,632	1,4167

Segue portanto que $\sum_{i=1}^{4} \tilde{\omega}_i^{-1} \log(\tilde{\psi}_i) = 6,7947$ e $\sum_{i=1}^{4} \tilde{\omega}_i^{-1} = 5,9342$. Assim, obtém-se as estimativas

$$\hat{\psi}_W = \exp\left(\frac{6,7947}{5,9342}\right) = 3,142 \;\text{ e }\; \hat{\mathrm{V}}\mathrm{ar}_{\mathrm{A}}\{\log(\hat{\psi}_{\mathrm{W}})\} = 1/\sum_{\mathrm{i}=1}^{4} \tilde{\omega}_{\mathrm{i}}^{-1} = 1/5,9342.$$

Consequentemente, tem-se que $\log(\hat{\psi}_W) = \log(3,142) = 1,145$ e $X^2_{HL} = (1,6-1,145)^2/0,5465 + (1,386-1,145)^2/0,916 + (0,827-1,145)^2/0,4335 + (0,632-1,145)^2/1,4167 = 0,861$, cujo nivel descritivo para uma distribuição qui-quadrado com 3 graus de liberdade é dado por $P = 0,84$, não rejeitando-se portanto a hipótese de ψ comum.

A estimativa intervalar de 95% para ψ comum fica dada por

$$\begin{aligned}
[\hat{\psi}_I, \hat{\psi}_S] &= \exp[\log(3,142) \pm 1,96\sqrt{1/5,9342}] \\
&= \exp[1,145 \pm 0,8046] \\
&= [1,4055; 7,0259].
\end{aligned}$$

Será aplicado a seguir o teste de Mantel-Haenszel para testar $\mathrm{H}_0 : \psi = 1$ contra $\mathrm{H}_1 : \psi \neq 1$. Ou seja, verificar se há asusência de associação em cada tabela 2×2. A estatística do teste com correção de continuidade fica expressa na seguinte forma:

$$\begin{aligned}
X^2_{MH} &= \frac{\{|\sum_{i=1}^{4} y_{1i} - \sum_{i=1}^{4} \mathrm{E}_{A_i}(1)| - 0,5\}^2}{\sum_{i=1}^{4} \mathrm{V}_{A_i}(1)} \\
&= \frac{(|11 - 5,2444| - 0,5)^2}{3,9983} = 6,9083.
\end{aligned}$$

Comparando com os quantis da distribuição qui-quadrado com 1 grau de liberdade obtém-se o nível descritivo $P = 0,0086$, rejeitando-se a hipótese nula. Esse resultado vai ao encontro da estimativa intervalar de ψ comum usando o estimador de Wolf.

4.5.2 Efeito de extrato vegetal

Considere agora parte dos dados de um experimento (ver Paula et al.,1988) conduzido para avaliar o efeito de diversos extratos vegetais na mortalidade de embriões de *Biomphalaria Glabrata* (hospedeiro da equistossomose). Para o extrato vegetal aquoso frio de folhas de P. Hyrsiflora foi considerado um total de $k = 7$ grupos sendo que os n_i embriões do i-ésimo grupo foram submetidos a uma dose x_i (ppm) do extrato vegetal, observando-se após o 20^{o} dia o número de embriões mortos. Os dados são resumidos na Tabela 4.3. Para aplicar o teste de tendência dado em (4.13), deve-se considerar que $n = 50 + \cdots + 50 = 350$, $n_1 = y_1 + \cdots + y_7 = 178$, $n_2 = n - n_1 = 172$ e $m_i = 50$, $\forall i$. Assim, obtem-se $\mathrm{E}_i(1) = 25,43$ para $i = 1, \ldots, 7$. A estatística do teste forneceu o valor $X^2_{HOM} = 131,82$, que é altamente significativo quando comparado aos quantis da distribuição qui-quadrado com 6 graus de liberdade, indicando uma forte tendência crescente para a proporção de mortes com o aumento da dose.

Tabela 4.3
Distribuição dos embriões segundo os níveis de exposição do estrato vegetal aquoso.

x_i	0	15	20	25	30	35	40
m_i	50	50	50	50	50	50	50
y_i	4	5	14	29	38	41	47

4.6 Regressão logística linear

4.6.1 Introdução

A regressão logística tem se constituído num dos principais métodos de modelagem estatística de dados. Mesmo quando a resposta de interesse não é originalmente do tipo binário, alguns pesquisadores têm dicotomizado a resposta de modo que a probabilidade de sucesso possa ser ajustada através da regressão logística. Isso ocorre, por exemplo, em análise de sobrevivência discreta em que a resposta de interesse é o tempo de sobrevivência, no entanto, em algumas pesquisas, a função de risco tem sido ajustada por modelos logísticos. Tudo isso se deve, principalmente, pela facilidade de interpretação dos parâmetros de um modelo logístico e também pela possibilidade do uso desse tipo de metodologia em análise discriminante com a construção, por exemplo, de curvas ROC.

Embora a regressão logística seja conhecida desde os anos 1950, foi através de Cox (1970) (ver também Cox e Snell, 1989) que a regressão logística ficou popular entre os usuários de Estatística. Nesta seção serão apresentados alguns resultados relacionados com o modelo logístico linear que completam os procedimentos apresentados no Capítulo 1, em que esse modelo foi descrito como um caso particular de modelos lineares generalizados.

4.6.2 Regressão logística simples

Conside inicialmente o modelo logístico linear simples em que $\pi(x)$, a probabilidade de sucesso dado o valor x de uma variável explicativa qualquer, é definida tal que

$$\log\left\{\frac{\pi(x)}{1-\pi(x)}\right\} = \alpha + \beta x, \tag{4.14}$$

em que α e β são parâmetros desconhecidos. Esse modelo poderia, por exemplo, ser aplicado para analisar a associação entre uma determinada doença e a ocorrência ou não de um fator particular. Seriam então amostrados, independentemente, n_1 indivíduos com presença do fator (x=1) e n_2 indivíduos com ausência do fator (x=0) e $\pi(x)$ seria a probabilidade de desenvolvimento da doença após um certo período fixo. Dessa forma, a chance de desenvolvimento da doença para um indivíduo com presença do fator fica dada por

$$\frac{\pi(1)}{1-\pi(1)} = e^{\alpha+\beta},$$

enquanto que a chance de desenvolvimento da doença para um indivíduo com ausência do fator é simplesmente

$$\frac{\pi(0)}{1-\pi(0)} = e^{\alpha}.$$

Logo, a razão de chances fica dada por

$$\psi = \frac{\pi(1)\{1-\pi(0)\}}{\pi(0)\{1-\pi(1)\}} = e^{\beta},$$

dependendo apenas do parâmetro β. Mesmo que a amostragem seja retrospectiva, isto é, são amostrados n_1 indivíduos doentes e n_2 indivíduos não doentes, o resultado acima continua valendo. Essa é uma das grandes vantagens da regressão logística, a possibilidade de interpretação direta dos coeficientes como medidas de associação. Esse tipo de interpretação pode ser estendido para qualquer problema prático.

Supor agora que tem-se dois estratos representados por x_1 ($x_1 = 0$ estrato 1, $x_1 = 1$ estrato 2) e que são amostrados do estrato 1 n_{11} indivíduos com presença do fator e n_{21} indivíduos com ausência do fator e n_{12} e n_{22}, respectivamente, do estrato 2. A probabilidade de desenvolvimento da doença será denotada por $\pi(x_1, x_2)$, com x_2 (x_2=1 presença do fator, $x_2 = 0$ ausência do fator). Tem-se aqui quatro parâmetros a serem estimados, $\pi(0,0), \pi(0,1), \pi(1,0)$

e $\pi(1,1)$. Logo, qualquer reparametrização deverá ter no máximo quatro parâmetros (modelo saturado).

Considere então a seguinte reparametrização:

$$\log\left\{\frac{\pi(x_1,x_2)}{1-\pi(x_1,x_2)}\right\} = \alpha + \gamma x_1 + \beta x_2 + \delta x_1 x_2,$$

em que γ representa o efeito do estrato, β o efeito do fator e δ a interação entre estrato e fator. Para entender melhor essa reparametrização, serão calculadas as razões de chances em cada estrato

$$\psi_1 = \frac{\pi(0,1)\{1-\pi(0,0)\}}{\pi(0,0)\{1-\pi(0,1)\}} = e^{\beta}$$

e

$$\psi_2 = \frac{\pi(1,1)\{1-\pi(1,0)\}}{\pi(1,0)\{1-\pi(1,1)\}} = e^{\beta+\delta}.$$

Assim, a hipótese de homogeneidade das razões de chances ($\mathrm{H}_0 : \psi_1 = \psi_2$) é equivalente à hipótese de não interação ($\mathrm{H}_0 : \delta = 0$). Portanto, a ausência de interação entre fator e estrato significa que a associação entre o fator e a doença não muda de um estrato para o outro. Contudo, pode haver efeito de estrato. Como ilustração nesse caso, supor que não rejeita-se a hipótese $\mathrm{H}_0 : \delta = 0$. Assim, o logaritmo da chance de desenvolvimento da doença fica dado por

$$\log\left\{\frac{\pi(x_1,x_2)}{1-\pi(x_1,x_2)}\right\} = \alpha + \gamma x_1 + \beta x_2,$$

ou seja, é o mesmo nos dois estratos a menos da quantidade γ. Isso quer dizer que mesmo não havendo interação entre os dois estratos (razão de chances constante), as probabilidades de desenvolvimento da doença podem estar em patamares diferentes. Num estrato essas probabilidades são maiores do que no outro estrato. Essas interpretações podem ser generalizadas para três ou mais tabelas.

Aplicação

Como ilustração, considere novamente o exemplo descrito na Seção 4.5.1, supondo que agora temos apenas os estratos macho e fêmea. Os dados são resumidos na Tabela 4.4 e no arquivo **camundongos.txt**.

Tabela 4.4
Classificação de camundongos segundo sexo, grupo e ocorrência de tumor.

	Macho		Fêmea	
Tumor	Tratado	Controle	Tratado	Controle
Sim	6	8	5	13
Não	26	158	28	159
Total	32	166	33	172

Denote por $\pi(x_1, x_2)$ a probabilidade de desenvolvimento de tumor dados x_1 (x_1=1 macho, x_1=0 fêmea) e x_2 (x_2=1 tratado, x_2=0 controle). Para testar a hipótese de ausência de interação ($H_0 : \delta = 0$) compara-se o desvio do modelo sem interação $D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) = 0,832$ com os quantis da distribuição qui-quadrado com 1 grau de liberdade (tem-se que o desvio do modelo saturado é zero). O nível descritivo obtido é dado por P= $0,362$, indicando pela não rejeição da hipótese de homogeneidade das razões de chances. Assim, ajusta-se o modelo sem interação. As estimativas resultantes são apresentadas na Tabela 4.5.

Tabela 4.5
Estimativas dos parâmetros do modelo logístico ajustado aos dados sobre ocorrência de tumor em camundongos.

Efeito	Estimativa	E/E.Padrão
Constante	-2,602	-9,32
Estrato	-0,241	-0,64
Tratamento	1,125	2,81

Os níveis descritivos dos testes para $H_0 : \beta = 0$ e $H_0 : \gamma = 0$ são, respectivamente, dados por P= $0,005$ e P= $0,520$, indicando fortemente pela presença de associação entre a exposição ao fungicida e o desenvolvimento de tumor e que as probabilidades de desenvolvimento de tumor não são diferentes entre os dois estratos.

Tem-se que $\hat{\psi} = e^{\hat{\beta}}$, logo um intervalo assintótico de confiança para ψ com coeficiente $(1 - \alpha)$, terá os limites

$$(\hat{\psi}_I, \hat{\psi}_S) = \exp\{\hat{\beta} \pm z_{(1-\alpha/2)} \sqrt{\hat{\mathrm{V}}\mathrm{ar}(\hat{\beta})}\}.$$

Para o exemplo acima e assumindo um intervalo de 95%, esses limites ficam dados por $[1,403; 6,759]$.

O valor observado da variável explicativa no modelo logístico dado em (4.14) pode representar o valor de alguma variável quantitativa qualquer como, por exemplo, a dose ou a log-dose de uma determinada droga. Nesse caso, faz sentido calcular a chance de um indivíduo que recebeu a dose x^*, ser curado, em relação a um outro indivíduo que recebeu a dose x. A razão de chances de cura, entre os dois níveis, fica dada por

$$\psi_{(x^*-x)} = \frac{\pi(x^*)\{1 - \pi(x)\}}{\pi(x)\{1 - \pi(x^*)\}} = \exp\{\beta(x^* - x)\}.$$

Portanto, $\log\{\psi_{(x^*-x)}\}$ é proporcional à diferença entre as duas doses. Se $\beta > 0$, tem-se que a chance de cura aumenta com o aumento da dose e se $\beta < 0$ ocorre o contrário. Essa interpretação pode ser estendida para qualquer variável explicativa quantitativa.

4.6.3 Regressão logística múltipla

Considere agora o modelo geral de regressão logística

$$\log\left\{\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right\} = \beta_1 + \beta_2 x_2 + \cdots + \beta_p x_p,$$

em que $\mathbf{x} = (1, x_2, \ldots, x_p)^\top$ contém os valores observados de variáveis explicativas. Como visto na Seção 1.6.1, o processo iterativo para obtenção de $\hat{\boldsymbol{\beta}}$ pode ser expresso como um processo iterativo de mínimos quadrados reponderados

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top \mathbf{V}^{(m)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{(m)} \mathbf{z}^{(m)},$$

em que $\mathbf{V} = \text{diag}\{\pi_1(1-\pi_1), \ldots, \pi_n(1-\pi_n)\}$, $\mathbf{z} = (z_1, \ldots, z_n)^\top$ é a variável dependente modificada, $z_i = \eta_i + (y_i - \pi_i)/\pi_i(1-\pi_i)$, $m = 0, 1, \ldots$ e $i = 1, \ldots, n$. Para dados agrupados (k grupos), n é substituído por k, $\mathbf{V} = \text{diag}\{n_1\pi_1(1-\pi_1), \ldots, n_k\pi_k(1-\pi_k)\}$ e $z_i = \eta_i + (y_i - n_i\pi_i)/\{n_i\pi_i(1-\pi_i)\}$. Assintoticamente, $n \to \infty$ no primeiro caso e para $\frac{n_i}{n} \to a_i > 0$ no segundo caso, $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim \text{N}_p(\mathbf{0}, (\mathbf{X}^\top \mathbf{V} \mathbf{X})^{-1})$.

Uma interpretação interessante pode ser dada para as razões de chances quando tem-se $(q-1)(q \leq p)$ das $(p-1)$ variáveis explicativas do tipo binário. Como ilustração, supor $q = 4$ e que x_2 ($x_2 = 1$ presença, $x_2 = 0$ ausência) e x_3 ($x_3 = 1$ presença, $x_3 = 0$ ausência) representam dois fatores. Supor ainda que $x_4 = x_2 x_3$ representa a interação entre os dois fatores. O modelo fica então dado por

$$\log\left\{\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\right\} = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \sum_{j=5}^{p} x_j \beta_j.$$

Denote por ψ_{ij} a razão de chances entre um indivíduo na condição ($x_2 = i, x_3 = j$) em relação a um indivíduo na condição ($x_2 = 0, x_3 = 0$), para $i, j = 0, 1$, supondo que os dois indivíduos têm os mesmos valores observados para as demais $(p-4)$ variáveis explicativas. Assim, pode-se mostrar facilmente que

$$\psi_{10} = \exp(\beta_2), \quad \psi_{01} = \exp(\beta_3) \quad \text{e} \quad \psi_{11} = \exp(\beta_2 + \beta_3 + \beta_4).$$

Portanto, testar a hipótese $\text{H}_0 : \beta_4 = 0$ (ausência de interação) é equivalente a testar a hipótese de efeito multiplicativo $\text{H}_0 : \psi_{11} = \psi_{10}\psi_{01}$. Em particular,

se x_3 representa dois estratos ($x_3 = 0$, estrato 1; $x_3 = 1$, estrato 2), a razão de chances no primeiro estrato entre presença e ausência do fator fica dada por $\psi_{10} = \exp(\beta_2)$, enquanto que no segundo estrato essa razão de chances vale $\psi_{11}/\psi_{01} = \exp(\beta_2 + \beta_4)$. Logo, testar $\mathrm{H}_0 : \beta_4 = 0$ equivale também a testar a hipótese de homogeneidade das razões de chances nos dois estratos.

4.6.4 Bandas de confiança

Como foi visto na Seção 2.7 uma banda assintótica de confiança de coeficiente $1 - \alpha$ pode ser construída para $\pi(\mathbf{z}), \forall \mathbf{z} \in I\!R^p$ (ver também Piegorsch e Casella, 1988). Assintoticamente $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim \mathrm{N}_p(\mathbf{0}, (\mathbf{X}^\top \mathbf{V}\mathbf{X})^{-1})$. Logo, uma banda assintótica de confiança de coeficiente $1 - \alpha$ para o preditor linear $\mathbf{z}^\top \boldsymbol{\beta}, \forall \mathbf{z} \in I\!R^p$, fica dada por

$$\mathbf{z}^\top \hat{\boldsymbol{\beta}} \pm \sqrt{c_\alpha} \{\mathbf{z}^\top (\mathbf{X}^\top \mathbf{V}\mathbf{X})^{-1} \mathbf{z}\}^{\frac{1}{2}}, \quad \forall \mathbf{z} \in I\!R^p,$$

em que c_α é tal que $Pr\{\chi^2_p \leq c_\alpha\} = 1 - \alpha$. Aplicando a transformação logito pode-se, equivalentemente, encontrar uma banda de confiança de coeficiente $1 - \alpha$ para $\pi(\mathbf{z})$, dada por

$$\frac{\exp[\mathbf{z}^\top \hat{\boldsymbol{\beta}} \pm \sqrt{c_\alpha} \{\mathbf{z}^\top (\mathbf{X}^\top \mathbf{V}\mathbf{X})^{-1} \mathbf{z}\}^{\frac{1}{2}}]}{1 + \exp[\mathbf{z}^\top \hat{\boldsymbol{\beta}} \pm \sqrt{c_\alpha} \{\mathbf{z}^\top (\mathbf{X}^\top \mathbf{V}\mathbf{X})^{-1} \mathbf{z}\}^{\frac{1}{2}}]}, \quad \forall \mathbf{z} \in I\!R^p.$$

É importante observar que $\mathbf{z}$ é um vetor $p \times 1$ que varia livremente no $I\!R^p$, enquanto $\mathbf{X}$ é uma matriz fixa com os valores das variáveis explicativas.

4.6.5 Seleção de modelos

Uma vez definido o conjunto de covariáveis (ou fatores) a ser incluído num modelo logístico, resta saber qual a melhor maneira de encontrar um modelo reduzido que inclua apenas as covariáveis e interações mais importantes para

explicar a probabilidade de sucesso $\pi(\mathbf{x})$. Esse problema poderia ser resolvido pelos métodos usuais de seleção de modelos discutidos na Seção 1.11. Contudo, a questão de interpretação dos parâmetros é crucial num modelo logístico, implicando que uma forma puramente mecânica de seleção pode levar a um modelo sem sentido e de difícil interpretação. Particularmente, a inclusão de certas interações impõe a permanência no modelo de seus respectivos efeitos principais de ordem inferior, na ótica do princípio hierárquico. Muitas vezes, variáveis consideradas biologicamente importantes não devem ser deixadas de lado pela sua falta de significância estatística. Assim, a seleção de um modelo logístico deve ser um processo conjugado de seleção estatística de modelos e bom senso.

Método *stepwise*

Um dos métodos mais aplicados em regressão logística é o método *stepwise*. O método, como foi visto na Seção 1.11, baseia-se num algoritmo misto de inclusão e eliminação de variáveis explicativas segundo a importância das mesmas de acordo com algum critério estatístico. Esse grau de importância pode ser avaliado, por exemplo, pelo nível de significância do teste da razão de verossimilhanças entre os modelos que incluem ou excluem as variáveis em questão. Quanto menor for esse nível de significância tanto mais importante será considerada a variável explicativa. Como a variável mais importante por esse critério não é necessariamente significativa do ponto de vista estatístico, deve-se impor um limite superior P_E (os valores usuais estão no intervalo $[0,15; 0,25]$) para esses níveis descritivos, a fim de atrair candidatos importantes em princípio à entrada.

Dado que a inclusão de novas variáveis explicativas num modelo pode tornar dispensáveis outras variáveis já incluídas, será feita a verificação da

importância dessas variáveis confrontando os seus respectivos níveis com um limite superior P_S. As variáveis explicativas com um nível descritivo maior do que P_S serão assim candidatas à remoção.

Descreve-se a seguir uma variante desse algoritmo aplicado em regressão logística (vide, por exemplo, Hosmer e Lemeshow, 1989). A etapa inicial consiste no ajuste do modelo apenas com o intercepto sendo completada pelos passos seguintes:

1. construir testes da razão de verossimilhanças entre o modelo inicial e os modelos logísticos simples formados com cada uma das variáveis explicativas do estudo. O menor dos níveis descritivos associados a cada teste será comparado com P_E. Se P_E for maior, a variável referente àquele nível é incluída no modelo indo ao passo seguinte. Caso contrário, a seleção é concluída e adota-se o último modelo;

2. partindo do modelo incluindo a variável explicativa selecionada no passo anterior, as demais variáveis são introduzidas individualmente. Cada um desses novos modelos é testado contra o modelo inicial desse passo. Novamente, o menor valor dos níveis descritivos é comparado com P_E. Se for menor do que P_E, implica na inclusão no modelo da variável correspondente e a passagem ao passo seguinte. Caso contrário, a seleção é finalizada;

3. compara-se o desvio do modelo logístico contendo as variáveis selecionadas nos passos anteriores com os desvios dos modelos que dele resultam por exclusão individual de cada uma das variáveis. Se o maior nível descritivo dos testes da razão de verossimilhanças for menor do que P_S, a variável explicativa associada a esse nível descritivo permanece no modelo. Caso contrário, a variável é removida. Em qualquer

circunstância, o algoritmo segue para o passo seguinte;

4. o modelo resultante do passo anterior será ajustado, no entanto, antes de tornar-se o modelo inicial da etapa 2 (seleção de interações de primeira ordem entre as variáveis explicativas incluídas), avalia-se a significância de cada um dos coeficientes das variáveis selecionadas, por exemplo através de um teste de Wald. Se alguma variável explicativa não for significativa pode ser excluí-la do modelo;

5. uma vez selecionadas as variáveis explicativas mais importantes, ou os efeitos principais, entra-se na etapa 2 com o passo 1 que agora envolve apenas interações de primeira ordem entre as variáveis selecionadas, e assim por diante.

É comum que algumas variáveis explicativas ou interações de interesse ou com algum significado no estudo sejam mantidas no modelo desde o início, mesmo que não sejam significativas. É também comum que a seleção de interações seja feita dentre aquelas de interesse ou com algum significado no problema.

Um aprimoramento desse procedimento tipo *stepwise* foi proposto posteriormente por Hosmer et al. (2013). Nesse novo algoritmo os autores sugerem que as variáveis explicativas eliminadas no passo 1 que causarem uma variação dsproporcional no(s) coeficiente(s) de alguma variável explicativa que permaneceu no modelo, devem ser trazidas de volta para o modelo. Os demais passos são similares, contudo o resultado final pode ser diferente, e segundo os autores em geral têm levado a resultados mais coerentes.

Uma desvantagem do procedimento descrito pelos passos 1-5 é de exigir as estimativas de máxima verossimilhança em cada passo, o que encarece o trabalho computacional, particularmente quando há muitas variáveis expli-

cativas (ou fatores). Alguns autores têm sugerido aproximações para esse processo de seleção. O aplicativo científico BMDP (Dixon, 1987) usa aproximações lineares nos testes da razão de verossimilhanças. Peduzzi et al. (1980) apresentam uma variante desse método baseada no uso da estatística de Wald.

Método de Akaike

Um procedimento mais simples para selecionar variáveis explicativas num modelo logístico é através do método de Akaike descrito na Seção 1.11. Uma sugestão é primeiro fazer uma seleção dos efeitos principais e depois num segundo passo, das interações de 1ª ordem. Para ilustrar uma aplicação do método, supor que as respostas binárias estejam armazenadas em `resp` e as variáveis explicativas sejam denotadas por `var1`, `var2` e `var3`. O ajuste do modelo logístico apenas com os efeitos principais pode ser realizado através dos comandos

```
ajuste <- glm(resp ~ var1 + var2 + var3, family=binomial).
```

A seleção dos efeitos principais pode ser realizada pelos comandos

```
require(MASS)
stepAIC(ajuste).
```

Eventualmente algumas variáveis explicativas selecionadas podem não ser significativas marginalmente e a retirada das mesmas do modelo poderá ser confirmada através de algum teste estatístico apropriado, como por exemplo o teste da razão de verossimilhanças. A inclusão de interações de 1ª ordem pode ser feita individualmente dentre aquelas interações de interesse ou de fácil interpretação.

4.6.6 Amostragem retrospectiva

Em muitas situações práticas, especialmente no estudo de doenças raras, pode ser mais conveniente a aplicação de uma amostragem retrospectiva em que um conjunto de n_1 casos (indivíduos com $y = 1$) e n_2 controles (indivíduos com $y = 0$) é selecionado aleatoriamente e classificado segundo os valores de $\mathbf{x} = (x_1, \ldots, x_p)^\top$. Esse tipo de planejamento é muitas vezes motivado por questões econômicas ligadas ao custo e a duração do experimento. A amostragem retrospectiva assim constituída levaria diretamente a um modelo para $Pr(\mathbf{X} = \mathbf{x}|y)$, ao contrário dos dados prospectivos que estão associados ao modelo $\pi(\mathbf{x}) = Pr(Y = y|\mathbf{x})$. Como o desenvolvimento de um modelo para $Pr(\mathbf{X} = \mathbf{x}|y)$ pode ficar muito complexo à medida que o valor $\mathbf{x}$ envolve um número maior de variáveis explicativas, particularmente contínuas, a proposta de uma abordagem alternativa através da especificação de um modelo para $Pr(Y = y|\mathbf{x})$, de modo a induzir um modelo para $Pr(\mathbf{X} = \mathbf{x}|y)$, tem sido utilizada.

Supor então um modelo logístico linear para explicar $\pi(\mathbf{x}) = Pr(Y = 1|\mathbf{x})$. Será mostrado a seguir que a probabilidade $\pi(\mathbf{x})$, a menos de uma constante adicionada ao intercepto do modelo, coincide com a probabilidade $\pi^*(\mathbf{x}) = Pr(Y = 1|\mathbf{x}, Z = 1)$ se a seleção amostral não depende de $\mathbf{x}$, em que Z é uma variável indicadora da classificação amostral (ver, por exemplo, Armitage, 1971). Denota-se $\gamma_1 = Pr(Z = 1|Y = 1)$ e $\gamma_2 = Pr(Z = 1|Y = 0)$, em que γ_1 é a probabilidade de um caso ser selecionado e γ_2 é a probabilidade de um controle ser selecionado da população global. A suposição é que γ_1 e γ_2 não dependem de $\mathbf{x}$. Portanto

$$\begin{aligned} \pi^*(\mathbf{x}) &= Pr(Y = 1|x, Z = 1) \\ &= \frac{Pr(Z = 1|Y = 1)Pr(Y = 1|\mathbf{x})}{\sum_{y=0,1} Pr(Z = 1|Y = y)Pr(Y = y|\mathbf{x})}, \end{aligned}$$

que pode ser expressa em função de $\pi(\mathbf{x})$, ou seja

$$\begin{aligned}\pi^*(\mathbf{x}) &= \frac{\gamma_1\pi(\mathbf{x})}{\gamma_2\{1-\pi(\mathbf{x})\}+\gamma_1\pi(\mathbf{x})} \\ &= \frac{\frac{\gamma_1}{\gamma_2}\left[\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\right]}{1+\frac{\gamma_1}{\gamma_2}\left[\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\right]}.\end{aligned}$$

Assim, obtém-se

$$\pi^*(\mathbf{x}) = \frac{e^{\log\{\gamma_1/\gamma_2\}+\eta}}{1+e^{\log\{\gamma_1/\gamma_2\}+\eta}},$$

em que $\eta = \sum_{j=1}^{p} x_j\beta_j$.

Portanto, fazendo uma amostragem retrospectiva e ajustando um modelo logístico como se fosse uma amostragem prospectiva, os coeficientes devem coincidir desde que a seleção tenha sido feita independente de $\mathbf{x}$. Se, no entanto, há interesse em estimar $\pi(\mathbf{x})$, isto é, fazer predições dado $\mathbf{x}$, deve-se corrigir a constante do modelo ajustado, obtendo um novo intercepto

$$\hat{\beta}_1 = \hat{\beta}_1^* - \log(\gamma_1/\gamma_2),$$

em que $\hat{\beta}_1^*$ é o intercepto do modelo ajustado.

4.6.7 Qualidade do ajuste

Como visto na Seção 4.4, quando o número de grupos k é fixo num experimento binomial e $\frac{n_i}{n} \to a_i > 0$ quando $n \to \infty$, o desvio $\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}})$ segue sob a hipótese do modelo adotado ser verdadeiro distribuição qui-quadrado com $(k-p)$ graus de liberdade. Esse resultado não vale quando $n \to \infty$ e $n_i\pi_i(1-\pi_i)$ fica limitado. Nesse caso, Hosmer e Lemeshow (1989) sugerem uma estatística alternativa para avaliação da qualidade do ajuste. Essa estatística é definida comparando o número observado com o número esperado de sucessos de g grupos formados. O primeiro grupo deverá conter

n_1' elementos correspondentes às n_1' menores probabilidades ajustadas, as quais serão denotadas por $\hat{\pi}_{(1)} \leq \hat{\pi}_{(2)} \leq \cdots \leq \hat{\pi}_{(n_1')}$. O segundo grupo deverá conter os n_2' elementos correspondentes às seguintes probabilidades ajustadas $\hat{\pi}_{(n_1'+1)} \leq \hat{\pi}_{(n_1'+2)} \leq \cdots \leq \hat{\pi}_{(n_1'+n_2')}$. E assim, sucessivamente, até o último grupo que deverá conter as n_g' maiores probabilidades ajustadas $\hat{\pi}_{(n_1'+\cdots+n_{g-1}'+1)} \leq \hat{\pi}_{(n_1'+\cdots+n_{g-1}'+2)} \leq \cdots \leq \hat{\pi}_{(n)}$. O número observado de sucessos no primeiro grupo formado será dado por $O_1 = \sum_{j=1}^{n_1'} y_{(j)}$, em que $y_{(j)} = 0$ se o elemento correspondente é fracasso e $y_{(j)} = 1$ se é sucesso. Generalizando, obtém-se $O_i = \sum_{j=n_1'+\cdots+n_{i-1}'+1}^{n_1'+\cdots+n_i'} y_{(j)}$, $2 \leq i \leq g$. A estatística é definida por

$$\hat{C} = \sum_{i=1}^{g} \frac{(O_i - n_i'\bar{\pi}_i)^2}{n_i'\bar{\pi}_i(1-\bar{\pi}_i)},$$

em que

$$\bar{\pi}_1 = \frac{1}{n_1'}\sum_{j=1}^{n_1'} \hat{\pi}_{(j)} \quad \text{e} \quad \bar{\pi}_i = \frac{1}{n_i'}\sum_{j=n_1'+\cdots+n_{i-1}'+1}^{n_i'+\cdots+n_i'} \hat{\pi}_{(j)},$$

para $2 \leq i \leq g$. Hosmer e Lemeshow sugerem a formação de $g = 10$ grupos de mesmo tamanho (aproximadamente), de modo que o primeiro grupo contenha n_i' elementos correspondentes às $[n/10]$ menores probabilidades ajustadas e assim por diante até o último grupo com n_{10}' elementos correspondentes às $[n/10]$ maiores probabilidades ajustados. Quando não há empates, isto é, $n_i = 1$, $\forall i$, fica relativamente fácil formar os 10 grupos com tamanhos aproximadamente iguais. No entanto, quando há empates, pode ser necessário que dois indivíduos com a mesma configuração de covariáveis sejam alocados em grupos adjacentes a fim de que os grupos formados não tenham tamanhos muito desiguais. Hosmer e Lemeshow verificaram através de simulações que a distribuição nula assintótica de $\hat{C}$ pode ser bem aproximada por uma distribuição qui-quadrado com $(g-2)$ graus de liberdade.

4.6.8 Técnicas de diagnóstico

Estudos de simulação (ver, por exemplo, Williams, 1984) têm sugerido o resíduo t_{D_i} para as análises de diagnóstico em modelos lineares generalizados, uma vez que o mesmo tem apresentado nesses estudos propriedades similares àquelas do resíduo t_i^* da regressão normal linear. Em particular, para os modelos binomiais, esse resíduo é expresso, para $0 < y_i < n_i$, na forma

$$t_{D_i} = \pm\sqrt{\frac{2}{1-\hat{h}_{ii}}}\left\{y_i \log\left(\frac{y_i}{n_i\hat{\pi}_i}\right) + (n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - n_i\hat{\pi}_i}\right)\right\}^{\frac{1}{2}},$$

em que o sinal é o mesmo de $y_i - \hat{y}_i$. Quando $y_i = 0$ ou $y_i = n_i$, o componente do desvio padronizado toma as formas

$$t_{D_i} = -\frac{\{2n_i|\log(1-\hat{\pi}_i)|\}^{\frac{1}{2}}}{\sqrt{1-\hat{h}_{ii}}} \quad \text{e} \quad t_{D_i} = \frac{\{2n_i|\log\hat{\pi}_i|\}^{\frac{1}{2}}}{\sqrt{1-\hat{h}_{ii}}},$$

respectivamente. O resíduo Studentizado t_{S_i}, também utilizado para avaliar a presença de observações aberrantes mesmo tendo em geral distribuição assimétrica acentuada, toma a forma

$$t_{S_i} = \frac{1}{\sqrt{1-\hat{h}_{ii}}}\frac{(y_i - n_i\hat{\pi}_i)}{\{n_i\hat{\pi}_i(1-\hat{\pi}_i)\}^{\frac{1}{2}}}.$$

Uma outra opção, conforme descrito na Seção 2.10, é o resíduo quantílico (Dunn e Smyth, 1996) definido para variáveis discretas por

$$r_{q_i} = \Phi^{-1}(u_i),$$

em que $\Phi(\cdot)$ denota a função de distribuição acumulada da $N(0,1)$ e u_i é um valor gerado no intervalo $(0,1)$ com base em $F(y_i;\hat{\boldsymbol{\beta}})$ (função de distribuição acumulada da distribuição discreta ajustada). Mostra-se para n grande que os resíduos $r_{q_1},\ldots,r_{q_n}$ são independentes e igualmente distribuídos $N(0,1)$. Assim, o gráfico entre os quantis amostrais $r_{q_{(1)}} \leq \cdots \leq r_{q_{(n)}}$ contra os

quantis teóricos da normal padrão é recomendado para avaliar afastamentos da distribuição postulada para a resposta.

O resíduo quantílico é disponibilizado na biblioteca `GAMLSS` do R (ver, por exemplo, Stasinopoulos et al., 2017) através dos comandos

```
require(gamlss)
plot(ajuste).
```

Aqui `ajuste` é o nome do objeto referente ao ajuste do modelo.

Contudo, no caso de variáveis discretas, o resíduo quantílico é aleatorizado e uma sugestão é gerar no GAMLSS m gráficos do `worm plot` (gráfico entre $r_{q_{(i)}} - \mathrm{E}(Z_{(i)})$ contra $\mathrm{E}(Z_{(i)})$) para avaliar com mais segurança a adequação do ajuste. Esse gráfico pode ser interpretado como um refinamento do gráfico normal de probabilidades podendo ser acionado para $m = 8$ gráficos através do comando

```
rqres.plot(ajuste, howmany=8, type=''wp'').
```

Por outro lado, para medir a influência das observações nas estimativas dos coeficientes, utiliza-se a distância de Cook aproximada dada por

$$\mathrm{LD}_i = \frac{\hat{h}_{ii}}{(1-\hat{h}_{ii})^2}\frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)}.$$

Hosmer e Lemeshow (1989) observam que $\hat{h}_{ii}$ depende das probabilidades ajustadas $\hat{\pi}_i$, $i = 1, \ldots, k$, e consequentemente os resíduos t_{S_i} e t_{D_i} e a medida de influência LD_i também dependem. Tem-se que

$$h_{ii} = n_i\pi_i(1-\pi_i)\mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{V}\mathbf{X})^{-1}\mathbf{x}_i,$$

com $\mathbf{V} = \mathrm{diag}\{n_1\pi_1(1-\pi_1), \ldots, n_k\pi_k(1-\pi_k)\}$. Hosmer e Lemeshow mostram através de um estudo numérico que o comportamento de $\hat{h}_{ii}$ numa regressão logística pode ser muito diferente do comportamento de $\hat{h}_{ii}$ na regressão linear para uma mesma matrix modelo $\mathbf{X}$.

Tabela 4.6
Possíveis valores para algumas medidas de diagnóstico segundo as probabilidades ajustadas.

	Probabilidade ajustada				
Medida	0,0-0,1	0,1-0,3	0,3-0,7	0,7-0,9	0,9-1,0
$t^2_{S_i}$	grande ou pequeno	moderado	moderado ou pequeno	moderado	grande ou pequeno
LD_i	pequeno	grande	moderado	grande	pequeno
$\hat{h}_{ii}$	pequeno	grande	moderado ou pequeno	grande	pequeno

A Tabela 4.6 descreve os possíveis valores de algumas medidas de diagnóstico em função das probabilidades ajustadas. A medida $\hat{h}_{ii}$ pode ser interpretada de maneira similar à medida h_{ii} da regressão normal linear para $0,1 \leq \hat{\pi}_i \leq 0,9$. No entanto, quando $\hat{\pi}_i$ é pequena ou alta, $\hat{h}_{ii}$ fica em geral pequeno o que pode dificultar a detecção de pontos que estejam mais afastados no subespaço gerado pelas colunas da matrix $\mathbf{X}$. A sugestão, portanto, são os gráficos de $t^2_{S_i}$, $t^2_{D_i}$ e LD_i contra as probabilidades ajustadas $\hat{\pi}_i$. Esses gráficos podem ser informativos a respeito do posicionamento dos pontos aberrantes e influentes com relação às probabilidades ajustadas. Os gráficos dessas quantidades contra $\hat{h}_{ii}$ podem ser complementares, pelo menos para verificar se as tendências apresentadas na Tabela 4.11 são confirmadas para o modelo ajustado.

Outros gráficos recomendados em regressão logística são os gráficos da variável adicionada e de $|\boldsymbol{\ell}_{max}|$ contra $\hat{\pi}_i$.

4.6.9 Aplicacões

Processo infeccioso pulmonar

Considere novamente os dados referentes a um estudo de caso-controle realizado no Setor de Anatomia e Patologia do Hospital Heliópolis em São Paulo,

no período de 1970 a 1982 (Paula e Tuder, 1986) discutido na Seção 2.10.2. Para simplicidade das análises, os níveis de HL e FF serão reagrupados de modo que os níveis de intensidade "ausente"e "discreto"sejam agora considerados como intensidade "baixa"e os níveis "moderado"e "intenso"sejam agora de intensidade "alta"conforme descrito na Tabela 4.7. Esses novos dados estão descritos no arquivo **canc3a.txt**.

Tabela 4.7
Descrição das novas variáveis referentes ao exemplo sobre processo infeccioso pulmonar.

Variável	Descrição	Valores
Y	Processo Infecioso	1:maligno
		0:benigno
IDADE	Idade	em anos
SEXO	Sexo	0:masculino
		1:feminino
HL	Intensidade de	1:alta
	Histiócitos-linfócitos	0:baixa
FF	Intensidade de	1:alta
	Fibrose-frouxa	0:baixa

Nesse estudo os pacientes foram amostrados retrospectivamente, sendo que os controles (processo benigno) foram formados por uma amostra de 104 pacientes de um grupo de 270, enquanto que os casos (processo maligno) foram todos os pacientes diagnosticados com processo infeccioso pulmonar maligno durante o período da pesquisa. Portanto, seguindo a notação da Seção 4.6.6 , tem-se que $\gamma_1 = 1$ e $\gamma_2 = 104/270$ [1].

O método de seleção *stepwise* proposto por Hosmer e Lemeshow (1989) será aplicado a seguir. Na etapa 1 considerou-se apenas os efeitos principais.

[1]Está sendo suposto que a razão $\gamma_1/\gamma_2 = 270/104$ vale também se as amostras tivessem sido extraídas diretamente da população

Foram considerados $P_E = 0,20$ (nível para inclusão de covariáveis) e $P_S = 0,25$ (nível para eliminação de covariáveis).

No passo 1 foi incluída a variável explicativa IDADE, uma vez que o nível descritivo dessa variável foi o menor dentre os níveis descritivos das demais variáveis explicativas e também foi menor do que P_E. No passo seguinte foi inbcluída a variável explicativa HL, e agora com duas variáveis incluídas no modelo verifica-se a possibilidade de eliminar uma das duas variáveis. O maior nível descritivo é da IDADE que encontra-se na Tabela 4.8 na linha de referência do passo 2. O nível descritivo dessa variável não é superior a P_S, logo IDADE é mantida no modelo. Seguindo essa lógica, tem-se os menores níveis descritivos em cada passo como sendo o elemento da diagonal principal de cada passo. No passo 3, por exemplo, entra a variável explicativa SEXO que tem o menor nível descritivo que por sua vez é menor do que P_E. Dado que SEXO entra no modelo, verifica-se a possibilidade de uma das duas variáveis incluídas no modelo ser retirada do modelo. Assim, no mesmo passo 3, nota-se que o maior nível descritivo (em asterisco) corresponde à variável explicativa HL que não deve sair do modelo, uma vez que o nível descritivo não é maior do que P_S. Seguindo essa mesma lógica todos os efeitos principais são incluídos no modelo. Em resumo, o modelo resultante na etapa 1 é o modelo com todos os efeitos principais.

De forma análoga procede-se a etapa 2, cujos níveis descritivos para tomada de decisão em cada passo encontram-se na Tabela 4.9. Por exemplo, no passo 1, entra a interação entre IDADE e HL que tem o menor nível descritivo que por sua vez é menor do que P_E. Não é verificado nessa etapa se algum efeito principal deve sair do modelo mesmo que fique não significativo com a inclusão das interações. Isso pode ser reavaliado após a seleção do modelo final. No passo 4, por exemplo, nota-se que a interação entre IDADE e FF

não entra no modelo pois o nível descritivo correspondente é maior do que P_E. Assim, como essa interação não entra no modelo, não é preciso verificar a retirada das demais interações já incluídas no modelo. Logo, tem-se apenas três interações de primeira ordem incluídas no modelo. Essas interações são IDADE $*$ HL, HL $*$ FF e SEXO $*$ FF.

Na etapa 3 nenhuma interação de segunda ordem foi selecionada, uma vez que o menor nível descritivo dos testes de inclusão foi menor do que P_E. Assim, o modelo resultante contém os efeitos principais e três interações de primeira ordem.

Tabela 4.8
Níveis descritivos referentes à etapa 1 do processo de seleção stepwise.

Passo	IDADE	HL	SEXO	FF
1	0,000	0,000	0,288	0,001
2	0,000	0,000	0,100	0,003
3	0,000	0,000*	0,050	0,125
4	0,000	0,000	0,072*	0,183
5	0,000	0,000	0,072	0,183*

O desvio do modelo foi de $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 146,22$ (167 graus de liberdade), indicando um ajuste adequado. As Figuras 4.1a-4.1d apresentam alguns gráficos de diagnóstico. Na Figura 4.1a temos o gráfico de $\hat{h}_{ii}$ contra os valores ajustados e nota-se dois pontos com maior destaque, #6 e #69. No gráfico de resíduos t_{D_i}, Figura 4.1b, a maioria dos pontos cai dentro do intervalo [-2,2], com apenas duas observações, #21 e #172, fora do intervalo, porém muito próximas aos limites. Já o gráfico de influência LD_i destaca novamente a observação #69 e a observação #172. O paciente #172 é do sexo feminino, tem processo maligno, idade 55 anos e níveis altos para HL e FF. Pelos resultados das estimativas seria mais provável esperar de um paciente com

esse perfil um processo benigno. O paciente #69 é também do sexo feminino, tem 78 anos, níveis altos para HL e FF e não tem processo maligno. Aqui seria um pouco menos provável processo benigno para o paciente. Perfil parecido tem o paciente #6. Já o paciente #21 tem processo benigno, 82 anos, é do sexo feminino e tem nível alto para HL e baixo para FF. Seria mais provável nesse caso processo maligno para o paciente.

Tabela 4.9
Níveis descritivos referentes à etapa 2 do processo de seleção stepwise.

Passo	IDA*HL	HL*FF	SEX*FF	IDA*FF	IDA*SEX	HL*SEX
1	0,013	0,014	0,059	0,056	0,657	0,063
2	0,023	0,027	0,060	0,231	0,218	0,099
3	0,028*	0,005	0,012	0,234	0,275	0,176
4				0,208	0,403	0,794

Finalmente, tem-se na Figura 4.1d o gráfico normal de probabilidades para o resíduo t_{D_i} e não apresentando nenhum indício de que a distribuição utilizada seja inadequada. Retirando cada uma das observações destacadas pelos gráficos de diagnóstico nota-se mudança inferencial quando a observação #172 é excluída, a interação SEXO $*$ FF deixa de ser significativa. Ou seja, a significância da interação SEXO $*$ FF é induzida pela observação #172. Logo, essa interação deve ser retirada do modelo.

As estimativas dos parâmetros do modelo final sem a interação SEXO $*$ FF bem como os valores padronizados pelos respectivos erros padrão aproximados encontram-se na Tabela 4.10.

Como há interesse em estudar a associação entre o tipo de processo infeccioso pulmonar e as covariáveis histológicas HL e FF, algumas razões de chances são construídas envolvendo essas covariáveis. Como ilustração, a razão de chances de processo infeccioso maligno entre um paciente no nível

alto de HL e um paciente no nível baixo de HL, denotada por ψ_{HL} e supondo que os pacientes tenham o mesmo sexo, idade e nível de FF, é estimada por

$$\hat{\psi}_{HL} = \exp\{-5,371 + 0,061\text{IDADE} + 2,255\text{FF}\}.$$

Tabela 4.10
Estimativas dos parâmetros referentes ao modelo logístico ajustado aos dados sobre processo infeccioso pulmonar.

Efeito	Parâmetro	Estimativa	E/E.Padrão
Constante	β_1^*	-1,247	-1,36
IDADE	β_2	0,038	2,23
HL	β_3	-5,371	-3,34
SEXO	β_4	0,765	1,60
FF	β_5	-2,090	-2,36
IDADE*HL	β_6	0,061	2,18
HL*FF	β_7	2,255	2,11

Logo, pode-se concluir que a chance de processo maligno é maior para pacientes com nível baixo de HL do que para pacientes com nível alto de HL, quando ambos estão no nível baixo de FF e também tenham a mesma idade. Por outro lado, quando ambos estão na categoria alta de FF, $\hat{\psi}_{HL}$ fica maior do que um após a idade de 52 anos (aproximadamente), indicando uma chance maior de processo maligno para pacientes no nível alto de HL após essa idade.

Analogamente, denota-se por ψ_{FF} a razão de chances de processo infeccioso maligno entre um paciente com nível alto de FF e um paciente com nível baixo de FF. Supondo que os pacientes são semelhantes nas demais covariáveis esse parâmetro é estimado por

$$\hat{\psi}_{FF} = \exp\{-2,090 + 2,255\text{HL}\}.$$

Dessa expressão pode-se deduzir que a chance de processo maligno é maior para pacientes com intensidade baixa de FF do que para pacientes com intensidade alta de FF, isso no grupo de pacientes com intensidade baixa de HL. Ocorre o contrário no grupo de pacientes com intensidade alta de HL. Bandas de confiança para ψ_{HL} e ψ_{FF} podem ser construídas com os procedimentos apresentados na Seção 4.6.4. Na comparação dos pacientes com relação ao sexo temos que a razão de chances de processo infeccioso pulmonar entre pacientes do sexo feminino e masculino é estimada por $\hat{\psi}_{FM} = \exp(0,765) = 2,15$.

Se o interesse em prever $Pr\{Y = 1|\mathbf{x}\}$, probabilidade de um paciente da população com um determinado conjunto de valores para as covariáveis estar com processo infeccioso maligno, deve-se antes estimar β_1 fazendo a correção

$$\hat{\beta}_1 = \hat{\beta}_1^* - \log(270/104) = -1,247 - 0,954 = -2,201.$$

Tabela 4.11
Discriminação do modelo logístico ajustado aos dados sobre processo infeccioso pulmonar.

Classificação	Classificação pelo modelo	
Correta	Benigno	Maligno
Benigno	81	23
Maligno	13	58

A regressão logística tem múltiplas utilidades, entre as quais a possibilidade de também ser utilizada em análise discriminante quando há apenas dois grupos para serem discriminados. O objetivo aqui é encontrar um modelo ajustado que melhor discrimine os dois grupos. Como aproximadamente 21% dos 341 pacientes foi diagnosticado com processo maligno pode-se verificar qual a taxa de acertos do modelo ajustado. Um critério seria classificarmos com processo maligno todo indivíduo com probabilidade ajustada de pelo

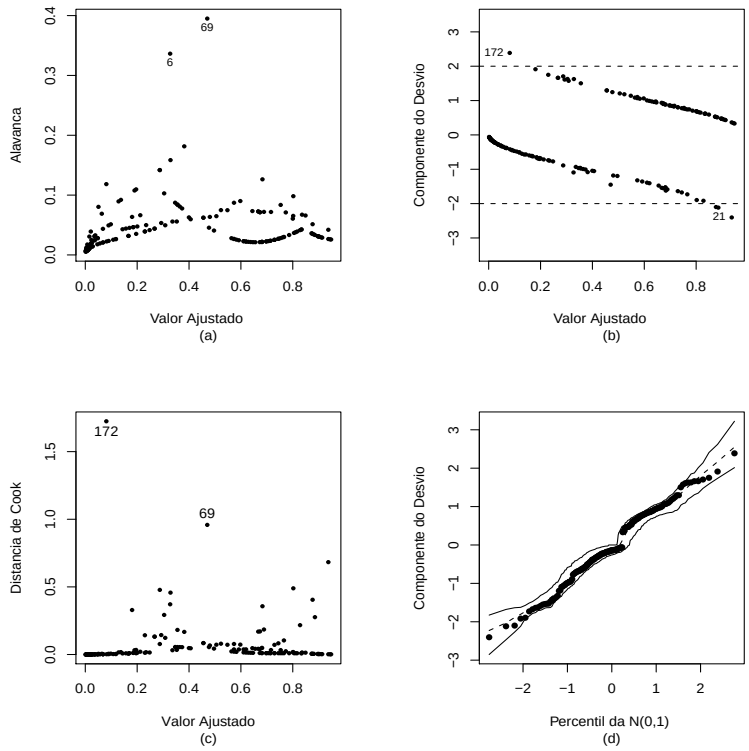

Figura 4.1: Gráficos de diagnóstico referentes ao modelo logístico ajustado aos dados sobre processo infeccioso pulmonar.

menos 0,21. Caso contrário o indivíduo seria classificado com processo benigno. A Tabela 4.11 apresenta a discriminaąão feita pelo modelo ajustado aos dados sobre processo infeccioso pulmonar. Tem-se que a taxa de acertos é de $139/175 = 0{,}795$ (79,5%). Uma outra forma de determinar o ponto de corte para a classificação dos paciente é através de curvas ROC.

Ocorrência de vaso-constrição

Como outra aplicação, conside os dados de um experimento desenvolvido para avaliar a influência da quantidade de ar inspirado na ocorrência de

vaso-constrição na pele dos dedos da mão (Finney, 1978; Pregibon, 1981). Os dados do experimento são descritos na Tabela 4.12 e também no arquivo **pregibon.txt**. A resposta, nesse exemplo, é a ocorrência ($Y = 1$) ou ausência ($Y = 0$) de compressão de vasos e as covariáveis são o logaritmo do volume e o logaritmo da razão de ar inspirado.

Tabela 4.12
Dados do experimento sobre a influência da razão e do volume de ar inspirado na ocorrência de vaso-constrição da pele dos dedos da mão.

Obs	Volume	Razão	Resposta	Obs.	Volume	Razão	Resposta
1	3,70	0,825	1	20	1,80	1,800	1
2	3,50	1,090	1	21	0,40	2,000	0
3	1,25	2,500	1	22	0,95	1,360	0
4	0,75	1,500	1	23	1,35	1,350	0
5	0,80	3,200	1	24	1,50	1,360	0
6	0,70	3,500	1	25	1,60	1,780	1
7	0,60	0,750	0	26	0,60	1,500	0
8	1,10	1,700	0	27	1,80	1,500	1
9	0,90	0,750	0	28	0,95	1,900	0
10	0,90	0,450	0	29	1,90	0,950	1
11	0,80	0,570	0	30	1,60	0,400	0
12	0,55	2,750	0	31	2,70	0,750	1
13	0,60	3,000	0	32	2,35	0,030	0
14	1,40	2,330	1	33	1,10	1,830	0
15	0,75	3,750	1	34	1,10	2,200	1
16	2,30	1,640	1	35	1,20	2,000	1
17	3,20	1,600	1	36	0,80	3,330	1
18	0,85	1,415	1	37	0,95	1,900	0
19	1,70	1,060	0	38	0,75	1,900	0
				39	1,30	1,625	1

Supor para a i-ésima unidade experimental que $Y_i \sim \text{Be}(\pi_i)$, em que

$$\log\left\{\frac{\pi_i}{1-\pi_i}\right\} = \beta_1 + \beta_2 \log\left(\text{volume}\right)_i + \beta_3 \log\left(\text{razão}\right)_i,$$

com π_i denotando a probabilidade de ocorrência de vaso-constrição.

As estimativas dos parâmetros são descritas na Tabela 4.13 e pode-se notar que as variáveis explicativas `log(volume)` e `log(razão)` são altamente significativas. O desvio do modelo foi de $\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}) = 29,36$ (com 36 graus de liberdade), indicando um ajuste adequado. As Figuras 4.2a-4.2d descrevem alguns dos gráficos sugeridos acima bem como o gráfico normal de probabilidades com envelope para o resíduo t_{D_i}. Na Figura 4.2a tem-se o gráfico de $\hat{h}_{ii}$ contra os valores ajustados e pode-se notar que a observação #31 é destacada mais do que as restantes.

Tabela 4.13
Estimativas dos parâmetros do modelo logístico ajustado aos dados sobre vaso-constrição.

Parâmetro	Estimativa	E/E.Padrão
β_1	-2,875	-2,18
β_2	5,179	4,85
β_3	4,562	2,49

Na Figura 4.2b tem-se o gráfico de LD_i contra os valores ajustados e pode-se notar duas observações mais discrepantes, #4 e #18, cujos valores ajustados são menores do que $0,11$. Uma tendência similar é exibida na Figura 4.2c onde tem-se o gráfico de $t^2_{S_i}$ contra os valores ajustados. A eliminação da observação #4 levou às novas estimativas $\hat{\beta}_1 = -5,204(2,17)$, $\hat{\beta}_2 = 7,452(2,93)$ e $\hat{\beta}_3 = 8,465(3,246)$ com variação, respectivamente, de -81%, 64% e 63%. O desvio do modelo reduziu para $\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}) = 22,42$ (35 g.l.), variação de 24%. Resultado parecido ocorreu com a eliminação da observação #18. Nesse caso obtém-se $\hat{\beta}_1 = -4,757(2,008)$, $\hat{\beta}_2 = 6,879(2,718)$ e $\hat{\beta}_3 = 7,669(2,937)$ com variação, respectivamente, de -66%, 48% e 51%. O desvio caiu para $\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}) = 23,58$ (35 g.l.), redução de 20%. Mesmo com as variações desproporcionais não houve mudança inferencial. Esses resultados indicam que os pontos #4 e #18 são influentes e aberrantes. Note que para os dois

casos houve ocorrência de ar inspirado, porém o valor do volume e da razão são relativamente baixos contrariando a tendência observada pelo modelo ajustado. O gráfico normal de probabilidades para o resíduo t_{D_i} (Figura 4.2d) não fornece indícios de afastamentos da suposição de distribuição binomial para a resposta. Pode-se notar que a maioria dos pontos caem dentro do envelope gerado.

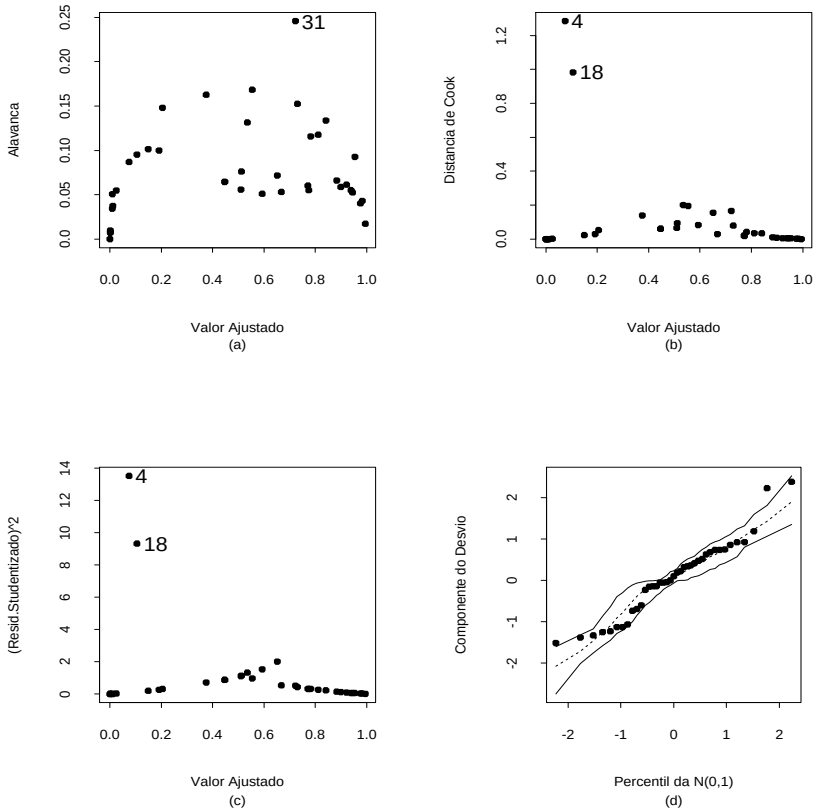

Figura 4.2: Gráficos de diagnóstico referentes ao modelo logístico ajustado aos dados sobre vaso-constrição.

Na Tabela 4.14 são apresentados os grupos formados com as observações

da Tabela 4.12 para o cálculo da estatística $\hat{C}$ proposta por Hosmer e Lemeshow (1989). Foram formados sete grupos com cinco observações cada e um grupo com quatro observações.

Tabela 4.14
Quantidades usadas para o cálculo da estatística $\hat{C}$ referente ao modelo logístico ajustado aos dados sobre vaso-constrição.

Grupo	Obervações	O_i	n'_i	$\bar{\pi}_i$
1	7,9,10,11,32	0	5	0,0024
2	4,18,21,26,30	2	5	0,0459
3	12,13,22,28,38	0	5	0,2737
4	8,19,23,29,37	1	5	0,5113
5	6,24,31,33,39	3	5	0,6728
6	5,15,34,35,36	5	5	0,7956
7	3,14,20,25,27	5	5	0,8974
8	1,2,16,17	4	4	0,9766

Os termos para o cálculo de $\hat{C}$ são dados abaixo

$$\begin{aligned} \hat{C} &= 0,0120 + 14,3157 + 1,8842 + 1,9391 \\ &+ 0,1203 + 1,2846 + 0,5716 + 0,0958 \\ &= 20,2233, \end{aligned}$$

cujo nível descritivo para uma qui-quadrado com 6 graus de liberdade é dado por P= $0,0025$, indicando que o ajuste não é adequado. Por outro lado, se eliminando as observações #4 e #18, obtém-se $\hat{C} = 5,9374$, que leva ao nível descritivo P= $0,4302$. Portanto, as duas observações destacadas pelas análises de diagnóstico têm grande influência na falta de ajuste detectada pela estatística $\hat{C}$.

Preferência de consumidores

Para ilustrar uma terceira aplicação com resposta binária será analisado parte dos dados descritos no arquivo **prefauto.txt** sobre a preferência de consu-

midores americanos com relação a automóveis. Uma amostra aleatória de 263 consumidores foi considerada. As seguintes variáveis foram observadas para cada comprador: preferência do tipo de automóvel (1: americano, 0: japonês), idade (em anos), sexo (0: masculino; 1: feminino) e estado civil (0: casado, 1: solteiro). Para maiores detalhes ver Foster et al.(1998, pgs. 338-339). Na Tabela 4.15 tem-se a distribuição da preferência do comprador segundo o sexo e estado civil, respectivamente.

Tabela 4.15
Distribuição da preferência do comprador de automóvel segundo o sexo e o estado civil.

	Masculino	Feminino
Americano	61 (42,4%)	54 (45,4%)
Japonês	83 (57,6%)	65 (54,6%)
Total	144	119
	Casado	Solteiro
Americano	83 (48,8%)	32 (34,4%)
Japonês	87 (51,2%)	65 (65,6%)
Total	170	93

Pode-se notar que para ambos os sexos a maior preferência é por carro japonês. Dentre os casados há pequena vantagem por carro japonês. Contudo, essa preferência é bem mais acentuada entre os solteiros. Pelos boxplots da Figura 4.3 nota-se que a idade mediana dos compradores de automóvel americano é ligeiramente superior à idade mediana dos compradores de automóvel japonês. Denotando por Y_i a preferência com relação ao tipo do automóvel pelo i-ésimo comprador (1: americano, 0: japonês), supor inicialmente um modelo logístico sem interação em que $Y_i \sim \text{Be}(\pi_i)$ com

$$\log\left\{\frac{\pi_i}{1-\pi_i}\right\} = \beta_1 + \beta_2 \text{idade}_i + \beta_3 \text{sexo}_i + \beta_4 \text{ecivil}_i,$$

sendo π_i a probabilidade do i-ésimo comprador preferir automóvel americano. Aplicando o método `AIC` a variável sexo é retirada do modelo. As estimativas dos parâmetros do modelo final sem interação são descritas na Tabela 4.16.

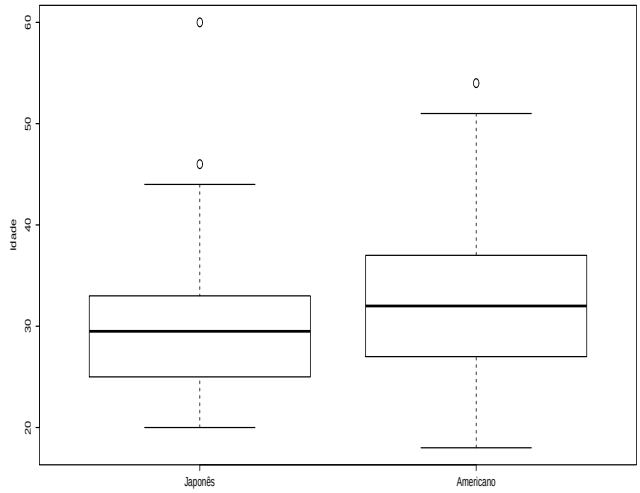

Figura 4.3: Boxplots da idade dos compradores de automóveis japonês e americano.

Assim, a probabilidade ajustada de preferência por automóvel americano fica expressa na forma

$$\hat{\pi} = \frac{\exp(-1,600 + 0,050 \times \text{Idade} - 0,526 \times \text{ECivil})}{1 + \exp(-1,600 + 0,050 \times \text{Idade} - 0,526 \times \text{ECivil})},$$

que é descrita na Figura 4.4 segundo a idade e o estado civil do comprador.

Tabela 4.16
Estimativas dos parâmetros referentes ao modelo logístico ajustado aos dados sobre preferência de compradores.

Efeito	Estimativa	E/E.Padrão
Constante	-1,600	-2,31
Idade	0,049	2,30
E.Civil	-0,526	-1,94

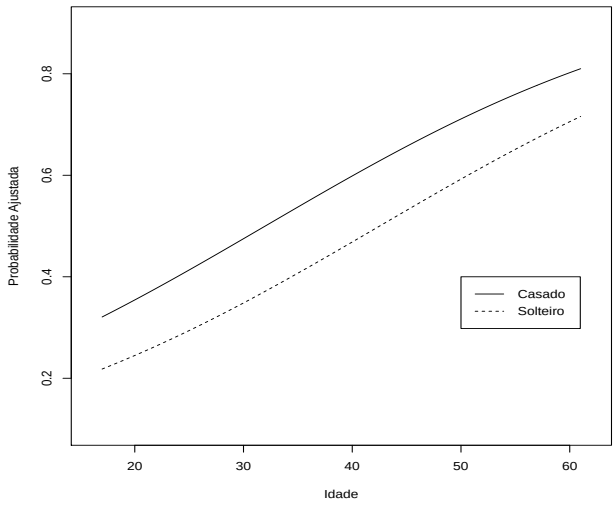

Figura 4.4: Probabilidade ajustada de preferência por caso americano segundo a idade e oestado civil do comprador.

Não foi significativa a inclusão no modelo da interação entre a idade e o estado civil do comprador. Assim, tem-se que a preferência por automóvel americano aumenta com a idade do comprador. Com relação ao estado civil nota-se que os casados preferem mais carro americano do que os solteiros. Essa razão de chances (entre casados e solteiros) por carro americano pode ser estimada por $\hat{\psi} = \exp(0,526) = 1,69$, enquanto uma estimativa intervalar aproximada de 90% para a razão de chances fica dada por

$$\begin{aligned} e^{0,526 \pm 1,65 \times 0,272} &= e^{0,526 \pm 0,449} \\ &= [1,080; 2,651][8,0\%; 165,1\%]. \end{aligned}$$

Portanto, um comprador casado tem uma chance entre 8% e 165,1% maior de preferir automóvel americano em relação a um comprador solteiro.

No gráfico da distância de Cook aproximada (Figura 4.5) a observação #99 (idade de 60 anos, solteira e prefere carro japonês) é destacada como

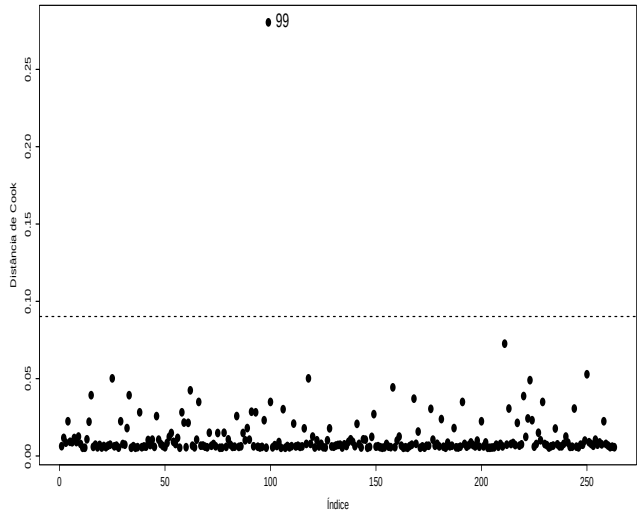

Figura 4.5: Distância de Cook referente ao modelo logístico ajustado aos dados sobre preferência de consumidores.

possívelmente influente, enquanto pela Figura 4.6 não há indícios de afastamentos importantes de suposição de distribuição binomial para a resposta. Tem-se na Tabela 4.17 as estimativas dos parâmetros sem a observação #99 e pode-se notar que, embora ocorram algumas variações desproporcionais, não há mudança inferencial. Essa compradora tem perfil com relação à idade de ter preferência por carro americano, e isso pode levado à discrepância com relação à distância de Cook.

Tabela 4.17
Estimativas dos parâmetros referentes ao modelo logístico ajustado aos dados sobre preferência de consumidores sem a observação #99.

Efeito	Estimativa	E/E.Padrão	Variação
Constante	-1,942	-2,65	-21,4%
Idade	0,060	2,65	22,4%
E.Civil	-0,474	-1,72	9,9%

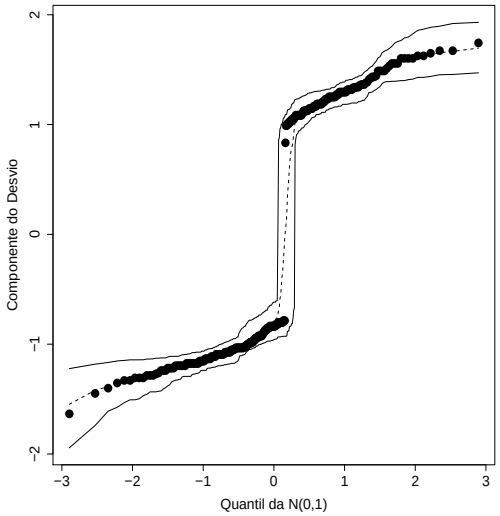

Figura 4.6: Gráfico normal de probabilidades para o resíduo componente do desvio referente ao modelo logístico ajustado aos dados sobre preferência de consumidores.

4.7 Curva ROC

A Curva Característica de Operação do Receptor (vide, por excemplo Agresti, 2013), também conhecida como curva ROC, é um procedimento gráfico de discriminação de dados binários que varia conforme variam taxas de verdadeiros positivos e falsos positivos. Assim, procura-se algum critério na curva ROC que maximize a taxa de verdadeiros positivos e minimize a taxa de falsos positivos.

Especificamente para regressão logística, denotando por $\hat{\pi}$ a probabilidade ajustada de sucesso de um modelo selecionado, o ojetivo principal é estabelecer algum critério para a probabilidade ajustada a fim de classificar um novo indivíduo como sendo sucesso ou fracasso. É esperado que esse novo indívio seja classificado como sendo sucesso à medida que $\hat{\pi}$ se aproxima de 1

e como fracasso à medida que $\hat{\pi}$ se aproxima de 0. Assim, definindo um ponto de corte para a probabilidade ajustada, pode-se construir para os dados da amostra uma tabela similar à Tabela 4.18, com as seguintes definções:

- Acurácia: proporção de predições corretas
  ACC = $\frac{\text{VP+VN}}{n}$.

- Sensibilidade: proporção de verdadeiros positivos
  SENS = $\frac{\text{VP}}{\text{VP+FN}}$
  1 - SENS: proporção de falsos negativos.

- Especificidade: proporção de verdadeiros negativos
  ESPEC = $\frac{\text{VN}}{\text{FP+VN}}$
  1 - ESPEC: proporção de falsos positivos.

Tabela 4.18
Tabela de classificação para dados binários.

Classificação	Classificação Correta		
pelo Modelo	Sucesso	Fracasso	Total
Sucesso	VP	FP	VP+FP
Fracasso	FN	VN	FN+VN
Total	VP+FN	FP+VN	n

A curva ROC para o exemple sobre preferência de consumidores é apresentada na Figura 4.7 e como pode ser observado a área sob a curva é pequena, dificultando encontrar um ponto de corte que corresponda a uma taxa de verdadeiros positivos alta e a uma taxa de falsos positivos pequena. Apenas, para ilustrar, supor ponto de corte de 0,44. Ou seja, classificar como comprador de automóvel americano se a probabilidade ajustada $\hat{\pi} \geq 0,44$ e como comprador de automóvel japonês se $\hat{\pi} < 0,44$. A classificação segundo esse critério para a amostra do exemplo de preferência de consumidores é descrita na Tabela 4.19 e nota-se taxas de acurácia, sensibilidade e especificidade, respectivamente, dadas por ACC = $\frac{68+86}{263} \cong 0,586(58,6\%)$, SENS

$= \frac{68}{115} \cong 0,591(59,1\%)$ e ESPEC $= \frac{86}{148} \cong 0,581(58,1\%)$, que podem ser consideradas baixas.

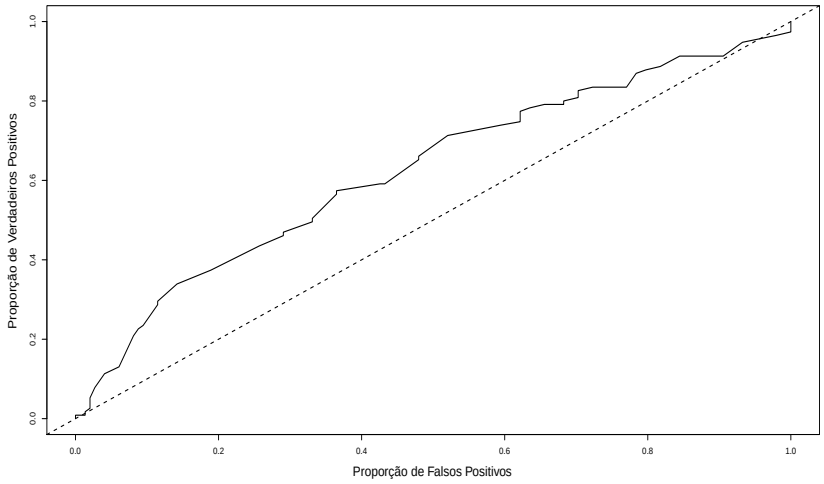

Figura 4.7: Curva ROC referente ao modelo logístico ajustado aos dados sobre preferência de consumidores.

Tabela 4.19
Tabela de classificação para o exemplo de preferência de consumidores.

Classificação	Classificação Correta		
pelo Modelo	Americano	Japonês	Total
Americano	68	62	130
Japonês	47	86	133
Total	115	148	263

4.8 Modelos de dose-resposta

O modelo logístico é frequentemente utilizado em Toxicologia no estudo do comportamento de determinados medicamentos, que é medido pela probabi-

lidade $\pi(x)$ de algum efeito produzido pelo medicamento em estudo, segundo a dose (ou a log-dose) x aplicada. Essa probabilidade pode ser escrita pela expressão geral

$$\pi(x) = \int_{-\infty}^{x} f(u)du, \tag{4.15}$$

em que $f(u)$ representa uma função densidade de probabilidade, também conhecida como função de tolerância. Como visto na Seção 2.3.1, alguns candidatos naturais para $f(u)$ são as funções de densidade da normal padrão, da distribuição logística e da distribuição do valor extremo, as quais levam aos modelos probito, logístico e complementar log-log, respectivamente. Utiliza-se o preditor linear $\eta = \beta_1 + \beta_2 x$ no lugar de x em (4.15) a fim de ampliar o leque de opções para $\pi(x)$.

Os modelos de dose-resposta visam não somente a predição da probabilidade de sucesso $\pi(x)$ para uma dosagem específica x, mas também a determinação da dosagem necessária para atingir uma probabilidade de sucesso p. Essa dosagem é chamada de dose letal. A notação usual para uma dose letal de $100p\%$ é dada por DL_{100p}. Logo,

$$p = \pi(\beta_1 + \beta_2 \mathrm{DL}_{100p}),\ 0 < p < 1.$$

A dose letal mais comum em Toxicologia é a dose mediana (DL_{50}), embora em certos casos sejam também de interesse doses extremas, tais como DL_1 ou DL_{99}. Deve-se observar que hoje em dia modelos de dose-resposta são definidos em várias áreas do conhecimento, em que a dose pode ser a idade, o peso, a resistência de um material, etc.

Supondo o modelo logístico com preditor linear $\eta = \beta_1 + \beta_2 x$, a estimativa de máxima verossimilhança de DL_{100p} fica, pela propriedade de invariância, dada por

$$\widehat{\mathrm{DL}}_{100p} = d(\hat{\beta}) = \frac{1}{\hat{\beta}_2}\left[\log\left(\frac{p}{1-p}\right) - \hat{\beta}_1\right],$$

em que $\hat{\boldsymbol{\beta}}$ é a estimativa de máxima verossimilhança de $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$.

A variância assintótica de $\widehat{\mathrm{DL}}_{100p}$ pode ser obtida após uma aproximação de primeira ordem por série de Taylor de $d(\hat{\boldsymbol{\beta}})$ em torno de $\boldsymbol{\beta}$, conhecido como método delta, levando ao seguinte resultado:

$$\mathrm{Var}_A[\widehat{\mathrm{DL}}_{100p}] = D(\boldsymbol{\beta})^\top (\mathbf{X}^\top \mathbf{V} \mathbf{X})^{-1} D(\boldsymbol{\beta}),$$

em que

$$D(\boldsymbol{\beta}) = \frac{\partial d(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \left[\frac{-1}{\beta_2}, \frac{1}{\beta_2^2} \left\{ \beta_1 - \log\left(\frac{p}{1-p} \right) \right\} \right]^\top.$$

Importante observar que $(\mathbf{X}^\top \hat{\mathbf{V}} \mathbf{X})^{-1}$ contém as variâncias e covariância estimadas de $\hat{\beta}_1$ e $\hat{\beta}_2$. Portanto, um intervalo de confiança assintótico de coeficiente $(1-\alpha)$ para DL_{100p} fica dado por

$$\widehat{\mathrm{DL}}_{100p} \pm z_{(1-\alpha/2)} \sqrt{\hat{\mathrm{Var}}_A[d(\hat{\boldsymbol{\beta}})]}.$$

4.8.1 Aplicações

Exposição de besouros

Em Bliss (1935) (ver também Silva, 1992) encontra-se uma situação típica para o ajuste de um modelo logístico de dose-resposta. O estudo baseia-se no comportamento de besouros adultos à exposição de disulfeto de carbono gasoso (CS_2) durante cinco horas. Os resultados obtidos a partir dos 481 besouros expostos segundo diferentes doses são apresentados na Tabela 4.20 e no arquivo **besouros.txt**.

Ajustando um modelo logístico do tipo $\mathrm{logit}\{\pi(x)\} = \beta_1 + \beta_2 x$ aos dados, em que x denota a dose de CS_2, obtém-se as estimativas $\hat{\beta}_1 = -60,72(5,18)$, $\hat{\beta}_2 = 34,27(2,91)$ e $\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -15,04$. O desvio do modelo foi de $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 11,23$ para 6 graus de liberdade, o que leva a um nível descritivo de P$= 0,0815$, indicando um ajuste razoável. O gráfico de envelope

descrito na Figura 4.8 confirma essa falta de ajuste. Talvez a inclusão de um termo quadrático ou mesmo o ajuste de um modelo logístico não linear (ver Silva, 1992) possam melhorar a qualidade do ajuste.

Tabela 4.20
Mortalidade de besouros expostos a disulfeto de carbono gasoso.

Dose $\log_{10} CS_2$	Besouros expostos	Besouros mortos
1,6907	59	6
1,7242	60	13
1,7552	62	18
1,7842	56	28
1,8113	63	52
1,8369	59	53
1,8610	62	61
1,8839	60	60

Uma vez conhecida a covariância assintótica entre $\hat{\beta}_1$ e $\hat{\beta}_2$, pode-se calcular a variância assintótica de $\widehat{\mathrm{DL}}_{100p}$ para alguns valores de p e consequentemente os intervalos assintóticos de confiança. Em particular, para $p = 0,50$, obtém-se a dose letal estimada

$$\begin{aligned}\widehat{\mathrm{DL}}_{50} &= \frac{1}{\hat{\beta}_2}\left[\log\left(\frac{0,5}{1-0,5}\right) - \hat{\beta}_1\right] \\ &= -\frac{\hat{\beta}_1}{\hat{\beta}_2} = \frac{60,72}{34,27} \\ &= 1,772.\end{aligned}$$

Um intervalo de confiança assintótico de 95% para DL_{50} fica então dado por

$$\begin{aligned}& 1,772 \ \pm \ 1,96\sqrt{(-0,029, -0,052)^\top (\mathbf{X}^\top \hat{\mathbf{V}} \mathbf{X})^{-1} \begin{pmatrix} -0,029 \\ -0,052 \end{pmatrix}} \\ = \ & 1,772 \ \pm \ 1,96\sqrt{0,00001488} \\ = \ & [1,764; 1,780].\end{aligned}$$

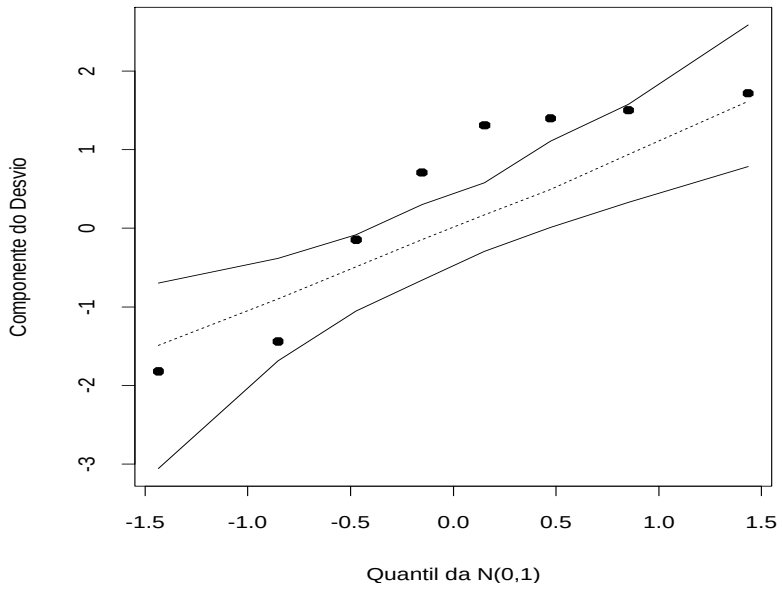

Figura 4.8: Gráfico normal de probabilidades referente ao modelo logístico ajustado aos dados sobre exposição de besouros.

A Figura 4.9 descreve a curva ajustada e as frequências observadas. Como pode-se observar os pontos abaixo de $\hat{\pi}(x) = 0,50$ parecem mais mal ajustados do que os pontos com resposta estimada acima desse valor. Isso sugere que um modelo binomial com ligação assimétrica poderia levar a um ajuste mais adequado. Uma opção poderia ser o modelo binomial com ligação complemento log-log, que é assimétrico em torno de $p = 0,50$ e cuja parte sistemática fica expressa na forma

$$\log\{-\log(1-\pi(x))\} = \beta_1 + \beta_2 x,$$

em que x denota a dose de CS_2. As estimativas paramétricas ficam dadas por $\hat{\beta}_1 = -39,57(3,24)$, $\hat{\beta}_2 = 22,04(1,80)$ e $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -5,82$. O desvio do modelo caiu para $\text{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 3,45$ com 6 graus de liberdade, que leva a um

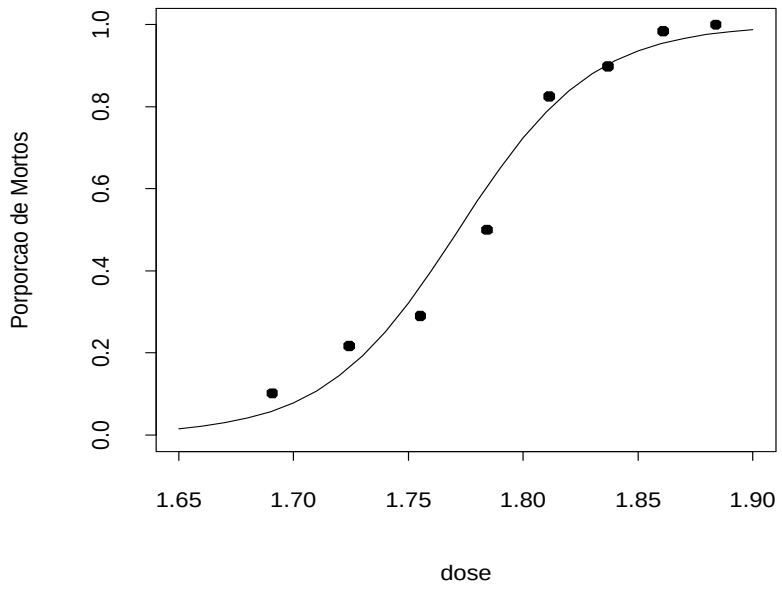

Figura 4.9: Modelo logístico ajustado à proporção de besouros mortos.

nível descritivo de P= $0,751$. Logo, não rejeita-se o modelo. O gráfico da curva ajustada (Figura 4.10a) e o gráfico normal de probabilidades (Figura 4.10b) confirmam essa indicação de modelo bem ajustado.

Para o modelo com ligação complemento log-log a estimativa de máxima verossimilhança de DL_{100p} fica dada por

$$\widehat{\mathrm{DL}}_{100p} = d(\hat{\beta}) = \frac{1}{\hat{\beta}_2}\left[\log\{-\log(1-p)\} - \hat{\beta}_1\right],$$

para a qual obtém-se a variância assintótica

$$\mathrm{Var}_A[\widehat{\mathrm{DL}}_{100p}] = D(\boldsymbol{\beta})^\top(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}D(\boldsymbol{\beta}),$$

em que

$$D(\boldsymbol{\beta}) = \frac{\partial d(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \left[\frac{-1}{\beta_2}, \frac{1}{\beta_2^2}\left\{\beta_1 - \log(-\log(1-p))\right\}\right]^\top,$$

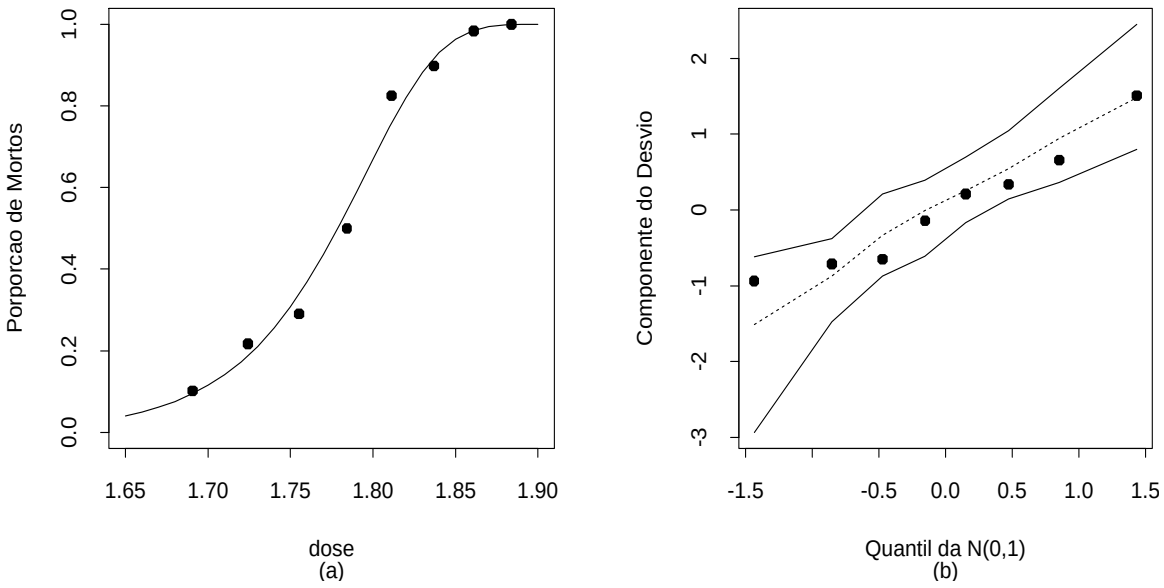

Figura 4.10: Curva ajustada para a proporção de besouros mortos (a) e gráfico normal de probabilidades sob o modelo complementar log-log (b).

com $\mathbf{W}$ sendo uma matriz diagonal de pesos dados por $\omega_i = n_i\pi_i^{-1}(1-\pi_i)\log^2(1-\pi_i)$ $i = 1, \ldots, 8$. Em particular, para $p = 0,50$, obtém-se

$$\begin{aligned}\widehat{\mathrm{DL}}_{50} &= \frac{1}{\hat{\beta}_2}\Big[\log\{-\log(1-0,5)\} - \hat{\beta}_1\Big] \\ &= \frac{1}{22,04}(-0,3665 + 39,57) \\ &= 1,779.\end{aligned}$$

Logo, um intervalo assintótico de 95% para DL_{50} fica dado por

$$\begin{aligned}& 1,779 \ \pm \ 1,96\sqrt{(-0,0454, -0,0807)^\top(\mathbf{X}^\top\hat{\mathbf{W}}\mathbf{X})^{-1}\begin{pmatrix}-0,0454\\-0,0807\end{pmatrix}} \\ = \ & 1,779 \ \pm \ 1,96\sqrt{0,00001606} \\ = \ & [1,771; 1,787].\end{aligned}$$

Pode-se notar que as estimativas intervalares para DL_{50} são praticamente as mesmas sob os dois modelos ajustados.

Garotas de Varsóvia

Os problemas de dose-resposta não se esgotam em Toxicologia. Milecer e Szczotka (1966) investigam a idade do início da menstruação em 3918 garotas de Varsóvia. Para 25 médias de idade foram observadas a ocorrência ($Y = 1$) ou não ($Y = 0$) do início de períodos de menstruação nas adolescentes. Os dados desse estudo são apresentados na Tabela 4.21 e no arquivo **meninas.txt**.

Tabela 4.21
Ocorrência do início da menstruação em garotas de Varsóvia.

	Número de garotas			Número de garotas	
Idade	Menstruadas	Entrevistadas	Idade	Menstruadas	Entrevistadas
9,21	0	376	13,08	47	99
10,21	0	200	13,33	67	106
10,58	0	93	13,58	81	105
10,83	2	120	13,83	88	117
11,08	2	90	14,08	79	98
11,33	5	88	14,33	90	97
11,58	10	105	14,58	113	120
11,83	17	111	14,83	95	102
12,08	16	100	15,08	117	122
12,33	29	93	15,33	107	111
12,58	39	100	15,58	92	94
12,83	51	108	15,83	112	114
			17,53	1049	1049

Considere o modelo logístico linear

$$\log\left\{\frac{\pi(x)}{1-\pi(x)}\right\} = \beta_1 + \beta_2 x,$$

em que $\pi(x) = Pr\{Y = 1|x\}$ e x denota a idade média. As estimativas de máxima verossimilhança deram $\hat{\beta}_1 = -21,23(0,769)$, $\hat{\beta}_2 = 1,63(0,059)$

e $\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) = -0,045$. Na Figura 4.11 são apresentadas a curva ajustada e as frequências observadas. O desvio do modelo foi de $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 26,80$ (23 graus de liberdade) para um nível descritivo de P$= 0,264$, indicando um ajuste adequado.

A estimativa da idade mediana de início do período de menstruação fica portanto dada por

$$\widehat{\mathrm{DL}}_{50} = \frac{21,23}{1,63} = 13,02,$$

com o seguinte intervalo assintótico de confiança de 95%:

$$13,02 \pm 1,96\sqrt{0,004524} = [12,89; 13,15].$$

Pelo gráfico de envelope descrito na Figura 4.12a nota-se que os resíduos apresentam uma tendência sistemática dentro do envelope gerado, sugerindo a inclusão de um termo quadrático na parte sitemática do modelo. O ajuste de um modelo com parte sistemática dada por

$$\eta(x) = \beta_1 + \beta_2 x + \beta_3 x^2$$

forneceu as seguintes estimativas: $\hat{\beta}_1 = -30,96(5,24)$, $\hat{\beta}_2 = 3,12(0,78)$ e $\hat{\beta}_3 = -0,06(0,03)$ com desvio $\mathrm{D}(\mathbf{y}, ; \hat{\boldsymbol{\mu}}) = 23,40$ (22 graus de liberdade) para um nível descritivo de P$= 0,38$. O gráfico de envelope descrito na Figura 4.12b confirma a adequação do modelo com termo quadrático.

Stukel (1988) (ver também Silva, 1992) mostra que o uso de um modelo logístico não linear pode melhorar substancialmente a qualidade do ajuste dos modelos de dose-resposta apresentados nesta seção.

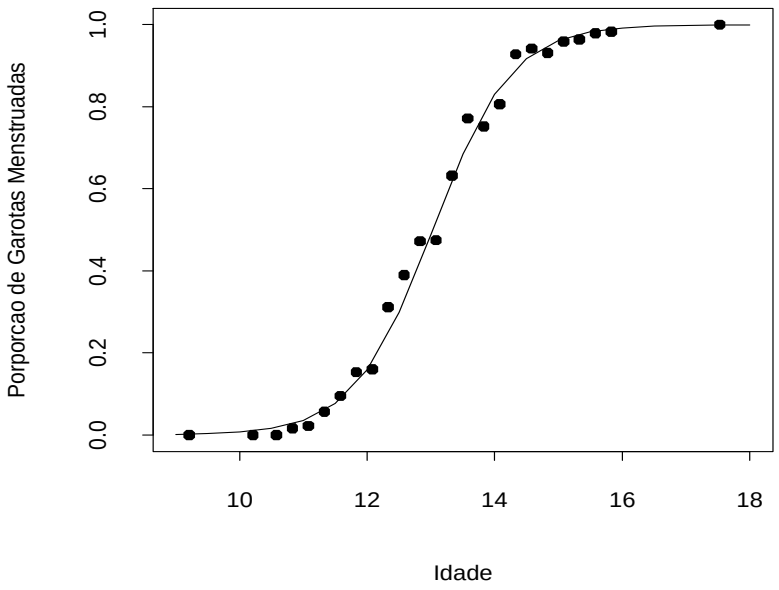

Figura 4.11: Curva ajustada pelo modelo logístico linear para a proporção de garotas de Varsóvia.

4.8.2 Estimação da dose letal

Intervalos de confiança aproximados para a dose letal DL_{100p} podem ser construídos utilizando a variância assintótica para $\widehat{\mathrm{DL}}_{100p}$, conforme descrito na seção anterior. Há, contudo, um outro método que é baseado no teorema de Fieller (1954) e será descrito a seguir.

Denote por $\rho = \frac{\beta_0}{\beta_1}$, em que β_0 e β_1 são estimados por $\hat{\beta}_0$ e $\hat{\beta}_1$, sendo assumido que essas estimativas são normalmente distribuídas com médias β_0 e β_1, variâncias v_{00} e v_{11} e covariância v_{01}. Defina a função $\hat{\psi} = \hat{\beta}_0 - \rho\hat{\beta}_1$. Então, se $\hat{\beta}_0$ e $\hat{\beta}_1$ são estimativas não viesadas de β_0 e β_1, obtém-se $\mathrm{E}(\hat{\psi}) = 0$. A variância de $\hat{\psi}$ fica, portanto, dada por

$$v = \mathrm{Var}(\hat{\psi}) = v_{00} + \rho^2 v_{11} - 2\rho v_{01}. \tag{4.16}$$

Desde que $\hat{\beta}_0$ e $\hat{\beta}_1$ são normalmente distribuídos, então $\hat{\psi}$ também é normal-

mente distribuído. Consequentemente, a variável

$$(\hat{\beta}_0 - \rho\hat{\beta}_1)/\sqrt{v}$$

segue distribuição normal padrão. Assim, um intervalo assintótico de confiança para ρ com coeficiente $(1-\alpha)$ é formado pelos valores de ρ tais que

$$\mid \hat{\beta}_0 - \rho\hat{\beta}_1 \mid \leq z_{(1-\alpha/2)}\sqrt{v}.$$

Os limites desse intervalo de confiança saem da equação quadrática

$$\hat{\beta}_0^2 + \rho^2\hat{\beta}_1^2 - 2\rho\hat{\beta}_0\hat{\beta}_1 - z_{(1-\alpha/2)}^2 v = 0,$$

que, após algumas manipulações algébricas e usando (4.16), fica dada por

$$(\hat{\beta}_1^2 - z_{(1-\alpha/2)}^2 v_{11})\rho^2 + (2v_{01}z_{(1-\alpha/2)}^2 - 2\hat{\beta}_0\hat{\beta}_1)\rho + \hat{\beta}_0^2 - v_{00}z_{(1-\alpha/2)}^2 = 0,$$

em que $z_{(1-\alpha/2)}$ denota o quantil $(1-\alpha/2)$ da distribuição normal padrão.

Portanto, as raízes da equação acima formam os limites inferior e superior do intervalo de confiança para ρ. Por exemplo, basta chamar $\rho = -\frac{\beta_1}{\beta_2}$ e aplicar os resultados acima para encontrar um intervalo assintótico de coeficiente $(1-\alpha)$ para a dose letal mediana DL_{50}.

4.8.3 Modelos de retas paralelas

Modelos de retas paralelas são comumente aplicados na área de Farmacologia para a comparação da eficiência de drogas do mesmo tipo, ou seja, com ação similar (ver, por exemplo, Finney, 1971; Collett, 1991). Nesses estudos, o interesse principal é comparar as potências entre as drogas definindo uma droga particular como nível base ou droga padrão. Para aplicar esses modelos em experimentos com respostas binárias é assumido que Y_{ijk}, o efeito

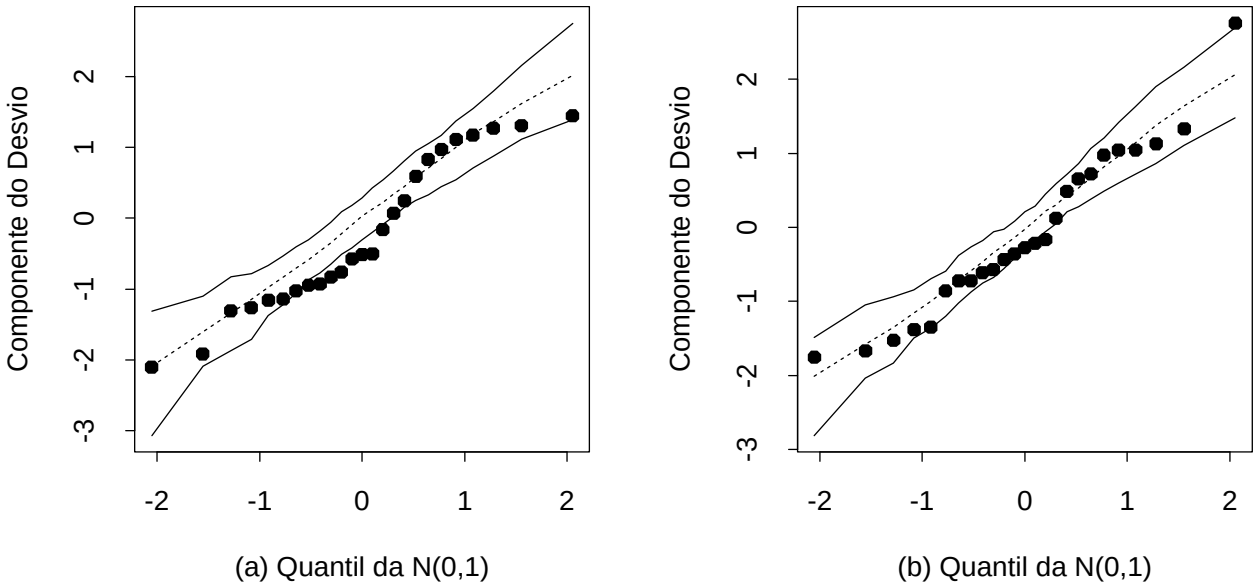

Figura 4.12: Gráficos normais de probabilidades para os modelos logístico com componente sistemática linear (a) e não linear (b) ajustados aos dados sobre garotas de Varsóvia.

produzido pela j-ésima dose correspondente à i-ésima droga no k-ésimo indivíduo, $i = 1, \ldots, g$, $j = 1, \ldots, d_i$ e $k = 1, \ldots, n_{ij}$, segue distribuição de Bernoulli com probabilidade de sucesso π_{ij} definida tal que

$$g(\pi_{ij}) = \alpha_i + \beta \log x_{ij}, \tag{4.17}$$

e que as variáveis Y_{ijk}'s são mutuamente independentes. Considerando a primeira droga como padrão, a potência ρ_i da i-ésima droga com relação à primeira é definida por

$$\log(\rho_i) = (\alpha_i - \alpha_1)/\beta,$$

$i = 1, \ldots, g$. Essa suposição leva à seguinte relação:

$$g(\pi_{ij}) = \alpha_1 + \beta \log(\rho_i x_{ij}),$$

isto é, x unidades da droga i têm o mesmo efeito que $\rho_i x$ unidades da primeira droga.

Aplicação

A Tabela 4.22 resume os resultados de um experimento (ver Collett, 1991) em que três inseticidas são aplicados num determinado tipo de inseto e é verificado o número de sobreviventes para cada dose aplicada. Esses dados estão também descritos no arquivo **insetic.txt**.

Tabela 4.22
Mortalidade de insetos segundo as doses de três inseticidas.

	Dose mg/cm^2					
Inseticida	2,00	2,64	3,48	4,59	6,06	8,00
DDT	3/50	5/49	19/47	19/50	24/49	35/50
γ-BHC	2/50	14/49	20/50	27/50	41/50	40/50
DDT + γ-BHC	28/50	37/50	46/50	48/50	48/50	50/50

Ajustando o modelo (4.17) com ligação logito aos dados, obtém-se as estimativas $\hat{\alpha}_1 = -4,555(0,361)$, $\hat{\alpha}_2 = -3,842(0,333)$, $\hat{\alpha}_3 = -1,425(0,285)$ e $\hat{\beta} = 2,696(0,214)$, com desvio dado por $\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}) = 21,282$, para 14 graus de liberdade, P= $0,0946$. Isso quer dizer que o ajuste do modelo de retas paralelas parece ser razoável.

Tem-se, portanto, os seguintes ajustes para as três drogas:

$$\log\left\{\frac{\hat{\pi}_1(x_j)}{1-\hat{\pi}_1(x_j)}\right\} = -4,555 + 2,696\log(x_j) \quad (\text{DDT});$$

$$\log\left\{\frac{\hat{\pi}_2(x_j)}{1-\hat{\pi}_2(x_j)}\right\} = -3,842 + 2,696\log(x_j) \quad (\gamma-\text{BHC}) \quad \text{e}$$

$$\log\left\{\frac{\hat{\pi}_3(x_j)}{1-\hat{\pi}_3(x_j)}\right\} = -1,425 + 2,696\log(x_j) \quad (\text{DDT} + \gamma-\text{BHC}),$$

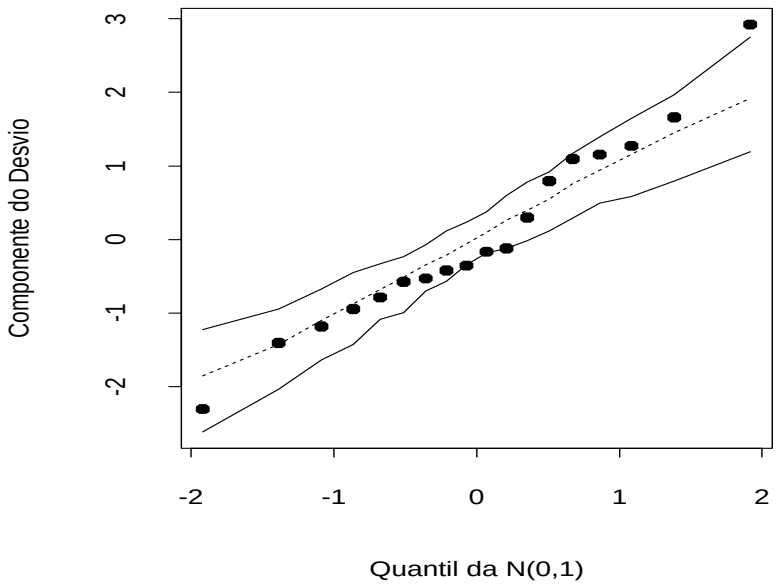

Figura 4.13: Gráfico normal de probabilidades para o modelo logístico de retas paralelas ajustado aos dados sobre três tipos de inseticida.

para $j = 1, \ldots, 6$. Nota-se, pelas estimativas, que há um aumento de potência quando as drogas DDT e γ-BHC são misturadas. Em particular, a potência da mistura com relação às drogas DDT e γ-BHC é estimada, respectivamente, por $\hat{\rho}_1 = \exp\{(-1,425+4,555)/2,696\} = 3,19$ e $\hat{\rho}_2 = \exp\{(-1,425+3,842)/2,696\} = 2,45$.

Pelo gráfico normal de probabilidades (Figura 4.13), nota-se que todos os resíduos caem dentro do envelope gerado. No entanto, parece haver uma tendência no gráfico, uma vez que os resíduos negativos apresentam-se ligeiramente abaixo da média enquanto os resíduos positivos apresentam-se ligeiramente acima. Isso pode ser um indício de sobredispersão, isto é, que as réplicas (para cada dose e cada inseticida) não são totalmente independentes. Em Collett (1991, Cap. 6) há uma discussão sobre o assunto. Apresenta-se a seguir uma abordagem para esse tipo de problema.

4.9 Sobredispersão

Sobredispersão ou variação extrabinomial é um fenômeno comum que ocorre na modelagem de dados binários agrupados e cuja ocorrência é caracterizada quando a variação observada excede aquela assumida pelo modelo (ver, por exemplo, Hinde e Demétrio, 1998). Em particular em regressão logística, quando o desvio $\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}})$ é maior do que o número de graus de liberdade $(g-p)$, pode haver indícios de sobredispersão, em que g é o número de grupos. Isso pode ser avaliado mais precisamente pelo nível descritivo do teste de ajustamento comparando $\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}})$ com os quantis da distribuição qui-quadrado com $(g-p)$ graus de liberdade.

Diferentes circunstâncias, entretanto, podem causar um valor alto para o desvio. Algumas delas representam uma sobredispersão aparente. Por exemplo, alguns pontos aberrantes podem aumentar substancialmente o valor do desvio e a simples eliminação desses pontos pode reduzir as evidências de sobredispersão. Outra causa aparente de sobredispersão é a ausência de algum termo extra na parte sistemática do modelo. Medidas de diagnóstico são ferramentas importantes para detectar o fenômeno. Em síntese, há duas possíveis causas de sobredispersão: correlação entre as réplicas binárias ou variação entre as probabilidades de sucesso de um mesmo grupo. Do ponto de vista prático é difícil distinguir entre os dois casos. Contudo, como será descrito a seguir, os procedimentos estatísticos para tratar a sobredispersão podem ser os mesmos.

4.9.1 Caso I

Supor inicialmente a existência de g grupos de modo que para o i-ésimo grupo sejam observadas n_i repetições de uma variável aleatória $Y_{ij} \sim \mathrm{Be}(\pi_i)$ (Bernoulli com probabilidade de sucesso π_i). O número total de sucessos no

i-ésimo grupo será definido por

$$Y_i = Y_{i1} + \cdots + Y_{in_i}.$$

Segue que $\mathrm{E}(Y_{ij}) = \pi_i$ e $\mathrm{Var}(Y_{ij}) = \pi_i(1-\pi_i)$. Supor adicionalmente a existência de correlação entre as repetições do i-ésimo grupo. Logo,

$$\mathrm{Var}(Y_i) = \sum_{j=1}^{n_i} \mathrm{Var}(Y_{ij}) + \sum_{j=1}^{n_i} \sum_{k=1,k\neq j}^{n_i} \mathrm{Cov}(Y_{ij}, Y_{ik}).$$

Se essa correlação é constante, $\mathrm{Corr}(Y_{ij}, Y_{ik}) = \delta$ para $j \neq k$, então tem-se que $\mathrm{Cov}(Y_{ij}, Y_{ik}) = \delta\pi_i(1-\pi_i)$. Daí obtém-se

$$\begin{aligned}
\mathrm{Var}(Y_i) &= \sum_{j=1}^{n_i} \pi_i(1-\pi_i) + \sum_{j=1}^{n_i} \sum_{k=1,k\neq j}^{n_i} \delta\pi_i(1-\pi_i) \\
&= n_i\pi_i(1-\pi_i) + n_i(n_i-1)\delta\pi_i(1-\pi_i) \\
&= \sigma_i^2 n_i\pi_i(1-\pi_i),
\end{aligned}$$

em que $\sigma_i^2 = 1 + (n_i - 1)\delta$. Se é exigido que $\sigma_i^2 > 0$, então deve-se ter

$$1 + (n_i - 1)\delta > 0,$$

que implica em $\delta > -1/(n_i - 1)$. Portanto, haverá a restrição

$$-\frac{1}{n_i - 1} \leq \delta \leq 1.$$

Assim, δ assumirá valores negativos apenas para n_i pequeno. Caso contrário, δ assumirá valores em geral positivos. Logo, tem-se em geral $\mathrm{Var}(Y_i) > n_i\pi_i(1-\pi_i)$ (sobredispersão).

4.9.2 Caso II

Supor agora que p_i representa a probabilidade de sucesso nas respostas do i-ésimo grupo tal que $\mathrm{E}(p_i) = \pi_i$ e $\mathrm{Var}(p_i) = \delta\pi_i(1-\pi_i)$, $\delta \geq 0$. Tem-se

portanto um modelo de efeito aleatório, que reduz ao modelo usual de efeito fixo fazendo $\delta = 0$. Assumindo ainda que $Y_{ij}|p_i \sim \text{Be}(p_i)$ de onde segue que $\text{E}(Y_{ij}|p_i) = p_i$ e $\text{Var}(Y_{ij}|p_i) = p_i(1-p_i)$. Daí obtém-se

$$\text{E}(Y_i) = \text{E}\{\text{E}(Y_i|p_i)\} = n_i\pi_i$$

e

$$\begin{aligned}
\text{Var}(Y_i) &= \text{E}\{\text{Var}(Y_i|p_i)\} + \text{Var}\{\text{E}(Y_i|p_i)\} \\
&= n_i\pi_i(1-\pi_i)(1-\delta) + n_i^2\delta\pi_i(1-\pi_i) \\
&= n_i\pi_i(1-\pi_i)\{1+(n_i-1)\delta\},
\end{aligned}$$

que coincidem com os resultados obtidos para o primeiro caso. No entanto aqui tem-se a restrição $\delta \geq 0$.

4.9.3 Estimação

A estimação de δ tem sido discutida em vários contextos. No primeiro caso, por exemplo, δ pode ser consistentemente estimado por

$$\tilde{\delta} = \sum_{i=1}^{g}\sum_{\ell'<\ell} \hat{r}_{P_{i\ell}}\hat{r}_{P_{i\ell'}}/(N-p), \tag{4.18}$$

em que $\hat{r}_{P_{i\ell}} = (y_{i\ell} - \hat{\pi}_i)/\sqrt{\hat{\pi}_i(1-\tilde{\pi}_i)}$ é o resíduo de Pearson estimado, $N = \frac{1}{2}\sum_{i=1}^{g} n_i(n_i-1)$ e $\hat{\pi}_i$ é a estimativa de máxima verossimilhança de π_i supondo $\delta = 0$. Contudo, deve-se estimar $\boldsymbol{\beta}$ e δ simultaneamente através de um processo iterativo. Uma proposta é o uso de equações de estimação generalizadas (Liang e Zeger, 1986) as quais serão discutidas no Capítulo 5. As novas estimativas, denotadas por $\hat{\boldsymbol{\beta}}_G$ e $\hat{\delta}$, saem do sistema de equações

$$\sum_{i=1}^{g}\{1+(n_i-1)\hat{\delta}\}^{-1}\mathbf{x}_i(y_i - n_i\hat{\pi}_i) = \mathbf{0}.$$

Dada uma estimativa inicial para δ, que pode ser $\tilde{\delta}$, tem-se o seguinte processo iterativo para obter $\hat{\boldsymbol{\beta}}_G$:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \{\sum_{i=1}^{g} \omega_i^{(m)} \mathbf{x}_i \mathbf{x}_i^\top\}^{-1} \sum_{i=1}^{g} \{\omega_i^{(m)} \mathbf{x}_i (y_i - n_i \pi_i^{(m)}) / n_i \pi_i^{(m)} (1 - \pi_i^{(m)})\}, \tag{4.19}$$

$m = 0, 1, 2 \ldots$, em que $\omega_i = n_i \pi_i (1 - \pi_i) / \{1 + (n_i - 1)\hat{\delta}\}$. O processo iterativo (4.19) é alternado com (4.18) até chegar à convergência. Pode-se mostrar que o estimador $\hat{\boldsymbol{\beta}}_G$ é consistente e assintoticamente normal. A variância assintótica de $\hat{\boldsymbol{\beta}}_G$ é dada por

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_G) = \{\sum_{i=1}^{g} \omega_i \mathbf{x}_i \mathbf{x}_i^\top\}^{-1}.$$

Há também uma proposta de variância assintótica robusta no caso da estrutura de correlação ter sido definida incorretamente, que é dada por

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_G) = \{\sum_{i=1}^{g} \omega_i \mathbf{x}_i \mathbf{x}_i^\top\}^{-1} \{\sum_{i=1}^{g} \nu_i \mathbf{x}_i \mathbf{x}_i^\top\} \{\sum_{i=1}^{g} \omega_i \mathbf{x}_i \mathbf{x}_i^\top\}^{-1},$$

em que $\nu_i = \{1 + (n_i - 1)\delta\}^{-2} \sum_{\ell,\ell'} (y_{i\ell} - \pi_i)(y_{i\ell'} - \pi_i)$. Um desvio corrigido para esse modelo fica dado por

$$\mathrm{D}(y; \hat{\mu}_G) = \sum_{i=1}^{g} \{1 + (n_i - 1)\hat{\delta}\} d_i^2,$$

em que d_i^2 é o i-ésimo componente do desvio de um modelo binomial, avaliado em $\hat{\boldsymbol{\beta}}_G$.

A seguir são descritos procedimentos para resolver (4.19) no `R`. Inicialmente é definida uma função `corpearson` para obter (4.18). Denota-se os vetores $(y_1/n_1, \ldots, y_g/n_g)^\top$, $(y_1, \ldots, y_g)^\top$ e $(n_1, \ldots, n_g)^\top$ por `fr`, `yt` e `nt`, respectivamente, e o número de parâmetros por `npar`. A função é definida por

```
corpearson = function(fr, yt, nt, npar) {
nt1 = 0.5*sum(nt*(nt-1))
sum1 = (0.5*yt*(yt-1) - fr*(nt-1)*yt +
0.5*fr*fr*nt*(nt-1))/(fr*(1-fr))
sum1 = sum(sum1)
rho = sum1/(nt1-npar)
rho }.
```

Supor que há duas variáveis explicativas representadas por `x1` e `x2` sem intercepto e que os resultados do ajuste do modelo supondo independência sejam colocados em `fit.model`. Em `fit.gee` são armazenados os resultados do processo iterativo dado em (4.19) e supor ainda 10 iterações. Seguem os comandos

```
fit.model = glm(resp ~ x1 + x2 - 1, family=binomial)
eta = predict(fit.model)
fr = fitted(fit.model)
rr = corpearson(fr, yt, nt, npar)
i = 1
while(i <= 10) {
fit.gee = glm(resp ~ x1 + x2 -1, family=binomial, start=
mu = exp(eta)/(1 + exp(eta)),
maxiter = 1,
weights = 1/(1 + (nt - 1)*rr))
eta = predict(fit.gee)
fr = fitted(fit.gee)
rr = corpearson(fr, yt, nt, npar)
i = i + 1 }.
```

A estimativa final da correlação está armazenada em `rr`. Para rodar os

programas descritos acima no R deve-se armazenar inicialmente a função corpearson num arquivo externo, por exemplo denominado corr.s, e executar o mesmo através do comando abaixo

```
source(‘‘corr.s’’).
```

Então a função corpearson estará instalada. Em seguida deve-se fazer o mesmo para ajustar o modelo colocando os demais comandos num arquivo externo, por exemplo denominado super.s, fazendo o seguinte:

```
source(‘‘super.s’’).
```

4.9.4 Teste de ausência de sobredispersão

Pode ser de interesse testar a hipótese de ausência de sobredispersão H_0 : $\delta = 0$ contra H_1 : $\delta > 0$. Como o conhecimento da distribuição de Y_{ij} é mais complexo sob a hipótese alternativa, dificultando a aplicação de testes tradicionais tais como razão de verossimilhanças, Wald e escore, a proposta de aplicar um teste tipo escore que requer apenas o conhecimento dos dois primeiros momentos de Y_{ij} com a estatística do teste sendo avaliada sob a hipótese nula (modelo binomial de respostas independentes) torna-se atrativo. Uma estatística do teste proposta por Paula e Artes (2000) é expressa na forma assume a forma

$$\xi_S = \frac{\sum_{i=1}^{g} \hat{M}_i}{\sqrt{\sum_{i=1}^{g} \hat{M}_i^2}},$$

em que $\hat{M}_i = \sum_{\ell<\ell'} \hat{r}_{P_{i\ell}} \hat{r}_{P_{i\ell'}}$ de modo que H_0 seja rejeitada quando $\xi_S > z_{(1-\alpha)}$. Pode-se mostrar que essa estatística corresponde à forma padronizada (sob H_0) de $\tilde{\delta}$. Para calcular ξ_S tem-se a função abaixo em que fr denota os valores ajustados sob a hipótese nula.

```
escore = function(fr,yt,nt) {
sum1 = (0.5*yt*(yt-1) - fr*(nt-1)*yt +
```

```
0.5*fr*fr*nt*(nt-1))/(fr*(1-fr))
sum2 = sum(sum1*sum1)
sum1 = sum(sum1)
escore = sum1/sqrt(sum2)
escore }.
```

4.9.5 Modelo beta-binomial

Uma outra possibilidade para estudar o fenômeno de sobredispersão é através do uso do modelo beta-binomial, em que variáveis aleatórias Y e Z são definidas tais que

$$Y|z \sim \text{B}(n, z) \text{ e } Z \sim \text{Beta}(\mu, \sigma),$$

com $0 < z, \mu < 1$ e $\sigma > 0$. Então, após algumas manipulações algébricas, pode-se mostrar que a distribuição marginal de Y é dada por

$$Y \sim \text{BB}(n, \mu, \sigma),\ y = 0, 1, \ldots, n,$$

com $\text{E}(Y) = n\mu$ e $\text{Var}(Y) = n\mu(1-\mu)\{1 + (n-1)\sigma^2\}$. Ou seja, tem-se a distribuição beta-binomial com mesmo domínio e mesma média da binomia, contudo com variância maior do que a variância da binomial. A distribuição beta-binomial não pertene à família exponencial, contudo pode ser ajustada através da biblioteca GAMLSS (Stasinopoulos et al., 2017) e pode contemplar os dois tipos de situações descritos na Seção 1.9 que geram sobredispersão com dados binários.

4.9.6 Quase-verossimilhança

Pode-se ainda supor $\sigma_i^2 = \phi^{-1}$ e estimar ϕ consistentemente dos dados ou do modelo ajustado substituindo a estimativa obtida nas quantidades que

envolvem ϕ. Quando n_i é grande, $\forall i$, pode-se estimar ϕ diretamente do desvio

$$\hat{\phi}^{-1} = \frac{\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})}{g-p}.$$

No caso de n_i pequeno, para algum i, recomenda-se a estimativa abaixo

$$\hat{\phi}^{-1} = \frac{1}{g-p} \sum_{i=1}^{g} \frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)},$$

em que p denota o número de parâmetros da parte sistemática do modelo e $\hat{\pi}_1, \ldots, \hat{\pi}_g$ são as probabilidades ajustadas nos g grupos. Sob a hipótese de que o modelo é verdadeiro, essa estimativa é também consistente para ϕ. Essa opção é um caso particular de modelos de quase-verossimilhança que serão discutidos no Capítulo 6.

No exemplo da seção anterior, envolvendo a comparação de três inseticidas, tem-se um total de 18 grupos com probabilidades ajustadas $\hat{\pi}_i(x_j)$, $i = 1, 2, 3$ e $j = 1, \ldots, 6$. Como $n_i = 50$ para a maioria dos grupos, pode-se estimar ϕ consistentemente através de

$$\hat{\phi}^{-1} = \frac{\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})}{g-p} = \frac{21,282}{14} = 1,52.$$

Algumas quantidades que envolvem ϕ deverão ser corrigidas,

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \hat{\phi}^{-1}(\mathbf{X}^\top \mathbf{V} \mathbf{X})^{-1} \text{ e } \mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \hat{\phi}\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})$$

com $t^*_{D_i} = \sqrt{\hat{\phi}}t_{D_i}$. O novo gráfico normal de probabilidades, agora com $t^*_{D_i}$, é apresentado na Figura 4.14 e não apresenta indícios de afastamentos sérios das suposições feitas para o modelo. É importante observar que o novo resíduo $t^*_{D_i}$ não corresponde ao componente do desvio de nenhum modelo particular. Nos modelos de quase-verossimilhança a distribuição da resposta

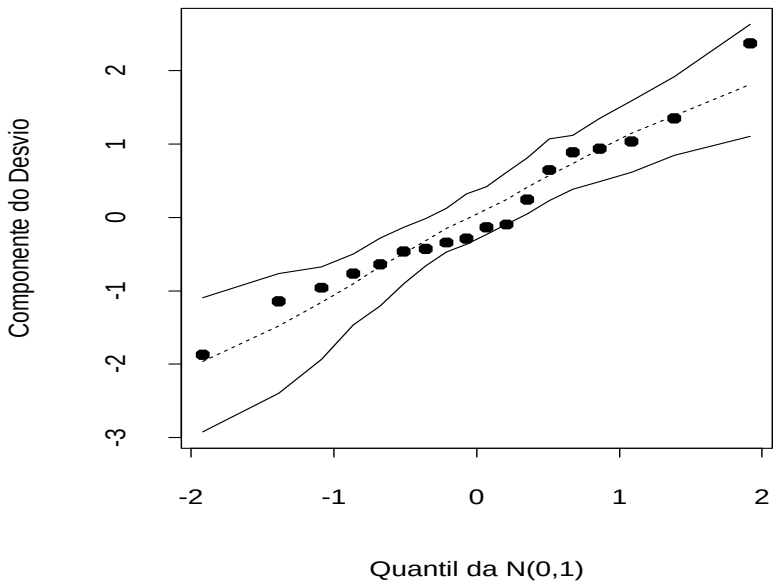

Figura 4.14: Gráfico normal de probabilidades para o resíduo corrigido para o modelo logístico de retas paralelas ajustado aos dados sobre três tipos de inseticida.

é em geral desconhecida e o uso de $D^*(\mathbf{y}; \boldsymbol{\mu})$ deve ser encarado de forma descritiva.

4.9.7 Aplicação

Collett (1991, Seção 6.9) descreve um experimento com duas espécies de *rotifers*, um tipo microscópico de invertebrado aquático. O objetivo do experimento é determinar a densidade relativa para cada uma das espécies. Foi utilizado um método indireto que consiste em centrifugar os animais em recipientes com densidades relativas de uma determinada substância e então utilizar uma regressão logística para ajustar a proporção de *rotifers* que permanecem suspensos segundo a densidade relativa. A densidade relativa de cada espécie pode ser estimada pela DL_{50}, que nesse caso representa a densidade relativa da substância que deixa suspenso 50% de *rotifers*.

Tabela 4.21
Distribuição de rotifers das duas espécies.

	Polyarthra major		Keratella cochlearis	
Densidade	Suspensos	Expostos	Suspensos	Expostos
1,019	11	58	13	161
1,020	7	86	14	248
1,021	10	76	30	234
1,030	19	83	10	283
1,030	9	56	14	129
1,030	21	73	35	161
1,031	13	29	26	167
1,040	34	44	32	286
1,040	10	31	22	117
1,041	36	56	23	162
1,048	20	27	7	42
1,049	54	59	22	48
1,050	20	22	9	49
1,050	9	14	34	160
1,060	14	17	71	74
1,061	10	22	25	45
1,063	64	66	94	101
1,070	68	86	63	68
1,070	488	492	178	190
1,070	88	89	154	154

Seja Y_{ij} o número de animais da i-ésima espécie que permanecem suspensos num recipiente com densidade relativa d_j da solução, onde foram colocados n_{ij} *rotifers*. É assumido inicialmente que $Y_{ij} \sim \mathrm{B}(n_{ij}, \pi_{ij})$, $i = 1, 2$ e $j = 1, \ldots, 20$, em que

$$\log\left\{\frac{\pi_{ij}}{1-\pi_{ij}}\right\} = \alpha_i + \beta_i d_j.$$

Na Tabela 4.21 e no arquivo **rotifers.txt** são apresentados para cada espécie a densidade relativa da substância, o número de *rotifers* expostos e o número de *rotifers* em suspensão. Para a espécie Polyathra as estimativas de máxima

verossimilhança são dadas por $\hat{\alpha}_1 = -109,72(5,22)$ e $\hat{\beta}_1 = 105,67(5,02)$, enquanto que para a espécie Keratella obtém-se $\hat{\alpha}_2 = -114,35(4,03)$ e $\hat{\beta}_2 = 108,75(3,86)$.

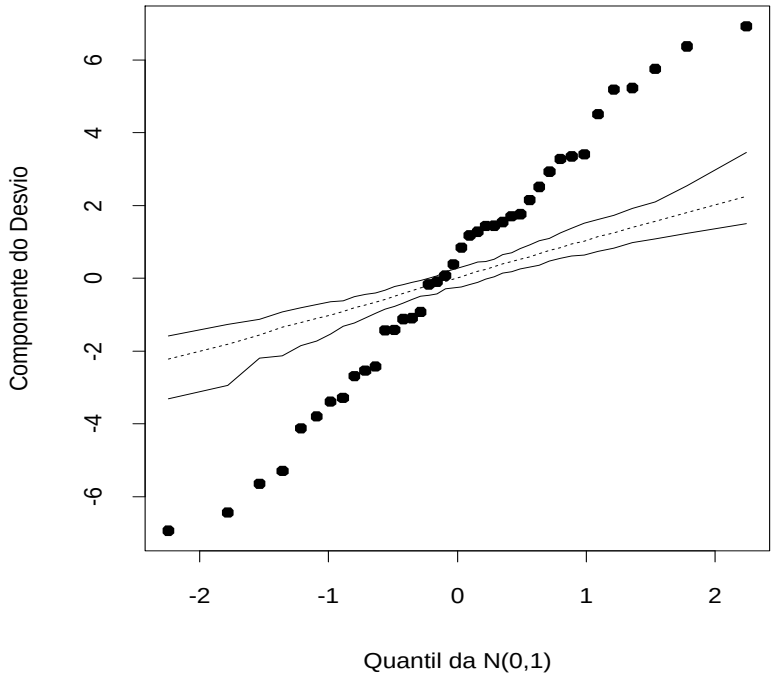

Figura 4.15: Gráfico normal de probabilidades do modelo logístico ajustado aos dados sobre *rotifers*.

Embora essas estimativas sejam altamente significativas, o desvio do modelo $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 434,02$ (36 graus de liberdade) indica para um ajuste inadequado. O gráfico normal de probabilidades descrito na Figura 4.15 confirma a sobredispersão. Segundo Collett (1991, Cap. 6) a sobredispersão nos dados pode ter sido causada por uma possível má distribuição dos animais nos recipientes, uma vez que *rotifers* mais jovens são menos densos do que os mais maduros. Collett (1991) propõe um modelo logístico com efeito aleatório

para ajustar a proporção de animais em suspensão e consegue uma redução substancial no valor do desvio. Alternativamente será assumido o modelo proposto na Seção 4.6.14, que com uma adaptação de notação corresponde a assumir $\mathrm{E}(Y_{ij}) = n_{ij}\pi_{ij}$ e $\mathrm{Var}(Y_{ij}) = n_{ij}\pi_{ij}(1-\pi_{ij})\{1+(n_{ij}-1)\delta\}$, em que δ denota a correlação intraunidade experimental.

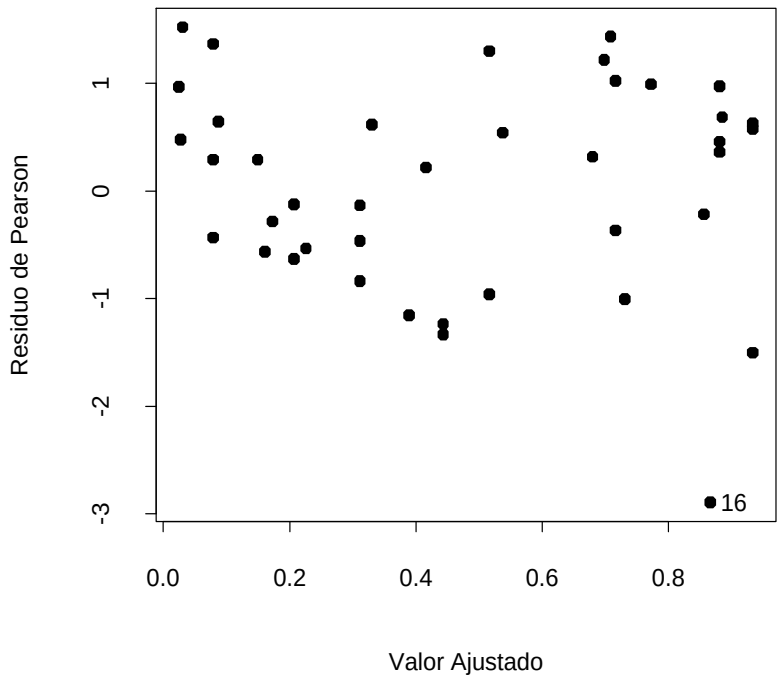

Figura 4.16: Gráfico de resíduos de Pearson contra os valores ajustados para o modelo logístico de sobredispersão ajustado aos dados sobre *rotifers*.

Usando o processo iterativo dado na seção anterior obtém-se as novas estimativas $\hat{\alpha}_1 = -90,64(13,18)$, $\hat{\beta}_1 = 87,22(12,66)$, $\hat{\alpha}_2 = -117,25(14,91)$, $\hat{\beta}_2 = 111,45(14,21)$ e $\hat{\delta} = 0,0815$. Pela Figura 4.16 nota-se que, exceto a observação #16 que corresponde a uma unidade experimental com baixa proporção de *rotifers* (10/22) para uma densidade alta, os demais resíduos

permanecem no intervalo [-2,2] e não apresentam nenhuma tendência sistemática contra os valores ajustados.

A aplicação da estatística ξ_S para testar $\mathrm{H}_0 : \delta = 0$ contra $\mathrm{H}_1 : \delta > 0$ forneceu o valor $\xi_S = 3,126$, com nível descritivo P$= 0,0009$, indicando fortemente pela rejeição da hipótese nula. Portanto, há indícios de sobredispersão nos dados.

4.10 Modelo logístico condicional

Em alguns estudos de caso e controle ou de seguimento o número de estratos formados pode ser relativamente grande. Isso ocorre em particular nos estudos emparelhados de caso e controle, em que a influência de fatores suspeitos de confundimento é controlada através de emparelhamentos de casos com controles, segundo alguns níveis desses fatores. Para cada emparelhamento tem-se um estrato. Assim, se é adotado um modelo logístico linear, além dos parâmetros correspondentes aos efeitos incluídos no modelo, tem-se um parâmetro (intercepto) para cada estrato. Nos casos de estratos com poucas observações, o número de parâmetros pode ser da mesma ordem do número total de observações, que pode levar a estimativas viesadas (ver Cox e Hinkley, 1974, p. 292).

Como ilustração, supor um estudo de caso e controle com k emparelhamentos do tipo 1:1 (1 caso por 1 controle) segundo os níveis de um fator binário de exposição representado pela variável X ($X = 1$ presença da exposição, $X = 0$ ausência da exposição). Denote por $Y_i(x)$ o resultado da resposta para o indivíduo do i-ésimo estrato com $X = x$ ($Y_i(x) = 1$ caso, $Y_i(x) = 0$ controle). Supor que $Y_i(x) \sim \mathrm{Be}\{\pi_i(x)\}$, em que

$$\log\left\{\frac{\pi_i(x)}{1-\pi_i(x)}\right\} = \alpha_i + \beta x.$$

A razão de chances de ser caso entre o indivíduo exposto e o indivíduo não exposto no i-ésimo estrato fica dada por

$$\psi = \frac{\pi_i(1)/\{1-\pi_i(1)\}}{\pi_i(0)/\{1-\pi_i(0)\}} = \exp(\beta)$$

sendo, portanto, constante ao longo dos estratos.

Para eliminar os parâmetros α_i's pode-se trabalhar com a distribuição condicional de $Y_i(1)$ dado $Y_i(1) + Y_i(0) = m$. Essa distribuição foi discutida na Seção 4.2.3. A função de probabilidade pode ser expressa na forma

$$f(a|m;\psi) = \frac{\binom{1}{a}\binom{1}{m-a}\psi^a}{\sum_{t=u}^{v}\binom{1}{t}\binom{1}{m-t}\psi^t},$$

em que $a = 0, 1$ e $m = 0, 1, 2$. É fácil mostrar que $f(a|0;\psi) = f(a|2;\psi) = 1$, havendo portanto informação a respeito de ψ somente nos estratos em que $Y_i(1)+Y_i(0) = 1$. A função de probabilidade nesse caso é definida para $a = 0$ e $a = 1$, sendo as probabilidades dadas por

$$f(0|1;\psi) = 1/(1+\psi)$$

e

$$f(1|1;\psi) = \psi/(1+\psi).$$

Definindo para o i-ésimo estrato duas novas variáveis binárias X_{1i} e X_{2i} representando, respectivamente, o nível de exposição do caso e do controle, é possível expressar as probabilidades condicinais na forma

$$f(a|1,\psi) = \frac{\exp(x_{1i} - x_{2i})\beta}{1+\exp(x_{1i} - x_{2i})\beta},$$

em que $a = 0, 1$. Assim, para k estratos, a função de verossimilhança conjunta condicional, que depende apenas de β e será denotada por $\ell(\beta)$, assume a forma

$$\ell(\beta) \;=\; \Pi_{i=1}^{k}\left[\frac{\exp\{(x_{i1} - x_{i2})\beta\}}{1+\exp\{(x_{i1} - x_{i2})\beta\}}\right].$$

Tem-se que a expressão acima coincide com a função de verossimilhança de uma regressão logística com k sucessos em k ensaios, com uma única covariável com valores observados $z_i = x_{i1} - x_{i2}$, $i = 1, \ldots, k$, e passando pela origem.

Generalizando para p covariáveis e supondo ainda emparelhamentos 1:1, tem-se o modelo

$$\log\left\{\frac{\pi_i(\mathbf{x})}{1-\pi_i(\mathbf{x})}\right\} = \alpha_i + \mathbf{x}^\top\boldsymbol{\beta},$$

em que $\mathbf{x} = (x_1, \ldots, x_p)^\top$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ e $\pi_i(\mathbf{x}) = Pr\{Y_i = 1|\mathbf{x}\}$, $i = 1, \ldots, k$. Observando no i-ésimo estrato os valores $\mathbf{x}_{i1} = (x_{i11}, \ldots, x_{i1p})^\top$ para o caso e os valores $\mathbf{x}_{i2} = (x_{i21}, \ldots, x_{i2p})^\top$ para o controle, a função de verossimilhança conjunta condicional assume a forma geral (ver, po exemplo, Breslow e Day, 1980, p. 205; Hosmer e Lemeshow, 1989, Cap. 7)

$$\ell(\boldsymbol{\beta}) \;=\; \Pi_{i=1}^{k}\left[\frac{\exp\{(\mathbf{x}_{i1}-\mathbf{x}_{i2})^\top\boldsymbol{\beta}\}}{1+\exp\{(\mathbf{x}_{i1}-\mathbf{x}_{i2})^\top\boldsymbol{\beta}\}}\right].$$

Logo, a estimação de $\boldsymbol{\beta}$ pode ser feita através do ajuste de uma regressão logística com k sucessos em k ensaios, com valores observados das covariáveis dados por $z_{ij} = x_{i1j} - x_{i2j}$, $i = 1, \ldots, k$ e $j = 1, \ldots, p$ e passando pela origem. Deve-se observar que embora algumas quantidades da regressão logística condicional para estudos emparelhados do tipo 1:1 coincidam com as quantidades de uma regressão logística não condicional passando pela origem, tais como estimativas dos parâmetros e erros padrão assintóticos, as distribuições dos modelos são diferentes. No primeiro caso tem-se o produto de hipergeométricas independentes, enquanto que no segundo caso tem-se o produto de binomiais independentes. Isso pode refletir na obtenção de alguns resultados, como por exemplo, geração de envelope para o resíduo componente do desvio que usa a distribuição da resposta no processo de geração dos dados.

4.10.1 Técnicas de diagnóstico

Moolgavkar et al.(1985) e Pregibon (1984) têm mostrado que a maioria das técnicas usuais de diagnóstico do modelo logístico não condicional podem ser estendidas para o modelo logístico condicional. Como a variável resposta no modelo logístico condicional sempre assume o valor 1, o resíduo componente do desvio é sempre positivo, sendo dado por

$$t_{D_i} = \frac{\sqrt{2}|\log \hat{\pi}_i|}{\sqrt{1-\hat{h}_{ii}}},$$

em que

$$\hat{\pi}_i = \frac{\exp(\mathbf{z}_i^\top \hat{\boldsymbol{\beta}})}{1+\exp(\mathbf{z}_i^\top \hat{\boldsymbol{\beta}})} \text{ e } \hat{h}_{ii} = \hat{\pi}_i(1-\hat{\pi}_i)\mathbf{z}_i^\top(\mathbf{Z}^\top \hat{\mathbf{V}}\mathbf{Z})^{-1}\mathbf{z}_i.$$

Os gráficos de t_{D_i} e $\hat{h}_{ii}$ contra os valores ajustados $\hat{\pi}_i$ podem revelar emparelhamentos discrepantes com algum tipo de influência nos resultados do modelo.

De forma similar, a distância de Cook no caso emparelhado fica dada por

$$\text{LD}_i = \frac{\hat{h}_{ii}}{(1-\hat{h}_{ii})^2}\hat{r}_{P_i}^2,$$

em que

$$\hat{r}_{P_i} = \frac{1-\hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}}$$

é o resíduo de Pearson. Tem-se que $\hat{r}_{P_1}$ assume sempre valores não negativos. O gráfico de LD_i contra os valores ajustados $\hat{\pi}_i$ pode revelar aqueles emparelhamentos com maior influência nas estimativas dos parâmetros. A geração de envelope, contudo, somente pode ser feita através do modelo logístico condicional.

Para ilustrar o ajuste no `R`, supor um estudo com $k = 20$ emparelhamentos do tipo 1:1 e que foram observados os valores de duas covariáveis $V1$ e $V2$. Os valores observados dos casos serão armazenados nos objetos `v11` e `v12` e

os valores observados dos controles nos objetos `v21` e `v22`. O ajuste segue os seguintes passos:

```
resp < rep(1, times=20)
z1 < v11 - v21
z2 < v12 - v22
fit.cond < glm(resp ~ z1+z2 - 1, family=binomial).
```

Pode-se analisar `fit.cond` em geral da mesma forma que é analisada a saída de um modelo logístico linear. Por exemplo, as estimativas e os erros padrão, como foi mostrado acima, coincidem com as estimativas e os erros padrão obtidos pelo modelo logístico condicional.

4.10.2 Aplicação

Como aplicação será discutido a seguir um estudo cujo objetivo foi avaliar o efeito da obesidade, do histórico familiar e de atividades físicas no desenvolvimento de diabetes não dependente de insulina. 30 indivíduos não diabéticos foram emparelhados com 30 indivíduos diabéticos não dependentes de insulina pela idade e pelo sexo. A obesidade foi medida através do índice de massa coporal (IMC), que é definida como sendo o peso (em kg) dividido pela altura (em metros quadrados). O histórico familiar com diabetes (HF) e as atividades físicas (ATF) foram tratadas como sendo variáveis binárias (HF=1 presença, HF=0 ausência; ATF=1 presença, ATF=0 ausência). Os dados são descritos em Lee (1991, p. 312) e reproduzidos na Tabela 4.22 e estão também no arquivo **diabetes.txt**.

Denote por x_{i11}, x_{i12} e x_{i13}, respectivamente, o valor da massa corporal (IMC), histórico familiar (HF) e atividades físicas (ATF) para o i-ésimo indivíduo diabético e por x_{i21}, x_{i22} e x_{i23} os valores dessas variáveis para o i-ésimo indivíduo não diabético. A função de verossimilhança do modelo

Tabela 4.22
Emparelhamento de 30 diabéticos não dependentes de insulina (casos) e 30 não diabéticos (controles).

	Casos			Controles		
Par	IMC	HF	ATF	IMC	HF	ATF
1	22,1	1	1	26,7	0	1
2	31,3	0	0	24,4	0	1
3	33,8	1	0	29,4	0	0
4	33,7	1	1	26,0	0	0
5	23,1	1	1	24,2	1	0
6	26,8	1	0	29,7	0	0
7	32,3	1	0	30,2	0	1
8	31,4	1	0	23,4	0	1
9	37,6	1	0	42,4	0	0
10	32,4	1	0	25,8	0	0
11	29,1	0	1	39,8	0	1
12	28,6	0	1	31,6	0	0
13	35,9	0	0	21,8	1	1
14	30,4	0	0	24,2	0	1
15	39,8	0	0	27,8	1	1
16	43,3	1	0	37,5	1	1
17	32,5	0	0	27,9	1	1
18	28,7	0	1	25,3	1	0
19	30,3	0	0	31,3	0	1
20	32,5	1	0	34,5	1	1
21	32,5	1	0	25,4	0	1
22	21,6	1	1	27,0	1	1
23	24,4	0	1	31,1	0	0
24	46,7	1	0	27,3	0	1
25	28,6	1	1	24,0	0	0
26	29,7	0	0	33,5	0	0
27	29,6	0	1	20,7	0	0
28	22,8	0	0	29,2	1	1
29	34,8	1	0	30,0	0	1
30	37,3	1	0	26,5	0	0

logístico condicional será dada por

$$\ell(\boldsymbol{\beta}) = \Pi_{i=1}^{30} \left\{ \frac{\exp(z_{i1}\beta_1 + z_{i2}\beta_2 + z_{i3}\beta_3)}{1 + \exp(z_{i1}\beta_1 + z_{i2}\beta_2 + z_{i3}\beta_3)} \right\},$$

em que $z_{i1} = x_{i11} - x_{i21}$, $z_{i2} = x_{i12} - x_{i22}$ e $z_{i3} = x_{i13} - x_{i23}$.

As estimativas de máxima verossimilhança (erro padrão aproximado) são dadas por $\hat{\beta}_1 = 0,090(0,065)$, $\hat{\beta}_2 = 0,968(0,588)$ e $\hat{\beta}_3 = -0,563(0,541)$, cujos níveis descritivos são, respectivamente, dados por $0,166$, $0,099$ e $0,298$, indicando indícios de efeito significativo ao nível de 10% apenas para o histórico familiar. Ou seja, indivíduos com histórico familiar de diabetes têm chance maior de desenvolvimento de diabetes dependente de insulina com relação a indivíduos sem histótico familiar de diabetes.

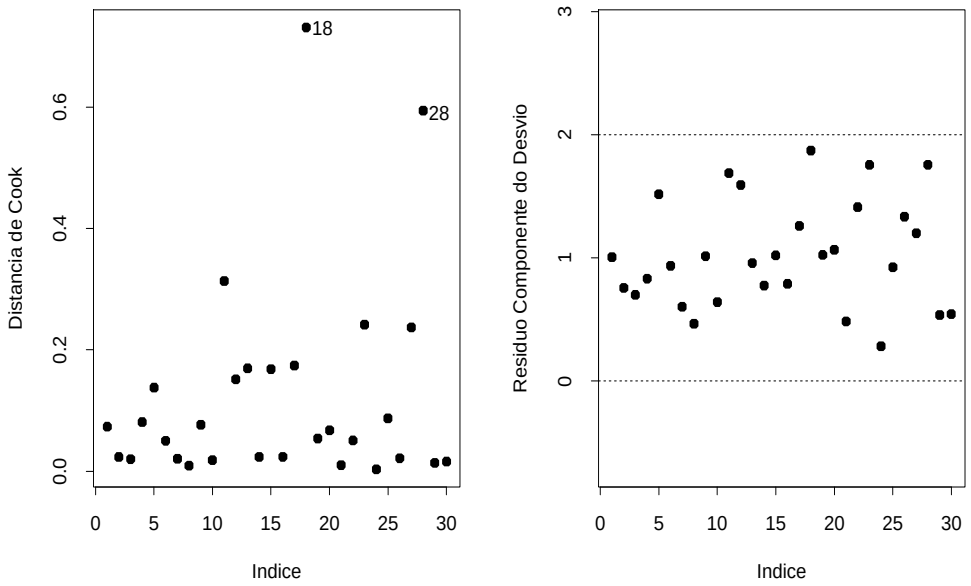

Figura 4.17: Gráficos de diagnóstico para o modelo logístico condicional aplicado aos dados sobre diabetes.

Na Figura 4.17 são apresentados alguns gráficos de diagnóstico em que

pode-se notar destaque para as observações #18 e #28 como possivelmente influentes nas estimativas dos parâmetros. A eliminação do emparelhamento #18 não muda os resultados inferenciais embora aumente a significância do histórico familiar. Já a eliminação do emparelhamento #28 muda os resultados inferenciais uma vez que o índice de massa corporal passa a ser significante ao nível de 10%. Nesse emparelhamento o controle tem histórico familiar e atividade física enquanto o caso não apresenta as duas características. Além disso, o caso tem um índice de massa corporal menor do que o controle.

4.10.3 Emparelhamento 1:M

Para emparelhamentos do tipo 1:M (M $\geq$ 2) e k estratos a função de verossimilhança (ver, por exemplo, Breslow e Day, 1980) para $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ fica dada por

$$\ell(\boldsymbol{\beta}) = \Pi_{i=1}^{k} \{\exp(\mathbf{x}_{i0}^\top \boldsymbol{\beta}) / \sum_{\ell=0}^{M} \exp(\mathbf{x}_{i\ell}^\top \boldsymbol{\beta})\}, \tag{4.20}$$

cujo logaritmo assume a forma

$$\mathrm{L}(\boldsymbol{\beta}) = \log \ell(\boldsymbol{\beta}) = \sum_{i=1}^{k} [\mathbf{x}_{i0}^\top \boldsymbol{\beta} - \log\{\sum_{\ell=0}^{M} \exp(\mathbf{x}_{i\ell}^\top \boldsymbol{\beta})\}], \tag{4.21}$$

em que $\mathbf{x}_{i0} = (x_{i01}, \ldots, x_{i0p})^\top$ denota os valores observados para o caso e $\mathbf{x}_{i\ell} = (x_{i\ell 1}, \ldots, x_{i\ell p})^\top$ denota os valores observados para o ℓ-ésimo controle.

A função de verossimilhança (4.21) coincide com a função de verossimilhança do modelo de regressão de Cox (Cox, 1972; Cox e Oakes, 1974) quando não há ocorrência de empates. Isso permite que os modelos logísticos condicionais para emparelhamentos 1:M ($M \geq 2$) sejam ajustados através de programas desenvolvidos para o modelo de Cox.

4.11 Exercícios

1. Supor a seguinte tabela de contingência 2×2:

	Fator		
Doença	A	B	
D	y_1	y_2	
$\bar{D}$	y_3	y_4	
			n

e que a amostragem foi realizada segundo distribuição multinomial, isto é, a função de probabilidade de $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)^\top$ é dada por

$$P(\mathbf{Y} = \mathbf{y}) = \frac{n!}{y_1!y_2!y_3!y_4!}\pi_1^{y_1}\pi_2^{y_2}\pi_3^{y_3}\pi_4^{y_4},$$

com $\mathbf{y} = (y_1, y_2, y_3, y_4)^\top$, $\sum_{i=1}^4 y_i = n$, $0 < \pi_i < 1$ e $\sum_{i=1}^4 \pi_i = 1$. Sabe-se que $\mathrm{E}(Y_i) = n\pi_i$, $\mathrm{Var}(Y_i) = n\pi_i(1-\pi_i)$ e $\mathrm{cov}(Y_i, Y_j) = -n\pi_i\pi_j$, para $i \neq j$. Mostre que as estimativas de máxima verossimilhança são dadas por $\hat{\pi}_i = \frac{y_i}{n}$, $i = 1, 2, 3, 4$, com $\mathrm{E}(\hat{\pi}_i) = \pi_i$, variâncias e covariâncias $\mathrm{Var}(\hat{\pi}_i) = \frac{\hat{\pi}_i(1-\hat{\pi}_i)}{n}$ e $\mathrm{cov}(\hat{\pi}_i, \hat{\pi}_j) = -\frac{\pi_i\pi_j}{n}$, para $i \neq j$.

2. A razão de chances sob amostragem multinomial é definida por $\psi = \pi_1\pi_4/\pi_2\pi_3$. Considere $\log(\widehat{\psi})$ e mostre, usando o método delta, que a variância assintótica de $\log(\widehat{\psi})$ fica dada por $\mathrm{Var}\{\log(\widehat{\psi})\} = [1/n\pi_1 + 1/n\pi_2 + 1/n\pi_3 + 1/n\pi_4]$. Lembre que a variância assintótica pode ser obtida através da expressão

$$\mathrm{V}ar\{\log(\widehat{\psi})\} = \left[\frac{\partial \log(\psi)}{\partial \boldsymbol{\pi}}\right]^\top \mathrm{V}ar(\hat{\boldsymbol{\pi}}) \left[\frac{\partial \log(\psi)}{\partial \boldsymbol{\pi}}\right],$$

em que $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4)^\top$ e $\mathrm{Var}(\hat{\boldsymbol{\pi}})$ denota a matriz de variâncias e covariâncias assintóticas de $\hat{\boldsymbol{\pi}}$. Neste caso as variâncias e covariâncias assintóticas são as mesmas obtidas em (1).

3. A tabela abaixo resume um estudo de caso e controle em que foram considerados como casos 200 homens adultos diagnosticados com câncer de esôfago num hospital de uma determinada comunidade. Os controles foram uma amostra de 775 homens adultos escolhidos aleatoriamente da lista de eleitores da comunidade. Esses dois grupos foram classificados segundo os níveis alto (mais de 80g/dia) e baixo (até 80g/dia) do fator exposição ao alcool.

	Alto	Baixo	Total
Caso	96	104	200
Controle	109	666	775
Total	205	770	975

Verifique, através de um teste apropriado, se há associação entre o fator de exposição e a doença. Encontre uma estimativa intervalar de 95% para a razão de chances. Indique as suposições utilizadas e interprete os resultados.

4. Considere a tabela 2×2 descrita abaixo.

	Fator		
Doença	A	B	Total
D	3	7	10
$\bar{D}$	6	9	15
Total	9	16	25

Aplicar o teste exato de Fisher para testar $H_0 : \psi = 1$ contra $H_1 : \psi \neq 1$. Obter inicialmente as probabilidades condicionais usando a distribuição hipergeométrica central correspondente. Comparar com o teste assintótico.

5. Os dados abaixo são provenientes de um estudo de seguimento cujo objetivo foi avaliar a associação de duas técnicas cirúrgicas, A e B, e a ocorrência de problemas graves pós-operatórios segundo duas faixas de idade.

	Faixa I		Faixa II	
Problema	A	B	A	B
Sim	6	7	7	4
Não	14	23	9	12

Obter estimativa intervalar de 95% para a razão de chances em cada estrato. Teste a hipótese de homogeneidade das razões de chances. Se a hipótese nula não for rejeitada ao nível de 5%, aplicar o teste de Mantel-Haenszel (com e sem correção para continuidade) para testar ausência de associação entre técnica cirúrgica e ocorrência de problemas graves pós-operatórios.

6. Suponha $Y_{ij} \sim \mathrm{B}(n_{ij}, \pi_{ij})$ mutuamente independentes, $i, j = 1, 2$ com as probabilidades π_{ij} sendo definidas por

$$\log\left\{\frac{\pi_{i1}}{1-\pi_{i1}}\right\} = \alpha_i - \Delta \quad \text{e} \quad \log\left\{\frac{\pi_{i2}}{1-\pi_{i2}}\right\} = \alpha_i + \Delta.$$

Interprete α_1, α_2 e Δ. Mostre que o teste de escore para testar $\mathrm{H}_0 : \Delta = 0$ contra $\mathrm{H}_1 : \Delta \neq 0$, coincide com o teste de Mantel-Hanszel (X^2_{MH}) para testar $\mathrm{H}_0 : \psi = 1$ contra $\mathrm{H}_1 : \psi \neq 1$, em que $\psi = \pi_{i2}(1-\pi_{i1})/\pi_{i1}(1-\pi_{i2})$, $i = 1, 2$ (Day e Byar, 1979).

7. Supor um modelo binomial quadrático de dose-resposta, em que $Y_i \overset{\text{ind}}{\sim} \mathrm{B}(n_i, \pi_i)$, com $g(\pi_i) = \alpha + \beta x_i + \gamma x_i^2$, $i = 1, \ldots, k$. Como fica expressa a estimativa $\widehat{\mathrm{DL}}_{100p}$? E a variância assintótica de $\widehat{\mathrm{DL}}_{100p}$?

8. Supor um modelo binomial quadrático de dose-resposta, em que $Y_i \overset{\text{ind}}{\sim} \text{B}(n_i, \pi_i)$, com $g(\pi_i) = \eta_i = \alpha + \beta x_i + \gamma x_i^2$, $i = 1, \ldots, k$. Assumir que $\partial^2\eta/\partial x^2 < 0$, $\forall x$. Qual a solução para $\partial\eta/\partial x = 0$? Denotando essa solução por x_0, interprete e encontre uma estimativa intervalar para x_0.

9. Os conjuntos de dados apresentados nos arquivos **dose1.txt**, **dose2.txt** e **dose3.txt** (Paula et al.,1988) são provenientes de um experimento de dose-resposta conduzido para avaliar a influência dos extratos vegetais "aquoso frio de folhas", "aquoso frio de frutos"e de um extrato químico, respectivamente, na morte de um determinado tipo de caramujo. Para cada conjunto, ajuste um modelo logístico linear simples e um modelo complementar log-log linear simples. Para o melhor ajuste (use envelopes como critério), encontre um intervalo assintótico de 95% para a dose letal DL_{50}, construa as bandas de confiança e verifique se há indícios de sobredispersão aplicando um teste apropriado.

10. Os dados abaixo (Collett, 1991, p.127) são provenientes de um experimento desenvolvido para avaliar a germinação de um determinado tipo de semente segundo três condições experimentais: nível da temperatura (21^oC, 42^oC e 62^oC); nível da umidade (baixo, médio e alto) e temperatura da germinação (11^oC e 21^oC). A tabela abaixo apresenta o número de sementes que germinaram após cinco dias para cada 100 sementes submetidas a cada condição experimental.

    Assuma um modelo logístico para explicar o número de sementes que germinaram. Aplique o método `AIC` para selecionar um modelo considerando interações de $1^{\underline{a}}$ ordem. Interprete os resultados. Faça uma análise de resíduos com o modelo selecionado. Esses dados estão descritos no arquivo **sementes.txt**.

Temperatura da Germinação	Nível da Umidade	Nível da Temperatura		
		21^oC	42^oC	62^oC
11^oC	baixo	98	96	62
11^oC	médio	94	79	3
11^oC	alto	92	41	1
21^oC	baixo	94	93	65
21^oC	médio	94	71	2
21^oC	alto	91	30	1

11. Mostre que a variância assintótica do estimador de máxima verossimilhança não condicional da razão de chances numa tabela 2×2 é dada por

$$\mathrm{Var_A}(\tilde{\psi}) = \psi^2 \left\{ \frac{1}{n_1\pi_1(1-\pi_1)} + \frac{1}{n_2\pi_2(1-\pi_2)} \right\}.$$

Lembre que: sob condições gerais de regularidade, os estimadores de máxima verossimilhança são assintoticamente normais e não viesados com variância assintótica igual à inversa da matriz de informação de Fisher.

12. A tabela abaixo descreve o resultado de um experimento em que vários pacientes foram submetidos a um de quatro níveis de exposição de um tratamento particular e foi observado, após 12 meses, se o paciente foi curado ou não curado.

	Nível de Exposição			
Resultado	E1	E2	E3	E4
Curado	20	16	12	5
Não-Curado	80	84	48	20

Seja Y_i o número de pacientes curados dentre os n_i submetidos ao nível de exposição Ei. Supor que $Y_i \sim \mathrm{B}(n_i, \pi_i)$, $i = 1, \ldots, 4$. Assunir o nível E1 como nível de referência e teste a hipótese de homogeneidade das razões de chances contra a alternativa de razões de chances diferentes.

13. Sejam Y_1 e Y_2 variáveis aleatórias independentes tais que $Y_1 \sim \mathrm{B}(n_1, \pi_1)$ e $Y_2 \sim \mathrm{B}(n_2, \pi_2)$. Seja $\mathrm{RR} = \pi_1/\pi_2$ o risco relativo. Aplique o método delta para obter a variância assintótica de $\widehat{\mathrm{RR}}$. Desenvolva o teste da Wald para testar $\mathrm{H}_0 : \mathrm{RR} = 1$ contra $\mathrm{H}_1 : \mathrm{RR} \neq 1$. Qual a distribuição nula assintótica do teste?

14. Sejam $Y_1, \ldots, Y_n$ variáveis aleatórias mutuamente independentes tais que $Y_i \sim \mathrm{B}(m, \pi_i)$, em que $\log\{\pi_i/(1-\pi_i)\} = \alpha$. (i) Encontre a estimativa de máxima verossimilhança de α. (ii) Calcule $\mathrm{Var}(\hat{\alpha})$. (iii) Como fica o teste da razão de verossimilhanças para testar $\mathrm{H}_0 : \alpha = 0$ versus $\mathrm{H}_1 : \alpha \neq 0$? Qual a distribuição nula assintótica da estatística do teste?

15. Supor $Y_{ij} \overset{\text{ind}}{\sim} \mathrm{Be}(\mu_i)$, $0 < \mu_i < 1$, para $i = 1, 2$ e $j = 1, 2, \ldots, r$, em que $\mathrm{arcsen}(\sqrt{\mu_1}) = \alpha + \Delta$ e $\mathrm{arcsen}(\sqrt{\mu_2}) = \alpha - \Delta$. Como fica a matriz $\mathbf{X}$? Obter $\hat{\alpha}$ e $\hat{\Delta}$ e as respectivas variâncias assintóticas, além de $\mathrm{Cov}(\hat{\alpha}, \hat{\Delta})$. Obter a estatística do teste de escore para testar $\mathrm{H}_0 : \Delta = 0$ contra $\mathrm{H}_1 : \Delta \neq 0$. Qual a distribuição nula assintótica da estatística do teste? Use o resultado: $\frac{d}{dx}\mathrm{arcsen}\{u(x)\} = \frac{1}{\sqrt{1-u^2}}\frac{du}{dx}$.

16. Considere uma aplicação de regressão logística em análise de sobrevivência (Lawless, 1982, p.389; Efron, 1988). Seja $\pi_i(t)$ a probabilidade de um equipamento do tipo i falhar no intervalo $I_t = (t-1, t]$ dado que o mesmo não falhou até o tempo $t-1$. Seja Y_{it} o número de falhas no intervalo I_t e seja n_{it} o número de equipamentos que não falharam até o tempo $t-1$ no i-ésimo grupo. Assumir que $Y_{it} \sim \mathrm{B}(n_{it}, \pi_i(t))$ e que as falhas são independentes. Ajustar um modelo logístico do tipo

$$\log\left\{\frac{\pi_i(t)}{1-\pi_i(t)}\right\} = \alpha_i + \beta_i t + \gamma_i t^2 \tag{4.22}$$

ao seguinte conjunto de dados:

	Tipo A		Tipo B		Tipo C	
Tempo	n_{1t}	y_{1t}	n_{2t}	y_{2t}	n_{3t}	y_{3t}
1	42	4	50	6	48	11
2	38	3	44	11	37	10
3	35	3	32	10	27	12
4	31	5	22	8	15	8
5	26	6	12	6	6	4

Apresente o gráfico com as curvas ajustadas e os valores observados. Tente selecionar um submodelo apropriado. Verifique a adequação do modelo adotado através de gráficos de resíduos. Interprete os resultados. Os dados estão descritos no arquivo **equipamentos.txt**.

17. No arquivo **matched.txt** (Hosmer e Lemeshow, 1989, Cap.7) estão os dados de um estudo de caso-controle com emparelhamentos do tipo 1:1, em que os casos foram mulheres com diagnóstico confirmado de tumor benigno na mama e os controles de mulheres sadias diagnosticadas no mesmo hospital e período dos casos. A variável de emparelhamento foi a idade da paciente na época da entrevista `AGMT`. Escolha três variáveis do arquivo mencionado e verifique através de uma regressão logística condicional a associação entre as variáveis escolhidas e o diagnóstico da doença (sim=1, não=0) representado pela variável `FNDX`. Interprete as estimativas dos parâmetros do modelo ajustado. Faça uma análise de diagnóstico. Obsevação: caso você escolha alguma variável com observações perdidas, exclua das análises as pacientes correspondentes.

18. Considere uma aplicação de regressão logística em transportes. Seja $\pi_i(t)$ a probabilidade de um caminhão do tipo i ser desativado durante o ano t dado que o mesmo não foi desativado durante o ano $t-1$.

Assuma que durante o ano t foram desativados y_{it} caminhões dentre os n_{it} existentes no começo do ano, $i = 1, 2$ e $t = 1, \ldots, k$. Supor que $Y_{it} \sim \mathrm{B}(n_{it}, \pi_i(t))$ e que são mutuamente independentes. Considere o modelo

$$\log\left\{\frac{\pi_1(t)}{1-\pi_1(t)}\right\} = \gamma_t \quad \text{e} \quad \log\left\{\frac{\pi_2(t)}{1-\pi_2(t)}\right\} = \gamma_t + \beta.$$

O que significa testar $\mathrm{H}_0 : \beta = 0$? Qual é a matriz X do modelo? Como fica $\mathrm{Var}(\hat{\beta})$? Mostre que a estatística do teste de escore para testar $\mathrm{H}_0 : \beta = 0$ contra $\mathrm{H}_1 : \beta \neq 0$ pode ser expressa na forma

$$\xi_{SR} = \left\{\sum_{t=1}^{k}\left(y_{2t} - \frac{y_t n_{2t}}{n_t}\right)\right\}^2 \Big/ \sum_{t=1}^{k}\frac{y_t n_{1t} n_{2t}(n_t - y_t)}{n_t^3},$$

em que $n_t = n_{1t} + n_{2t}$ e $y_t = y_{1t} + y_{2t}$. Qual é a distribuição nula assintótica da estatística do teste?

19. Sejam $Y_1, \ldots, Y_k$ variáveis aleatórias independentes tais que a função de probabilidade de Y_i seja dada por

$$f(y_i; \psi_i) = \frac{\binom{1}{y_i}\binom{1}{1-y_i}\psi_i^{y_i}}{\sum_{t=0}^{1}\binom{1}{t}\binom{1}{1-t}\psi_i^t},$$

em que $y_i = 0, 1$. Supor a parte sistemática $\log(\psi_i) = \beta$. (i) Encontre a estimativa de máxima verossimilhança de β; (ii) encontre a informação de Fisher para β; (iii) como fica o teste de escore para testar $\mathrm{H}_0 : \beta = 0$ contra $\mathrm{H}_1 : \beta \neq 0$? Qual a distribuição nula assintótica da estatística do teste? (iv) Expresse o resíduo $r_i = (y_i - \hat{\mu}_i)/\sqrt{\hat{\mathrm{V}}\mathrm{ar}(Y_i)}$ em função de y_i e $\hat{\beta}$; (v) Como você faria para gerar valores de Y_i da distribuição dada acima?

20. Construir o logaritmo da função de verossimilhança de um modelo logístico $\log\{\pi(x)/(1-\pi(x))\} = \alpha + \beta x$ de duas binomiais independentes, em que tem-se y_0 sucessos em n_0 ensaios para $x = 0$ e y_1 sucessos

em n_1 ensaios para $x = 1$. Mostre que a estimativa de máxima verossimilhança $\hat{\beta}$ coincide com log(razão de chances) (Agresti, 2013, p.205).

21. Os dados do arquivo **leuce.txt** (Everitt, 1994) referem-se a um estudo com 51 pacientes adultos, previamente diagnosticados com um tipo agudo de leucemia, que receberam um tipo de tratamento e foi verificado após um certo período a eficiência ou não do tratamento. Algumas variáveis explicativas pré-tratamento foram também observadas. As variáveis em estudo são as seguintes: (i) idade do paciente na época do diagnóstico (em anos), (ii) mancha diferencial da doença (em %), (iii) infiltração na medula (em %), (iv) células com leucemia na medula (em %), (v) malignidade da doença ($\times 10^3$), (vi) temperatura máxima antes do tratamento ($\times 10^o F$), (vii) tratamento (1: satisfatório, 0: não satisfatório), (viii) tempo de sobrevivência após o diagnóstico (em meses) e (ix) situação (1: sobrevivente, 0: não sobrevivente). Considere um modelo logístico linear para explicar a probabilidade de eficiência do tratamento dadas as seis variáveis explicativas. Selecionar as variáveis explicativas bem como as interações de primeira ordem através do método *stepwise*. Usar $\mathrm{P}_E = \mathrm{P}_S = 0,20$. Fazer uma análise de diagnóstico com o modelo selecionado e interpretar algumas razões de chances.

22. No arquivo **hearthd.txt** (Hosmer et al., 2013, Cap.1) são descritos os dados de $n = 100$ pacientes com ausência (HD=0) e evidência (HD=1) de doença arterial coronariana, além da idade (Age) do paciente e a faixa etária (FE). Para ler os dados use o comando

    ```
    heart = read.table(‘‘heart.txt’’, header=TRUE)
    ```

    Fazer uma análise descritiva dos dados, por exemplo boxplots robustos

da idade para cada um dos grupos, comente. Construa uma tabela de contigência com as frequências relativas de pacientes com evidência e ausência da doença segundo as faixas etárias, comente. Ajustar um modelo logístico para explicar a probabilidade Pr(HD=1) dado Age. Comente as estimativas. Fazer uma análise de diagnóstico como gráfico de resíduos e distância de Cook. Avalie o impacto das observações destacadas como possivelmente influentes. Construa uma banda de confiança de 95% para Pr(HD=1) dado Age. Encontre uma estimativa intervalar de 95% para a razão de chances entre um paciente com Age+1 e um paciente com Age ter presença da doença. Construa a curva ROC e estabeleça um critério para classificar pacientes como suspeitos de terem presença da doença. Para esse critério obter as taxas de positivo positivo e de falso positivo. Ajustar o modelo pelo GAMLSS através dos comandos

```
y.heart = cbind(HD, 1-HD)

ajuste = gamlss(y.heart ~ Age, family=BI)

plot(ajuste)

rqres.plot(ajuste, howmany=8, ylim.all=1)

rqres.plot(ajuste, howmany=40, plot="all")
```

Comente os gráficos de resíduos.

23. Cinquenta e quatro indivíduos considerados idosos são submetidos a um exame psiquiátrico para avaliar a ocorrência ou não de sintoma de caduquice (Agresti, 1990, pgs. 122-123). Acredita-se que o escore obtido num exame psicológico feito previamente esteja associado com a ocorrência ou não do sintoma. Os dados são apresentados abaixo (`score`: escala no exame psicológico e `resp`: ocorrência (resp=1) ou

não ocorrência (resp=0) do sintoma). Esses dados estão descritos no arquivo **caduquice.txt**.

Score	Resp	Score	Resp	Score	Resp	Score	Resp	Score	Resp
9	1	7	1	7	0	17	0	13	0
13	1	5	1	16	0	14	0	13	0
6	1	14	1	9	0	19	0	9	0
8	1	13	0	9	0	9	0	15	0
10	1	16	0	11	0	11	0	10	0
4	1	10	0	13	0	14	0	11	0
14	1	12	0	15	0	10	0	12	0
8	1	11	0	13	0	16	0	4	0
11	1	14	0	10	0	10	0	14	0
7	1	15	0	11	0	16	0	20	0
9	1	18	0	6	0	14	0		

Ajustar um modelo logístico para explicar a probabilidade de ocorrência do sintoma em função do escore. Interpretar os resultados. Calcule a estatística de Hosmer-Lemeshow. Faça uma análise de diagnóstico com o modelo ajustado.

24. No arquivo **grahani.txt** (McCullagh e Nelder, 1989, pgs. 128-135) estão os dados referentes à distribuição de duas espécies de lagarto (grahani e opalinus) segundo quatro fatores: (i) período do dia (manhã, meio-dia, tarde), (ii) comprimento da madeira (curta, comprida), (iii) largura da madeira (estreita, larga) e (iv) local de ocupação (claro, escuro). Supor que o número de lagartos encontrados da espécie grahani tenha distribuição binomial. Responda às seguintes questões: (i) proponha um modelo logístico (sem interação) para explicar a proporção de lagartos da espécie grahani. Ajuste o modelo e verifique através do teste da razão de verossimilhanças quais efeitos são significativos ao

nível de 10%. (ii) Verifique separadamente se cada interação de primeira ordem pode ser incluída no modelo ao nível de 5%. Construa o `ANODEV`.(iii) Interprete os resultados tentando falar de uma forma não técnica sobre as preferências dos dois tipos de lagarto.

25. Em um estudo para investigar a incidência de dengue numa determinada cidade da costa mexicana (Neter et el., 1996, pgs. 582-584), um total de 196 indivíduos, escolhidos aleatoriamente em dois setores da cidade, respondeu às seguintes perguntas: (i) `idade`, idade do entrevistado (em anos), (ii) `nivel`, nível sócio-econômico (`nivel=1`, nível alto; `nivel=2`, nível médio; `nivel=3`, nível baixo) e (iii) `setor`, setor da cidade onde mora o entrevistado (`setor=1`, setor 1; `setor=2`, setor 2) e (iv) `caso`, se o entrevistado contraiu (`caso=1`) ou não (`caso=0`) a doença recentemente. Um dos objetivos do estudo é tentar prever ou explicar a probabilidade de um indivíduo contrair a doença dadas as variáveis explicativas `idade`, `nivel` e `setor`. Os dados estão descritos no arquivo **dengue.txt**.

    Inicialmene fazer uma análise descritiva dos dados com tabelas de contingência e boxplots. Em seguida, selecionar um submodelo no R através do método `stepAIC` considerando interações até 1ª ordem. Para o modelo selecionado construir o envelope para o resíduo componente do desvio e o gráficos da distância de Cook. Comentar. Interpretar os coeficientes do modelo ajustado através de razões de chances. Repetir o ajuste no GAMLLSS. Construir os gráficos de resíduos e interpretar o `term.plot`.

    Para ler o arquivo no R

    ```
    dengue = read.table("dengue.txt", header=TRUE)
    ```

```
summary(dengue)

attach(dengue)
```

Ajuste no R

```
nivel = factor(nivel)

setor = factor(setor)

table(caso,nivel)

table(caso,setor)

boxplot(split(idade,caso), xlab="Dengue", ylab="Idade", names
= c("Não","Sim"))

fit1.dengue = glm(caso ~ idade + nivel + setor + idade*nivel
+ idade*setor + nivel*setor, family = binomial)

require(MASS)

fit2.dengue = stepAIC(fit1.dengue)

fit.model = fit2.dengue

source("envel_bino.txt")

source("diag_cook_bino.txt")
```

Ajuste no GAMLSS

```
require(gamlss)

fit3.dengue = gamlss(caso ~ idade + setor, family=BI)

plot(fit3.dengue)

rqres.plot(fit3.dengue, howmany=8, type="wp")

term.plot(fit3.dengue, pages=1).
```

26. No arquivo **olhos.txt** (McCullagh e Nelder, 1989, p.144) são apresentados dados referentes a 78 famílias com pelo menos seis filhos cada uma. Na primeira coluna tem-se a classificação dos olhos dos pais segundo a cor (1: ambos claros, 2: ambos castanhos, 3: ambos escuros, 4: claro e castanho, 5: claro e escuro e 6: castanho e escuro), na segunda coluna a classificação dos olhos dos avós segundo a cor (1: todos claros, 2: todos castanhos, 3: todos escuros, 4: três claros e um castanho, 5: três claros e um escuro, 6: um claro e três castanhos, 7: um escuro e três castanhos, 8: um claro e três escuros, 9: um castanho e três escuros, 10: dois claros e dois castanhos, 11: dois claros e dois escuros, 12: dois castanhos e dois escuros, 13: dois claros, um castanho e um escuro, 14: um claro, dois castanhos e um escuro e 15: um claro, um castanho e dois escuros), na terceira coluna tem-se o número de filhos na família e na última coluna o número de filhos com olhos claros. Seja Y_i o número de filhos com olhos claros pertencentes à i-ésima família. Assuma inicialmente que $Y_i \sim \mathrm{B}(n_i, \pi_i)$, $i = 1, \ldots, 78$. Responda às seguintes questões:

    (i) Ajustar inicialmente um modelo logístico linear apenas com o fator 'cor dos olhos dos pais'. Construir gráficos de resíduos. Identificar os pontos aberrantes. Quais as mudanças nos resultados com a eliminação desses pontos. Há indícios de sobredispersão? Ajustar um modelo de quase-verossimilhança com e sem os pontos aberrantes. Comente.

    (ii) Incluir agora o fator cor dos olhos dos avós. Refazer todos os passos acima. Comente os resultados.

27. A tabela abaixo (Morgan, 1992, p.90) descreve os resultados de um

experimento em que a toxicidade de três concentrações (R-rotenine, D-deguelin e M-mistura, essa última como uma mistura das duas primeiras) é investigada. As concentrações foram testadas em insetos e observado para cada dose o número de insetos mortos. Os dados estão descritos no arquivo **morgan.txt**.

Concentração	Dose	Expostos	Mortos
R	0,41	50	6
R	0,58	48	16
R	0,71	46	24
R	0,89	49	42
R	1,01	50	44
D	0,71	49	16
D	1,00	48	18
D	1,31	48	34
D	1,48	49	47
D	1,61	50	47
D	1,70	48	48
M	0,40	47	7
M	0,71	46	22
M	1,00	46	27
M	1,18	48	38
M	1,31	46	43
M	1,40	50	48

Supor inicialmente o modelo $\log\{\pi_i(x)/(1-\pi_i(x))\} = \alpha_i + \beta_i x$, $i = 1, 2, 3$, em que $\pi_i(x)$ é a proporção esperada de insetos mortos sob a concentração i e dose x. Faça uma análise de diagnóstico e verifique se há indícios de sobredispersão aplicando um teste apropriado. Teste a hipótese de paralelismo com todos os pontos e sem as observações discrepantes. Comente.

28. No arquivo **pulso.txt** são descritas as variáveis pulsação em repouso (1: normal, 2: alta), hábito de fumar (1: sim, 2: não) e peso (em kg) de 92 adultos do sexo masculino. Ajustar um modelo logístico linear para explicar a probabilidade de pulsação alta dadas as demais variáveis. Faça uma análise de diagnóstico. Apresente as curvas ajustadas para cada grupo de hábito de fumar com as respectivas bandas de confiança de 95%.

29. Na tabela abaixo (Agresti, 2013, Seção 6.4) é apresentado um resumo dos resultados de um ensaio clínico em que dois tipos de pomada (droga ativa e controle) para um tipo de infecção foram aplicadas em voluntários de 4 centros clínicos. O objetivo do estudo é avaliar se há associação entre tratamento e cura da infecção nos 4 centros clinicos.

Centro	Tratamento	Sucessos	Fracassos	Total
1	Droga	11	25	36
	Controle	10	27	37
2	Droga	16	4	20
	Controle	22	10	32
3	Droga	14	5	19
	Controle	7	12	19
4	Droga	2	14	16
	Controle	1	16	17

Fazer inicialmente uma análise descritiva dos dados calculando para cada centro a razão de chances e a estimativa intervalar de 95%. Em seguida, ajustar no R um modelo logístico binomial com os efeitos e interação entre tratamento e centro. Testar a ausência de interação. Selecionar um submodelo, apresentar estimativas intervalares para a razão de chances e fazer análise de resíduos. Ajustar o modelo selecio-

nado no GAMLSS, fazer análise de resíduos e obter o `term.plot`. Os dados estão no arquivo **creme.txt**

Para ler o arquivo no R

```
creme = read.table("creme.txt", header=TRUE)
summary(creme)
attach(creme)
```

Ajuste no R

```
yresp = cbind(sucessos,fracassos)
fit1.creme = glm(yresp ~ tratamento + centro + tratamento*creme,
family = binomial)
summary(fit1.creme)
fit2.creme = glm(yresp ~ tratamento + centro, family = binomial)
summary(fit2.creme)
anova(fit2.creme, fit1.creme)
```

Análise de resíduos no R

```
fit.model = fit2.creme
ntot=total
source("envelr_bino.txt")
```

Ajuste no GAMLSS

```
require(gamlss)
fit3.creme = gamlss(yresp ~ tratamento + centro, family = BI)
summary(fit3.creme)
plot(fit3.creme)
```

```
rqres.plot(fit3.creme, howmany=8, type="wp")

term.plot(fit3.creme, terms=1)

term.plot(fit3.creme, terms=2).
```

30. Os dados a serem analisados e disponíveis no arquivo `PimaIndiansDiabetes2` da biblioteca `mlbench` do R, referem-se a uma amostra de $n = 768$ mulheres de pelo menos 21 anos de idade descendentes do povo nativo `Pima` dos Estados Unidos que viviam às margens dos rios Gila e Sal, na parte sul do estado do Arizona. A variável resposta de principal interesse é o diagnóstico de diabetes (1:positivo, 0:negativo), que será relacionada com as seguintes variáveis explicativas:

    - pregnant: número de vezes que engravidou
    - glucose: concentração de glicose no plasma
    - pressure: pressão diastólica (em mm Hg)
    - triceps: espessura da dobra cutânea do tríceps (em mm)
    - insulin: insulina sérica de 2-horas (mu U/ml)
    - mass: índice de massa corporal
    - pedigree: função de pedigree de diabetes (medida hereditária de diabetes)
    - age: idade em anos.

    Para ler o arquivo no R

    ```
    require(mlbench)

    data("PimaIndiansDiabetes",package="mlbench")

    data("PimaIndiansDiabetes2",package="mlbench")
    ```

```
PimaIndiansDiabetes2
```

```
summary(PimaIndiansDiabetes2)
```

```
attach(PimaIndiansDiabetes2)
```

Inicialmente fazer uma análise descritiva através de boxplots comparando os dois grupos de interesse com relação a diabetes, positivo e negativo, para cada variável explicativa. Como há variáveis explicativas assimétricas usar também o boxplot robusto que ajusta a indicação de pontos extremos. Por exemplo, para a variáel explicativa glucose use os comandos

```
require(robustbase)
```

```
boxplot(split(glucose,diabetes),col="gray",
names=c("Negativo","Positivo"), main="Boxplot Tradicional")
```

```
adjbox(split(glucose,diabetes),col="gray",
names=c("Negativo","Positivo"), main="Boxplot Ajustado")
```

Usar apenas o GAMLSS nas análises. Como há muitas observações incompletas, serão consideradas apenas as mulheres com valores para todas as variáveis. Para ajustar inicialmente um modelo logíitico binomial apenas com efeitos principais use o comando

```
fit1.pima = gamlss(formula = diabetes ∼ ·, family = BI, data
= na.omit(PimaIndiansDiabetes2))
```

```
summary(fit1.pima)
```

```
plot(fit1.pima)
```

```
rqres.plot(fit1.pima, howmany=8, type="wp")
```

Use o comando `stepGAIC` do GAMLSS para fazer uma seleção de variáveis explicativas

```
fit2.pima = stepGAIC(fit1.pima)

summary(fit2.pima)

plot(fit2.pima)

rqres.plot(fit2.pima, howmany=8, type="wp")
```

Interpretar os coeficientes ajustados do modelo selecionado e comentar sobre os gráficos de resíduos. Gerar e interpretar o `term.plot` para cada variável explicativa incluída no modelo.

Agora construir a curva ROC. Há vários pacotes no R, e abaixo segue uma aplicação com o pacote ROCR.

```
require(ROCR)

pred = prediction(fitted(fit2.pima),
na.omit(PimaIndiansDiabetes2)$diabetes)

perf = performance(pred, "tpr", "fpr")

plot(perf, xlab="Proporção de Falsos Positivos", ylab="Proporção
de Verdadeiros Positivos", cex=2, cex.axis=1.5, cex.lab=1.5)

abline(0,1,lty=2)

plot(perf,print.cutoffs.at=c(0.30),xlab="Proporção de Falsos
Positivos", ylab="Proporção de Verdadeiros Positivos", cex=2,
cex.axis=1.5, cex.lab=1.5)

abline(0,1,lty=2)
```

Para calcular a área sob a curva ROC usar os comandos

```
area = performance(pred,measure="auc")

area = area@y.values[[1]]

area.
```

Capítulo 5

Modelos para Dados de Contagem

5.1 Introdução

Neste capítulo serão apresentados alguns métodos para a análise de dados de contagem. Inicialmente são apresentados os principais métodos tradicionais e em seguida a modelagem através de regressão. Duas situações de interesse são consideradas. Na primeira delas, muito comum em estudos de seguimento, as unidades amostrais são classificadas segundo os níveis de categorias, tais como sexo, faixa etária e tipo de tratamento e são acompanhadas por um período fixo pré-estabelecido ou até a ocorrência de um determinado evento. Tem-se, portanto, um tempo particular de observação para cada unidade amostral, o qual deverá ser incorporado nas análises. Na segunda situação, o interesse é estudar o número de ocorrências de um evento particular segundo os níveis de categorias, de modo que seja possível construir uma tabela típica de contingência. Aqui, a suposição de distribuição de Poisson para o número de ocorrências do evento em cada configuração de níveis das categorias leva a resultados equivalentes à suposição de distribuição multinomial para as caselas da tabela de contingência formada. Assim, muitas

tabelas de contingência que seriam originalmente analisadas através de um modelo log-linear multinomial podem ser analisadas, alternativamente, por um modelo log-linear de Poisson. A vantagem disso é o fato do modelo log-linear de Poisson ser ajustado mais facilmente do que o modelo log-linear multinomial, além da possibilidade de todos os procedimentos desenvolvidos para os MLGs serem diretamente estendidos para o modelo log-linear de Poisson. Não é discutido, contudo, aspectos particulares na análise de tabelas de contingência, tais como testes ou modelos multinomiais mais específicos.

Discute-se também neste capítulo o fenômeno de sobredispersão que pode ocorrer com dados de contagem quando a variância da variável resposta é maior do que a média. Nesses casos, a suposição de distribuição de Poisson para a resposta é inadequada sendo necessário o uso de modelos alternativos. O modelo de quase-verossimilhança com parâmetro de dispersão leva às mesmas estimativas do modelo de Poisson, porém corrige a variabilidade das estimativas. Em especial será dada atenção aos modelos com resposta binomial negativa, os quais permitem uma análise mais completa dos dados do que os modelos de quase-verossimilhança. Finalmente, será abordado de forma mais sucinta os modelos de Poisson e binomial negativo com excesso de zeros.

5.2 Métodos clássicos: uma única tabela 2×2

Considere inicialmente a tabela abaixo resultante de um estudo de seguimento, em que indivíduos expostos e não expostos são acompanhados ao longo do tempo por um período fixo ou até a ocorrência de um evento.

	E	$\bar{E}$
Casos	y_1	y_2
Pessoas-Tempo	t_1	t_2

Assumir que Y_1 e Y_2 seguem, respectivamente, distribuição de Poisson com parâmetros λ_1 e λ_2, em que λ_1 é a taxa média de casos (por unidade de tempo) no grupo exposto e λ_2 é a taxa média de casos no grupo não exposto. O parâmetro de interesse nesse tipo de estudo é a razão entre as taxas, denotada por $\psi = \frac{\lambda_1}{\lambda_2}$. O objetivo principal é fazer inferências a respeito do parâmetro ψ.

5.2.1 Modelo probabilístico não condicional

A função de probabilidade conjunta de (Y_1, Y_2) fica então dada por

$$\begin{aligned} f(y_1, y_2; \lambda_1, \lambda_2) &= \frac{e^{-\lambda_1 t_1}(\lambda_1 t_1)^{y_1}}{y_1!}\frac{e^{-\lambda_2 t_2}(\lambda_2 t_2)^{y_2}}{y_2!} \\ &= \exp\{-\psi\lambda_2 t_1 - \lambda_2 t_2 + y_1\log(\psi) + (y_1+y_2)\log(\lambda_2) + \\ &\quad y_1\log(t_1) + y_2\log(t_2) - \log(y_1!) - \log(y_2!)\}, \end{aligned}$$

e consequentemente o logaritmo da função de verossimilhança pode ser expresso na forma

$$\begin{aligned} L(\psi, \lambda_2) &= -\psi\lambda_2 t_1 - \lambda_2 t_2 + y_1\log(\psi) + (y_1+y_2)\log(\lambda_2) + \\ &\quad y_1\log(t_1) + y_2\log(t_2) - \log(y_1!) - \log(y_2!). \end{aligned}$$

Pode-se mostrar que a maximização de $L(\psi, \lambda_2)$ leva às estimativas de máxima verossimilhança $\tilde{\lambda}_2 = \frac{y_2}{t_2}$ e $\tilde{\psi} = \frac{y_1 t_2}{y_2 t_1}$. Para obter a variância assintótica $\mathrm{Var}_A(\tilde{\psi})$ pode-se aplicar o método delta

$$\mathrm{Var}_A(\tilde{\psi}) = \left[\frac{\partial\psi}{\partial\boldsymbol{\lambda}}\right]^\top \mathrm{Var}_A(\tilde{\boldsymbol{\lambda}}) \left[\frac{\partial\psi}{\partial\boldsymbol{\lambda}}\right],$$

em que $[\partial\psi/\partial\boldsymbol{\lambda}] = [1/\lambda_2, -\psi/\lambda_2]^\top$ e $\mathrm{Var}_A(\tilde{\boldsymbol{\lambda}}) = \mathrm{diag}\left\{\frac{\lambda_1}{t_1}, \frac{\lambda_2}{t_2}\right\}$ com $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$. Daí obtém-se facilmente

$$\mathrm{Var}_A(\tilde{\psi}) = \frac{\psi}{\lambda_2}\left\{\frac{1}{t_1} + \frac{\psi}{t_2}\right\}.$$

5.2.2 Modelo probabilístico condicional

Pelo teorema da fatorização tem-se que as estatísticas (Y_1, Y_1+Y_2) são suficientes minimais para (ψ, λ_2). Logo, condicionando em $Y_1+Y_2=m$, obtém-se

$$\begin{aligned}
f(a|m;\psi) &= Pr\{Y_1=a \mid Y_1+Y_2=m\} \\
&= \frac{Pr\{Y_1=a, Y_2=m-a\}}{Pr\{Y_1+Y_2=m\}} \\
&= \frac{Pr\{Y_1=a\}Pr\{Y_2=m-a\}}{Pr\{Y_1+Y_2=m\}} \\
&= \frac{e^{-\lambda_1 t_1}(\lambda_1 t_1)^a e^{-\lambda_2 t_2}(\lambda_2 t_2)^{(m-a)}}{\frac{a!e^{-\lambda_1 t_1-\lambda_2 t_2}(\lambda_1 t_1+\lambda_2 t_2)^m (m-a)!}{m!}} \\
&= \binom{m}{a}\frac{(\lambda_1 t_1)^a(\lambda_2 t_2)^{(m-a)}}{(\lambda_1 t_1+\lambda_2 t_2)^m} \\
&= \binom{m}{a}\left(\frac{\lambda_1 t_1}{\lambda_1 t_1+\lambda_2 t_2}\right)^a\left(\frac{\lambda_2 t_2}{\lambda_1 t_1+\lambda_2 t_2}\right)^{(m-a)} \\
&= \binom{m}{a}\pi^a(1-\pi)^{(m-a)},
\end{aligned}$$

em que $\pi = \psi t_1/\{t_2+\psi t_1\} = \psi/\{t_2/t_1+\psi\}$, sendo π a probabilidade de um caso ter sido exposto. Equivalentemente, tem-se que

$$\psi = \frac{\pi t_2}{(1-\pi)t_1}.$$

Mostra-se facilmente que $\hat{\pi} = \frac{a}{m} = \frac{y_1}{y_1+y_2}$ e consequentemente que $\hat{\psi} = \frac{at_2}{bt_1} = \frac{y_1 t_2}{y_2 t_1}$, que coincide com a estimativa $\tilde{\psi}$ (não condicional). Além disso, segue a variância assintótica $\mathrm{Var}_A(\hat{\pi}) = \frac{\pi(1-\pi)}{m}$ e portanto aplicando o método delta obtém-se a variância assintótica

$$\mathrm{Var}_A(\hat{\psi}) = \left[\frac{d\psi}{d\pi}\right]^2 \mathrm{Var}_A(\hat{\pi}) = \left[\frac{t_2}{t_1}\right]^2 \frac{\pi}{m(1-\pi)^3},$$

em que $d\psi/d\pi = \frac{t_1}{t_2}(1-\pi)^{-2}$. Após algumas manipulações algébricas mostra-se que $\tilde{\mathrm{V}}\mathrm{ar}_A(\tilde{\psi}) = \hat{\mathrm{V}}\mathrm{ar}_A(\hat{\psi}) = \left[\frac{t_2}{t_1}\right]^2 \frac{y_1(y_1+y_2)}{y_2^3}$. Assim, as inferências para ψ são

equivalentes sob os modelos não condicional e condicional, diferentemente das inferências para a razão de chances descritas no Capítulo 4. A justificativa é que no caso do produto de duas binomiais independentes a estatística Y_1+Y_2 é suficiente para o parâmetro π_2, porém não é ancilar para ψ. Logo, há perda de informação para ψ com a distribuição condicional (hipergeométrica não central). Mesmo assim muitas inferências para a razão de chances são desenvolvidas sob o modelo condicional, em particular o teste exato de Fisher. No caso do produto de duas Poissons independentes a estatística $Y_1 + Y_2$ é suficiente para λ_2 e ancilar para a razão de taxas ψ. Assim, as inferências para os modelos não condicional e condicional são equivalentes. Fica-se então com o modelo condicional que é mais siomples.

Inferência exata

Aqui o interesse é testar $\mathrm{H}_0 : \psi = 1$ contra $\mathrm{H}_1 : \psi \neq 1$, que é equivalente a testar $\mathrm{H}_0 : \pi = \pi_0$ contra $\mathrm{H}_1 : \pi \neq \pi_0$, em que $\pi_0 = t_1/(t_1 + t_2)$.

O nível descritivo exato para testar a hipótese H_0 contra H_1 é dado por $\mathrm{P} = 2\min\{\mathrm{P}_I, \mathrm{P}_S\}$, em que

$$\mathrm{P}_I = \sum_{x=0}^{a} \binom{m}{x} \pi_0^x (1-\pi_0)^{(m-x)}$$

e

$$\mathrm{P}_S = \sum_{x=a}^{m} \binom{m}{x} \pi_0^x (1-\pi_0)^{(m-x)}.$$

Pode-se usar o resultado abaixo (ver, por exemplo, Leemis e Trivedi, 1996) para expressar a distribuição binomial em função de uma distribuição $F_{u,v}$, ou seja uma distribuição F com u e v graus de liberdade. Supondo $Y \sim \mathrm{B}(n, p)$, tem-se que

$$Pr(Y \geq y) = Pr\{F_{2y,2(n-y+1)} < (n-y+1)p/y(1-p)\}, \qquad (5.1)$$

com $0 < p < 1$. Daí tem-se, sob $\mathrm{H}_0 : \pi = \pi_0$, que

$$\begin{aligned} \mathrm{P}_I &= 1 - \sum_{x=a+1}^{m} \binom{m}{x} \pi_0^x (1-\pi_0)^{(m-x)} \\ &= 1 - Pr\left\{ F_{u,v} < \frac{(m-a-1+1)\pi_0}{(a+1)(1-\pi_0)} \right\} \\ &= 1 - Pr\left\{ F_{u,v} < bt_1/(a+1)t_2 \right\}, \end{aligned}$$

com $b = m-a$, $u = 2(a+1)$ e $v = 2b$. Similarmente, obtém-se sob $\mathrm{H}_0 : \pi = \pi_0$, que

$$\mathrm{P}_S = Pr\{F_{u,v} < (b+1)t_1/at_2\},$$

com $u = 2a$ e $v = 2(b+1)$. De (5.1) tem-se que que os limites exatos de confiança para p, para um coeficiente de confiança $(1-\alpha)$, são tais que

$$\frac{\alpha}{2} = \sum_{t \geq y} Pr(Y = t; \hat{p}_I) = Pr(Y \geq y; \hat{p}_I)$$

e

$$\frac{\alpha}{2} = \sum_{t \leq y} Pr(Y = t; \hat{p}_S) = 1 - Pr(Y \geq y+1; \hat{p}_S).$$

Logo, usando (5.1) obtém-se

$$\hat{p}_I = \frac{1}{1 + \frac{n-y+1}{yF_{2y,2(n-y+1)}(\alpha/2)}}$$

e

$$\hat{p}_S = \frac{1}{1 + \frac{n-y}{(y+1)F_{2(y+1),2(n-y)}(1-\alpha/2)}},$$

em que $F_{u,v}(\alpha/2)$ denota o quantil $\alpha/2$ de uma distribuição F com u e v graus de liberdade. Portanto, tem-se para π, fazendo $y = a$ e $m = a+b$, o limite inferior exato de confiança

$$\begin{aligned} \hat{\pi}_I &= \frac{1}{1 + \frac{b+1}{aF_{u,v}(\alpha/2)}} \\ &= aF_{u,v}(\alpha/2)/\{b+1+aF_{u,v}(\alpha/2)\}, \end{aligned}$$

em que $u = 2a$ e $v = 2(b+1)$. De forma análoga obtém-se o limite superior exato

$$\begin{aligned}\hat{\pi}_S &= \frac{1}{1+\frac{b}{aF_{u,v}(1-\alpha/2)}} \\ &= aF_{u,v}(1-\alpha/2)/\{b+aF_{u,v}(1-\alpha/2)\},\end{aligned}$$

em que $u = 2(a+1)$ e $v = 2b$. A estimativa de máxima verossimilhança para ψ considerando a distribuição condicional fica dada por

$$\hat{\psi} = \frac{\hat{\pi}t_2}{(1-\hat{\pi})t_1} = \frac{y_1 t_2}{y_2 t_1}.$$

Portanto, a estimativa intervalar exata de coeficiente de confiança $(1-\alpha)$ para ψ fica denotada por $[\hat{\psi}_I, \hat{\psi}_S]$, em que

$$\hat{\psi}_I = \frac{\hat{\pi}_I t_2}{(1-\hat{\pi}_I)t_1} \quad \text{e} \quad \hat{\psi}_S = \frac{\hat{\pi}_S t_2}{(1-\hat{\pi}_S)t_1}.$$

Inferência assintótica

Embora a inferência exata para a razão de taxas tenha um custo computacional bem menor do que para a razão de chances, tem-se também a opção da inferência assintótica para a razão de taxas quando λ_1 e λ_2 são grandes no modelo não condicional ou quando m é grande no modelo condicional. Similarmente ao caso da razão de chances a aproximação para a distribuição normal é mais rápida para $\log(\tilde{\psi})$ do que para $\tilde{\psi}$. Assim, aplicando o método delta tem-se que

$$\begin{aligned}\mathrm{Var}_A\{\log(\tilde{\psi})\} &= \left[\frac{d\log(\psi)}{d\psi}\right]^2 \mathrm{Var}_A(\tilde{\psi}) \\ &= \frac{1}{\lambda_1 t_1} + \frac{1}{\lambda_2 t_2},\end{aligned}$$

em que $\frac{d\log(\psi)}{d\psi} = \frac{1}{\psi}$. Daí segue que uma estimativa intervalar assintótica de coeficiente de confiança $(1-\alpha)$ para ψ fica dada por

$$\exp[\log(\tilde{\psi}) \pm z_{(1-\alpha/2)}\sqrt{\tilde{\mathrm{Var}}\{\log(\tilde{\psi})\}}],$$

em que $\tilde{\text{V}}\text{ar}\{\log(\tilde{\psi})\} = \frac{1}{y_1} + \frac{1}{y_2}$. O teste de Wald para testar $\text{H}_0 : \psi = 1$ contra $\text{H}_1 : \psi \neq 1$ fica dado por

$$\xi_W = \frac{\{\log(\tilde{\psi})\}^2}{\tilde{\text{V}}\text{ar}\{\log(\tilde{\psi})\}}$$

que segue assintoticamente sob H_0 distribuição qui-quadrado com 1 grau de liberdade.

Aplicação

Considere, como aplicação, os dados apresentados em Boice e Monson (1977) referentes a um estudo de seguimento com dois grupos de mulheres com tuberculose, um grupo exposto a radiação e o outro grupo não exposto, sendo observado ao longo do tempo o desenvolvimento ou não de câncer de mama. Os resultados desse estudo são resumidos na Tabela 5.1.

Tabela 5.1
Casos de câncer de mama em mulheres com tuberculose.

	Radiação	
	Exposto	Não Exposto
Casos	41	15
Pessoas-anos	28010	19017

Tem-se, portanto, que $a = 41$, $b = 15$, $t_1 = 28010$ e $t_2 = 19017$. Os níveis descritivos correspondentes ao teste exato para testar $\text{H}_0 : \psi = 1$ contra $\text{H}_1 : \psi \neq 1$ ficam dados por

$$\text{P}_I = 1 - Pr\{F_{84,30} < 0,526\} = 0,988$$

e

$$\text{P}_S = Pr\{F_{82,32} < 0,575\} = 0,026,$$

e obtém-se o nível descritivo P= $0,052$ que indica pela rejeição de H_0. Isso quer dizer que há indícios de que mulheres com tuberculose e expostas a radiação têm uma chance maior de desenvolvimento de câncer de mama do que mulheres não expostas com a mesma doença. Uma estimativa pontual de máxima verossimilhança para ψ fica dada por $\hat{\psi} = \frac{0,732 \times 19017}{0,268 \times 28010} = 1,86$, que corresponde à estimativa da razão de médias (por ano) de casos de câncer de mama entre mulheres com tuberculose que foram expostas à radiação e mulheres com tuberculose não expostas à radiação. Uma estimativa intervalar exata de de 95% para π tem os limites

$$\begin{aligned}
\hat{\pi}_I &= 41 \times F_{84,30}(0,025)/\{16 + 41 \times F_{84,30}(0,025)\} \\
&= 0,595 \text{ e} \\
\hat{\pi}_S &= 41 \times F_{82,32}(0,975)/\{15 + 41 \times F_{82,32}(0,975)\} \\
&= 0,836,
\end{aligned}$$

em que $F_{84,30}(0,025) = 0,574$ e $F_{82,32}(0,975) = 1,866$. Desses limites obtém-se os limites exatos de confiança para a razão de tazas ψ

$$\begin{aligned}
\hat{\psi}_I &= \frac{\hat{\pi}_I t_2}{(1-\hat{\pi}_I)t_1} = \frac{0,595 \times 19017}{(1-0,595) \times 28010} \\
&= 0,997 \text{ e} \\
\hat{\psi}_S &= \frac{\hat{\pi}_S t_2}{(1-\hat{\pi}_S)t_1} = \frac{0,836 \times 19017}{(1-0,836) \times 28010} \\
&= 3,461.
\end{aligned}$$

Esse intervalo $[0,997; 3,461]$ cobre ligeiramente o valor 1 uma vez que o nível descritivo do teste $H_0 : \psi = 1$ contra $H_1 : \psi \neq 1$ foi P= $0,052$.

5.2.3 Estratificação: k tabelas 2×2

Se o dados são estratificados segundo um fator com k níveis, cada tabela resultante pode ser expressa na forma abaixo.

	E	$\bar{E}$
Casos	y_{1i}	y_{2i}
Pessoas-Tempo	t_{1i}	t_{2i}

Tem-se aqui as suposições $Y_{1i} \sim \mathrm{P}(\lambda_{1i}t_{1i})$ e $Y_{2i} \sim \mathrm{P}(\lambda_{2i}t_{2i})$, $i = 1, \ldots, k$. Consequentemente, a distribuição condicional de Y_{1i} dado $Y_{1i} + Y_{2i} = m_i$ é uma $\mathrm{B}(m_i, \pi_i)$, em que $\pi_i = \psi_i/\{t_{2i}/t_{1i} + \psi_i\}$, ou equivalentemente

$$\psi_i = \frac{\pi_i t_{2i}}{(1-\pi_i)t_{1i}}.$$

Se há interesse em testar a homogeneidade das razões de taxas $\mathrm{H}_0 : \psi_1 = \ldots = \psi_k$ contra a alternativa de pelo menos duas diferentes, a estimativa comum $\hat{\psi}$, sob H_0, sai do sistema de equações

$$\sum_{i=1}^{k} y_{1i} = \hat{\psi} \sum_{i=1}^{k} m_i/\{\hat{\psi} + t_{2i}/t_{1i}\},$$

que tem no máximo uma raiz positiva. Alternativamente, de forma análoga aos estudos de caso e controle, pode-se construir uma versão da estimativa de Mantel-Haenszel dada por

$$\hat{\psi}_{MH} = \frac{\sum_{i=1}^{k} y_{1i}t_{2i}/t_i}{\sum_{i=1}^{k} y_{2i}t_{1i}/t_i},$$

em que $t_i = t_{1i} + t_{2i}$. Segundo Breslow e Day (1987), $\hat{\psi}_{MH}$ é consistente e assintoticamente normal com variância assintótica estimada por

$$\hat{\mathrm{V}}\mathrm{ar}_A(\hat{\psi}_{MH}) = \frac{\hat{\psi}_{MH} \sum_{i=1}^{k} t_{1i}t_{2i}m_i/t_i^2}{\left\{\sum_{i=1}^{k} \frac{t_{1i}t_{2i}m_i}{t_i(t_{1i}+\hat{\psi}_{MH}t_{2i})}\right\}^2}.$$

A estatística sugerida para testar H_0 é definida por

$$X^2 = \sum_{i=1}^{k} \left\{ \frac{(y_{1i} - \hat{y}_{1i})^2}{\hat{y}_{1i}} + \frac{(y_{2i} - \hat{y}_{2i})^2}{\hat{y}_{2i}} \right\},$$

em que $\hat{y}_{1i} = m_i\hat{\pi}_i$, $\hat{y}_{2i} = m_i(1-\hat{\pi}_i)$ e

$$\hat{\pi}_i = \frac{\hat{\psi}_{MH}}{t_{2i}/t_{1i} + \hat{\psi}_{MH}}.$$

A distribuição nula assintótica de X^2 é uma qui-quadrado com $k-1$ graus de liberdade. Quando a hipótese de homogeneidade das razões de chances não é rejeitada, pode-se testar a hipótese de associação entre o fator e a doença levando em conta o efeito de estrato. Isso equivale a testar $\mathrm{H}_0 : \psi = 1$ contra $\mathrm{H}_1 : \psi \neq 1$. O teste qui-quadrado apropriado, com correção de continuidade, é dado por

$$\begin{aligned} X^2 &= \frac{\{|\sum_{i=1}^k y_{1i} - \sum_{i=1}^k \mathrm{E}(Y_{1i}|m_i, \psi = 1)| - 0,5\}^2}{\sum_{i=1}^k \mathrm{Var}(Y_{1i}|m_i, \psi = 1)} \\ &= \frac{\{|\sum_{i=1}^k y_{1i} - \sum_{i=1}^k m_i t_{1i}/(t_{1i}+t_{2i})| - 0,5\}^2}{\sum_{i=1}^k m_i t_{1i} t_{2i}/(t_{1i}+t_{2i})^2}. \end{aligned} \tag{5.2}$$

Para m_i grandes a distribuição nula assintótica da estatística X^2 é uma qui-quadrado com $(k-1)$ graus de liberdade.

A distribuição assintótica de $\log(\hat{\psi})$ converge mais rapidamente para a distribuição normal do que a distriuição assintótica de $\hat{\psi}$. Assim, quando a hipótese de homogeneidade de razão de taxas não é rejeitada é mais conveniente, similarmente à razão de chances, obter uma estimativa intervalar para $\log(\psi)$ comum e daí extrair a estimativa intervalar de ψ comum nos estratos.

Aplicando-se o método delta, tem-se que a variância assintótica de $\log(\hat{\psi}_{MH})$ é estimada por

$$\hat{\mathrm{V}}\mathrm{ar}_A\{\log(\hat{\psi}_{MH})\} = \hat{\psi}_{MH}^{-2}\hat{\mathrm{V}}\mathrm{ar}_A(\hat{\psi}_{MH}).$$

Assim, um intervalo assintótico de confiança com coeficiente $(1-\alpha)$ para $\log(\psi)$ fica dado por

$$\log(\hat{\psi}_{MH}) \pm z_{(1-\alpha/2)}\hat{\psi}_{MH}^{-1}\{\hat{\mathrm{V}}\mathrm{ar}_A(\hat{\psi}_{MH})\}^{\frac{1}{2}}$$

levando aos limites de confiança superior e inferior dados abaixo

$$\begin{aligned}\hat{\psi}_I &= \exp\{\log(\hat{\psi}_{MH}) - z_{(1-\alpha/2)}\hat{\psi}_{MH}^{-1}\sqrt{\hat{\mathrm{Var}}_A(\hat{\psi}_{MH})}\} \quad \text{e} \\ \hat{\psi}_S &= \exp\{\log(\hat{\psi}_{MH}) + z_{(1-\alpha/2)}\hat{\psi}_{MH}^{-1}\sqrt{\hat{\mathrm{Var}}_A(\hat{\psi}_{MH})}\}.\end{aligned}$$

Esse intervalo deve ser construído quando a aplicação da estatística (5.2) levar à rejeição da hipótese $\mathrm{H}_0 : \psi = 1$.

Aplicação

Como ilustração, na Tabela 5.2 tem-se um resumo do número de avarias causadas por ondas em navios de carga e os respectivos tempos de exposição (em navios-meses) para dois tipos de navios e dois períodos de operação.

Tabela 5.2
Número de avarias por ondas em navios de carga segundo dois tipos de navios e dois períodos de operação.

Período de operação		Tipo de navio	
		Tipo E	Tipo A
P1	avarias	12	9
	n-meses	1991	2734
P2	avarias	20	33
	n-meses	3140	6755

As estimativas pontuais para a razão de taxas entre os tipos E e A são, respectivamente, dadas por

$$\hat{\psi}_1 = \frac{12 \times 2734}{9 \times 1991} = 1,83 \quad \text{e} \quad \hat{\psi}_2 = \frac{20 \times 6755}{33 \times 3140} = 1,30.$$

Para obter a estimativa intervalar de 95% para a razão de taxas entre os tipos E e A para o período 1 de operação é preciso que calcular inicialmente

as probabilidades

$$\begin{aligned}
\hat{\pi}_{1I} &= 12F_{24,20}(0,025)/\{10 + 12F_{24,20}(0,025)\} \\
&= 0,340 \quad \text{e} \\
\hat{\pi}_{1S} &= 12F_{26,18}(0,975)/\{9 + 12F_{26,18}(0,975)\} \\
&= 0,768.
\end{aligned}$$

Logo, obtém-se a estimativa intervalar de 95%

$$\begin{aligned}
\hat{\psi}_{1I} &= \frac{\hat{\pi}_{1I}t_{12}}{(1-\hat{\pi}_{1I})t_{11}} = \frac{0,340 \times 2734}{0,660 \times 1991} = 0,707 \quad \text{e} \\
\hat{\psi}_{1S} &= \frac{\hat{\pi}_{1S}t_{12}}{(1-\hat{\pi}_{1S})t_{11}} = \frac{0,768 \times 2734}{0,232 \times 1991} = 4,546.
\end{aligned}$$

De forma similar, para o período 2 de operação, obtém-se

$$\begin{aligned}
\hat{\pi}_{2I} &= 20F_{40,68}(0,025)/\{34 + 20F_{40,68}(0,025)\} \\
&= 0,248 \quad \text{e} \\
\hat{\pi}_{2S} &= 20F_{42,66}(0,975)/\{33 + 20F_{42,66}(0,975)\} \\
&= 0,509.
\end{aligned}$$

A estimativa intervalar de 95% fica dada por

$$\begin{aligned}
\hat{\psi}_{2I} &= \frac{\hat{\pi}_{2I}t_{22}}{(1-\hat{\pi}_{2I})t_{21}} = \frac{0,248 \times 6755}{0,752 \times 3140} = 0,709 \quad \text{e} \\
\hat{\psi}_{2S} &= \frac{\hat{\pi}_{2S}t_{22}}{(1-\hat{\pi}_{2S})t_{21}} = \frac{0,509 \times 6755}{0,491 \times 3140} = 2,230.
\end{aligned}$$

Nota-se que ambas as estimativas intervalares cobrem o valor 1, indicando pela não rejeição da mesma taxa de avarias entre os dois tipos de navios em cada período de operação.

Para aplicar o teste de homogeniedade de razão de taxas entre os dois tipos de navios, deve-se inicialmente obter a estimativa de razão de taxas

comum de Mantel-Haenszel

$$\hat{\psi}_{MH} = \left(\frac{12 \times 2734}{1991 + 2734} + \frac{20 \times 6755}{3140 + 6755}\right) / \left(\frac{9 \times 1991}{1991 + 2734} + \frac{33 \times 3140}{3140 + 6755}\right) = 1,44.$$

As estimativas da probabilidade da avaria ter sido de navio do tipo E ficam, respectivamente, sob a hipótese de ψ constante dadas por

$$\begin{aligned}
\hat{\pi}_1 &= \hat{\psi}_{MH} t_{11}/(t_{12} + \hat{\psi}_{MH} t_{11}) \\
&= 1,44 \times 1991/(2734 + 1,44 \times 1991) = 0,512 \quad \text{e} \\
\hat{\pi}_2 &= \hat{\psi}_{MH} t_{21}/(t_{22} + \hat{\psi}_{MH} t_{21}) \\
&= 1,44 \times 3140/(6755 + 1,44 \times 3140) = 0,401.
\end{aligned}$$

Assim, tem-se os valores esperados de avarias para os dois tipos de navios e dois períodos de operação sob a hipótese de homogeneidade da razão de taxas: $\hat{y}_{11} = m_1\hat{\pi}_1 = 21 \times 0,512 = 10,752$, $\hat{y}_{12} = m_1(1 - \hat{\pi}_1) = 21 \times 0,488 = 10,248$ $\hat{y}_{21} = m_2\hat{\pi}_2 = 53 \times 0,401 = 21,253$ e $\hat{y}_{22} = m(1 - \hat{\pi}_2) = 53 \times 0,599 = 31,747$. A estatística para testar as hipóteses $\mathrm{H}_0 : \psi_1 = \psi_2$ contra $\mathrm{H}_1 : \psi_1 \neq \psi_2$ fica dada por

$$\begin{aligned}
X^2 &= \frac{(12 - 10,752)^2}{10,752} + \frac{(9 - 10,248)^2}{10,248} \\
&\quad + \frac{(20 - 21,253)^2}{21,253} + \frac{(33 - 31,747)^2}{31,747} \\
&= 0,420,
\end{aligned}$$

que comparado com os quantis da distribuição qui-quadrado com 1 grau de liberdade leva ao nível descritivo $\mathrm{P} = 0,52$, indicando pela não rejeição da hipótese nula.

Finalmente, deve-se testar as hipóteses $\mathrm{H}_0 : \psi = 1$ contra $\mathrm{H}_1 : \psi \neq 1$, em que ψ denota a razão de taxas comum. A estatística do teste de Mantel-

Hanszel com correção de continuidade fica dada por

$$\begin{aligned} X^2 &= \frac{\{|y_{11}+y_{12}-\{m_1t_{11}/(t_{11}+t_{21})+m_2t_{12}/(t_{12}+t_{22})\}|-0,5\}^2}{m_1t_{11}t_{21}/(t_{11}+t_{21})^2+m_2t_{12}t_{22}/(t_{12}+t_{22})^2} \\ &= \frac{(|12+20-(8,85+16,82)|-0,5)^2}{5,12+11,48} = 2,05, \end{aligned}$$

cujo nível descritivo, quando comparado com os quantis da distribuição qui-quadrado com 1 grau de liberdade é dado por $\mathrm{P} = 0,15$, não rejeitando-se a hipótese nula.

5.3 Modelos de Poisson

5.3.1 Propriedades da Poisson

Supor que $Y \sim \mathrm{P}(\lambda)$ cuja função de probabilidade é dada por

$$Pr(Y=y) = \frac{e^{-\lambda}\lambda^y}{y!}, \quad y = 0,1,2,\ldots.$$

Pode-se mostrar (ver, por exemplo, McCullagh e Nelder, 1989, p. 195) que quando $\lambda \to \infty$

$$(Y-\lambda)/\sqrt{\lambda} \to_d \mathrm{N}(0,1).$$

Em outras palavras, para λ grande tem-se que Y segue aproximadamente distribuição normal de média λ e desvio padrão $\sqrt{\lambda}$. Se, no entanto, o interesse é aplicar um modelo normal linear para explicar λ, tem-se o incoveniente do desvio padrão depender da média, inviabilizando o uso de um modelo normal linear homocedástico. Uma maneira de contornar esse problema é através da aplicação de uma transformação na resposta Y de modo a alcançar a normalidade e a constância de variância, mesmo que aproximadamente. Nesse sentido, tem-se que se Y é Poisson, segue quando $\lambda \to \infty$ o seguinte resultado:

$$\{\sqrt{Y} - \mathrm{E}(\sqrt{Y})\} \to_d \mathrm{N}(0,1/4).$$

Portanto, quando λ é grande, a variável aleatória $2\{\sqrt{Y} - \mathrm{E}(\sqrt{Y})\}$ segue aproximadamente distribuição $\mathrm{N}(0,1)$. Assim, para uma amostra aleatória $Y_1, \ldots, Y_n$ tal que $Y_i \sim \mathrm{P}(\lambda_i)$ se o interesse é explicar λ_i através de variáveis explicativas, pode-se propor para λ_i grande, $\forall i$, o modelo normal linear

$$\sqrt{Y_i} = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i,$$

em que $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$, $i = 1, \ldots, n$. Isso foi feito na Seção 2.10.3 no exemplo sobre sobrevivência de bactérias.

5.3.2 Modelos log-lineares: k tabelas 2×2

Como visto no Capítulo 2, os modelos log-lineares são recomendados para a análise de dados de contagem, mesmo quando o tempo de observação não é o mesmo para cada unidade amostral. Em particular, para um conjunto de k tabelas 2×2, uma modelagem possível para a taxa média por unidade de tempo em cada casela é supor que $Y_{1i} \sim \mathrm{P}(\lambda_{1i} t_{1i})$ e $Y_{2i} \sim \mathrm{P}(\lambda_{2i} t_{2i})$, $i = 1, \ldots, k$, mutuamente independentes e com a seguinte parte sistemática:

$$\begin{aligned}
\log(\lambda_{11}) &= \alpha, \\
\log(\lambda_{21}) &= \alpha + \beta, \\
\log(\lambda_{1i}) &= \alpha + \gamma_i \text{ e} \\
\log(\lambda_{2i}) &= \alpha + \beta + \gamma_i + \delta_i,
\end{aligned}$$

para $i = 2, \ldots, k$. Portanto, tem-se a reparametrização $(\lambda_{11}, \lambda_{21}, \ldots, \lambda_{1k}, \lambda_{2k})$ $\rightarrow$ $(\alpha, \beta, \gamma_2, \delta_2, \ldots, \gamma_k, \delta_k)$. A razão de taxas na i-ésima tabela fica definida por $\psi_i = \lambda_{2i}/\lambda_{1i} = \exp(\beta + \delta_i)$, com $\delta_1 = 0$. Assim, testar $\mathrm{H}_0 : \psi_1 = \cdots = \psi_k$ contra H_1 : pelo menos dois ψ's diferentes é o mesmo que testar na nova parametrização $\mathrm{H}_0 : \delta_2 = \cdots = \delta_k = 0$ contra H_1 : pelo menos dois δ_i's diferentes, que é equivalente a ausência de interação entre as tabelas. Deve-se lembrar que γ_i é o efeito da i-ésima tabela com relação à primeira tabela.

Logo, testar $\mathrm{H}_0 : \gamma_2 = \cdots = \gamma_k$, dado que $\delta_i = 0$, significa testar a ausência de efeito de estrato.

Aqui t_{ij} denota o total de unidades de tempo na casela (i, j), $i = 1, 2$ e $j = 1, \ldots, k$. Assim, tem-se que $\log(\mu_{ij}) = \log(t_{ij}) + \log(\lambda_{ij})$, em que $\log(t_{ij})$ desempenha o papel de um *offset*. Pela propriedade de que os totais marginais $Y_{1i} + Y_{2i}$ são estatísticas suficientes para os parâmetros $\lambda_{21}, \ldots, \lambda_{2k}$ e ancilares para $\psi_1, \ldots, \psi_k$, deve-se esperar que as estimativas de máxima verossimilhança não condicionais $\hat{\psi}_i = \exp(\hat{\beta} + \hat{\delta}_i)$, $i = 1, \ldots, k$, coincidam com as estimativas condicionais.

Uma maneira de verificar se é razoável a suposição de distribuição de Poisson nas unidades de tempo é tratar $\log(T_{ij})$ como sendo uma variável explicativa, isto é, ajustar o modelo com parte sistemática dada por $\log(\mu_{ij}) = \theta \log(t_{ij}) + \log(\lambda_{ij})$. Assim, ao testar $\mathrm{H}_0 : \theta = 1$ contra $\mathrm{H}_1 : \theta \neq 1$, a não rejeição de H_0 indica que a suposição de distribuição de Poisson nas unidades de tempo não é inadequada. Como será mostrado a seguir isso significa que os tempos têm distribuição exponencial.

Relação com a exponencial

O logaritmo da função de verossimilhança do modelo de Poisson para a análise de k tabelas 2×2 é dado por

$$\mathrm{L}(\boldsymbol{\lambda}) \propto \sum_{i=1}^{2} \sum_{j=1}^{k} (y_{ij} \log(\lambda_{ij}) - \lambda_{ij} t_{ij}), \tag{5.3}$$

em que $\boldsymbol{\lambda} = (\lambda_{11}, \lambda_{21}, \ldots, \lambda_{k1}, \lambda_{k2})^\top$. Tem-se, portanto, para cada casela (i, j) um estudo de seguimento em que as unidades amostrais foram observadas um total de t_{ij} unidades de tempo. Sem perda de generalidade, supor que $t_{ij} = N$ e que nesse subestrato foram acompanhadas I unidades amostrais cujos tempos de observação foram, respectivamente, $N_1, N_2, \ldots, N_I$. Considerar

$u_\ell = 1$ se o evento sob estudo ocorrer para a ℓ-ésima unidade amostral antes de um tempo pré-fixado T. Quando o evento não ocorrer para a ℓ-ésima unidade amostral durante o período de estudo ($u_\ell = 0$) não há censura, sendo aqui o tempo de observação dado por $N_\ell = T$. Supor ainda que a taxa de ocorrência do evento, que é definida por

$$\xi = \lim_{\Delta t \to 0} \frac{Pr\{\text{o evento ocorrer em } (t, t + \Delta t)\}}{\Delta t},$$

dado que o evento não ocorreu até o tempo t, permanece constante durante o período de observação. Finalmente, assumir que as ocorrências são independentes entre as unidades amostrais. Sob essas condições, mostra-se que a distribuição conjunta das variáveis (N_ℓ, u_ℓ), $\ell = 1, \ldots, I$, é um produto de I exponenciais independentes de parâmetro ξ. Se o evento ocorrer antes do tempo T para a ℓ-ésima unidade amostral ($N_\ell < T, u_\ell = 1$) a mesma contribui com o fator $\xi e^{-\xi N_\ell}$ na função de verossimilhança. Caso contrário ($N_\ell = T, u_\ell = 0$), o fator é dado por $e^{-\xi T}$. O logaritmo da função de verossimilhança conjunta fica então dado por

$$\begin{aligned} \mathrm{L}(\xi) &= \sum_{\ell=1}^{I} \{u_\ell \log(\xi) - N_\ell \xi\} \\ &= \log(\xi) \sum_{\ell=1}^{I} u_\ell - \xi \sum_{\ell=1}^{I} N_\ell. \end{aligned} \tag{5.4}$$

Se considerar que para a casela (i, j) o evento ocorreu y_{ij} vezes, as unidades amostrais foram observadas um total de t_{ij} unidades de tempo e a taxa de ocorrência do evento é λ_{ij}, então (5.4) fica reexpressa na forma

$$\mathrm{L}(\lambda_{ij}) = y_{ij} \log(\lambda_{ij}) - \lambda_{ij} t_{ij},$$

que coincide com o termo geral da expressão (5.3). Portanto, a suposição de modelo de regressão log-linear de Poisson com *offset* $\log(t_{ij})$ equivale à

suposição de tempos exponenciais para as unidades amostrais. No entanto, é importante ressaltar que as inferências exatas para ξ no modelo exponencial são bastante complexas em virtude da ocorrência de censura (ver discussão, por exemplo, em Breslow e Day, 1987, p. 132). Já os resultados assintóticos são equivalentes àqueles obtidos para o modelo de Poisson.

Aplicação

A Tabela 5.3 resume os resultados de um estudo de seguimento em que doutores Britânicos foram acompanhados durante a década de 1950 e observado, em particular, a ocorrência de mortes por câncer de pulmão segundo o consumo médio diário de cigarros e a faixa etária. Esses dados estão disponíveis no arquivo **breslow.txt**.

Tabela 5.3
Número de casos de morte por câncer de pulmão e pessoas-anos de observação em doutores Britânicos segundo a faixa etária e o consumo médio diário de cigarros.

Consumo médio diário de cigarros		Faixa Etária			
		40-49	50-59	60-69	70-80
0	mortes	0	3	0	3
	p-anos	33679	21131,5	10599	4495,5
1-9	mortes	0	1	3	3
	p-anos	6002,5	4396	2813,5	1664,5
10-30	mortes	7	29	41	45
	p-anos	34414,5	25429	13271	4765,5
+ 30	mortes	3	16	36	11
	p-anos	5881	6493,5	3466,5	769

Denotar por Y_{ij} o número de mortes para o i-ésimo nível de consumo e j-ésima faixa etária, $i, j = 1, \ldots, 4$. Supor que $Y_{ij} \sim \mathrm{P}(\lambda_{ij} t_{ij})$, em que λ_{ij} é a

taxa média de mortes por unidade de tempo para o consumo i e faixa etária j. O modelo saturado nesse caso é dado por

$$\log(\lambda_{ij}) = \alpha + \beta_i + \gamma_j + \delta_{ij},$$

em que $\beta_1 = 0$, β_i é o efeito da i-ésima classe de consumo de cigarros com relação à classe de não fumantes, $i = 2, 3, 4$, $\gamma_1 = 0$, γ_j é o efeito da j-ésima faixa etária com relação à faixa etária de $40 - 49$ anos e δ_{ij} denota a interação entre faixa etária e consumo de cigarros, em que $\delta_{i1} = \delta_{1j} = 0$, para $i, j = 1, \ldots, 4$.

O teste de ausência de interação, $H_0 : \delta_{ij} = 0, \forall ij$, contra a alternativa de pelo menos um parâmetro diferente de zero forneceu $\xi_{RV} = 11,91$ (9 graus de liberdade) que equivale a um nível descritivo P$= 0,218$. Adota-se, portanto, um modelo sem interação entre faixa etária e consumo de cigarros.

Tabela 5.4
Estimativas dos parâmetros do modelo log-linear de Poisson para explicar a taxa média de morte de doutores Britânicos com câncer de pulmão.

Efeito	Parâmetro	Estimativa	E/E.Padrão
Constante	α	-11,424	-22,44
C(1-9)	β_2	1,409	2,53
C(10-20)	β_3	2,866	6,86
C(+30)	β_4	3,758	8,80
F(50-59)	γ_2	1,769	5,10
F(60-69)	γ_3	2,897	8,62
F(70-80)	γ_4	3,791	11,12

As estimativas são apresentadas na Tabela 5.4. Nota-se claramente que as estimativas são significativamente diferentes de zero e que há fortes indícios de um aumento (exponencial) da taxa média de mortes com o aumento da faixa etária e/ou com o aumento do consumo médio diário de cigarros. O

ajuste do modelo com $\log(T_{ij})$ como variável explicativa forneceu a estimativa de máxima verossimilhança $\hat{\theta} = 1,839(0,610)$. O teste de Wald para testar $\mathrm{H}_0 : \theta = 1$ contra $\mathrm{H}_1 : \theta \neq 1$ forneceu o valor

$$\xi_W = \frac{(1,839-1)^2}{0,610^2} = 1,89,$$

cujo nível descritivo é dado por P$= 0,17$, indicando que o modelo pode ser ajustado com $\log(t_{ij})$ como sendo *offset*. O gráfico normal de probabilidades descrito na Figura 5.1 indica que o modelo está bem ajustado.

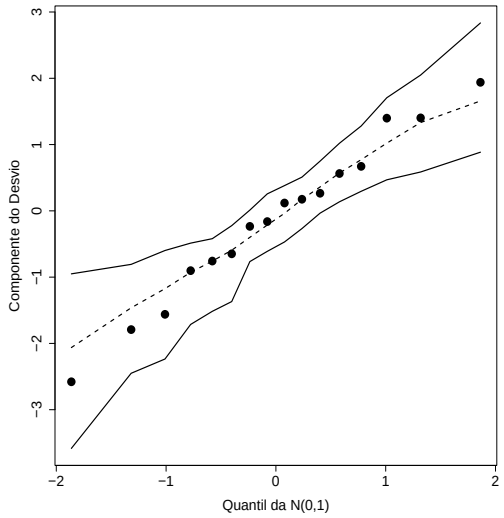

Figura 5.1: Gráfico normal de probabilidades referente ao modelo log-linear de Poisson ajustado aos dados sobre morte por câncer de pulmão de doutores Britânicos.

5.3.3 Modelos gerais de Poisson

Supor agora que Y_i's são variáveis aleatórias independentes distribuídas tais que $Y_i \sim \mathrm{P}(\mu_i)$, com parte sistemática dada por $g(\mu_i) = \eta_i$, em que $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ contém valores de variáveis explicativas, para

$i = 1, \ldots, n$, e $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ é um vetor de parâmetros desconhecidos. As ligações mais utilizados são logarítmica ($g(\mu_i) = \log(\mu_i)$), raiz quadrada ($g(\mu_i) = \sqrt{\mu_i}$) e identidade ($g(\mu_i) = \mu_i$).

O processo iterativo para a estimação de $\boldsymbol{\beta}$, como foi visto na Seção 1.6.1, é dado por

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{z}^{(m)},$$

$m = 0, 1, \ldots$, com variável dependente modificada $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{-\frac{1}{2}} \mathbf{V}^{-\frac{1}{2}} (\mathbf{y} - \boldsymbol{\mu})$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$, $\mathbf{y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, $\mathbf{V} = \text{diag}\{\mu_1, \ldots, \mu_n\}$ e $\mathbf{W} = \text{diag}\{\omega_1, \ldots, \omega_n\}$ com $\omega_i = (d\mu_i/d\eta_i)^2/\mu_i$. Em particular tem-se $\omega_i = \mu_i$ para ligação logarítmica, $\omega_i = 4$ para ligação raiz quadrada e $\omega_i = \mu_i^{-1}$ para ligação identidade.

No caso das unidades experimentais serem observadas em tempos distintos t_i's e for assumido que $Y_i \overset{\text{ind}}{\sim} \text{P}(\lambda_i t_i)$, $i = 1, \ldots, n$, a parte sistemática do modelo para ligação logarítmica fica dada por

$$\log(\mu_i) = \log(t_i) + \mathbf{x}_i^\top \boldsymbol{\beta},$$

em que $\log(t_i)$ desempenha o papel de *offset* e isso deve ser informado ao sistema. Outra possibilidade é incluir os tempos t_i's como valores da variável explicativa $\log(T_i)$. Nesse caso, a parte sistemática assume a forma

$$\log(\mu_i) = \theta \log(t_i) + \mathbf{x}_i^\top \boldsymbol{\beta}.$$

O teste de $\text{H}_0 : \theta = 1$ contra $\text{H}_1 : \theta \neq 1$ verifica se $\log(t_i)$ deve ser incluído no modelo como *offset*. A não rejeição da hipótese nula significa a suposição de tempos exponenciais nas unidades experimentais.

O estimador de máxima verossimilhança $\hat{\boldsymbol{\beta}}$ é consistente, eficiente e tem distribuição assintótica dada por

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \ \sim \text{N}_p(\mathbf{0}, (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1}),$$

portanto, assintoticamente, $\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1}$.

5.3.4 Qualidade do ajuste

A função desvio de um modelo de Poisson supondo $y_i > 0, \forall i$, é definida por

$$\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\{y_i\log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)\}.$$

Porém, se $y_i = 0$, o i-ésimo termo de $\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}})$ fica dado por $2\hat{\mu}_i$.

Em particular, para ligação logarítmica e se o modelo inclui uma constante na parte sistemática, mostra-se facilmente que $\sum_{i=1}^{n}(y_i - \hat{\mu}_i) = 0$, ficando a função desvio reexpressa na forma $\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} y_i\log(y_i/\hat{\mu}_i)$. Logo, particionando o vetor de parâmetros tal que $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, em que $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$ são subvetores de dimensão $p-q$ e q, respectivamente, a estatística do teste da razão de verossimilhanças para testar $\mathrm{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ contra $\mathrm{H}_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$ em modelos log-lineares fica dada por

$$\begin{aligned}\xi_{RV} &= \mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}_0) - \mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}) \\ &= 2\sum_{i=1}^{n} y_i\log(\hat{\mu}_{0i}/\hat{\mu}_i).\end{aligned}$$

Sob H_0 e para grandes amostras $\xi_{RV} \sim \chi^2_q$. Os resultados assintóticos para os modelos de Poisson valem tanto para p fixo e $n \to \infty$ como para n fixo e $\mu_i \to \infty, \forall i$.

5.3.5 Técnicas de diagnóstico

Um dos resíduos mais recomendados para modelos com resposta de Poisson é o componente do desvio padronizado, que para $y_i > 0$, fica dado por

$$t_{D_i} = \pm\sqrt{\frac{2}{1-\hat{h}_{ii}}}\{y_i\log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)\}^{\frac{1}{2}},$$

em que h_{ii} é o i-ésimo elemento da diagonal principal da matriz de projeção $\mathbf{H} = \mathbf{W}^{\frac{1}{2}}\mathbf{X}(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{W}^{\frac{1}{2}}$. Quando $y_i = 0$ o resíduo componente do desvio padronizado assume a forma $t_{D_i} = \pm\sqrt{2\hat{\mu}_i}/\sqrt{1-\hat{h}_{ii}}$.

Estudos de simulação (ver Williams, 1984) mostram que em geral a distribuição de t_{D_i} não se afasta muito da distribuição normal padrão, podendo ser usadas nas análises de diagnóstico as mesmas interpretações da regressão normal linear. Em particular, a construção de envelopes é fortemente recomendada para t_{D_i}.

Conforme descrito na Seção 2.8.2, uma outra opção é o resíduo quantílico (Dunn e Smyth, 1996) definido para variáveis discretas por

$$r_{q_i} = \Phi^{-1}(u_i),$$

em que $\Phi(\cdot)$ denota a função de distribuição acumulada da $N(0,1)$ e u_i é um valor gerado no intervalo $(0,1)$ com base em $F(y_i;\hat{\boldsymbol{\beta}})$ (função de distribuição acumulada da distribuição discreta ajustada). Tem-se para n grande que os resíduos $r_{q_1},\ldots,r_{q_n}$ são independentes e igualmente distribuídos $N(0,1)$. Logo, o gráfico entre os quantis amostrais $r_{q_{(1)}} \leq \cdots \leq r_{q_{(n)}}$ contra os quantis teóricos da normal padrão é recomendado para avaliar afastamentos da distribuição postulada para a resposta. Esse resíduo é disponibilizado na biblioteca `GAMLSS` do R (ver, por exemplo, Stasinopoulos et al., 2017) e como é aleatorizado para variáveis discretas, uma sugestão é gerar m gráficos do `worm plot` (gráfico entre $r_{q_{(i)}} - \mathrm{E}(Z_{(i)})$ contra $\mathrm{E}(Z_{(i)})$) para avaliar com mais segurança a adequação do ajuste.

Por exemplo, se o ajuste é armazenado no arquivo `fit`, a geração do resíduo quantílico e de m gráficos do `worm plot` podem ser obtidos por meio dos comandos

```
plot(fit)
rqres.plot(fit, howmany=8, type=''wp'').
```

A Figura 5.1 apresenta o gráfico normal de probabilidades para o resíduo t_{D_i} correspondente ao modelo ajustado aos dados da Tabela 5.2. Como pode-se notar, todos os resíduos cairam dentro do envelope gerado sem apresen-

tarem nenhuma tendência sistemática, indicando que a suposição de distribuição de Poisson parece ser bastante razoável. O programa utilizado para gerarmos o gráfico de envelopes é apresentado no Apêndice B.

5.3.6 Aplicação

Como ilustração considere os dados apresentados em Neter et al. (1996, p. 613) sobre o perfil dos clientes de uma determinada loja oriundos de 110 áreas de uma cidade. O objetivo do estudo é relacionar o número esperado de clientes em cada área com as seguintes variáveis explicativas em cada área: número de domicílios (em mil), renda média anual (em mil USD), idade média dos domicílios (em anos), distância ao concorrente mais próximo (em milhas) e distância à loja (em milhas). Portanto, a área é a unidade experimental. Esses dados estão também descritos no arquivo **store.txt**.

Na Figura 5.2 são apresentados os diagramas de dispersão entre o número de clientes (variável resposta) e as variáveis explicativas renda e idade média, distância ao concorrente mais próximo (dist1) e distância à loja (dist2). Indícios mais evidentes de relação linear podem ser observados entre a resposta e as distâncias dist1 e dist2. Ou seja, há indícios de que o número de clientes aumenta à medida que a distância ao concorrente mais próximo aumenta e a distância à loja diminui.

Denote por Y_i o número de clientes da i-ésima área que foram à loja no período determinado. Supor que $Y_i \overset{\text{ind}}{\sim} \text{P}(\mu_i)$ com parte sistemática dada por

$$\log(\mu_i) = \alpha + \beta_1 \text{domic}_i + \beta_2 \text{renda}_i + \beta_3 \text{idade}_i + \beta_4 \text{dist1}_i + \beta_5 \text{dist2}_i.$$

Tem-se que a variável número de domicílios (domic) deve ser incluída no modelo uma vez que as áreas não têm o mesmo número de domicílios. As estimativas dos parâmetros são apresentadas na Tabela 5.5 e como pode-se

notar todas as estimativas são altamente significativas. O desvio do modelo foi de $\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}) = 114,98$ (104 graus de liberdade) que equivale a um nível descritivo P$= 0,35$ indicando um ajuste adequado. Nota-se pela tabela que o número esperado de clientes na loja cresce com o aumento do número de domicílios na área e da distância ao concorrente mais próximo, porém diminui com o aumento da renda média e da idade média dos domicílios bem como da distância da área à loja. Isso sugere que deve ser uma loja de conveniência.

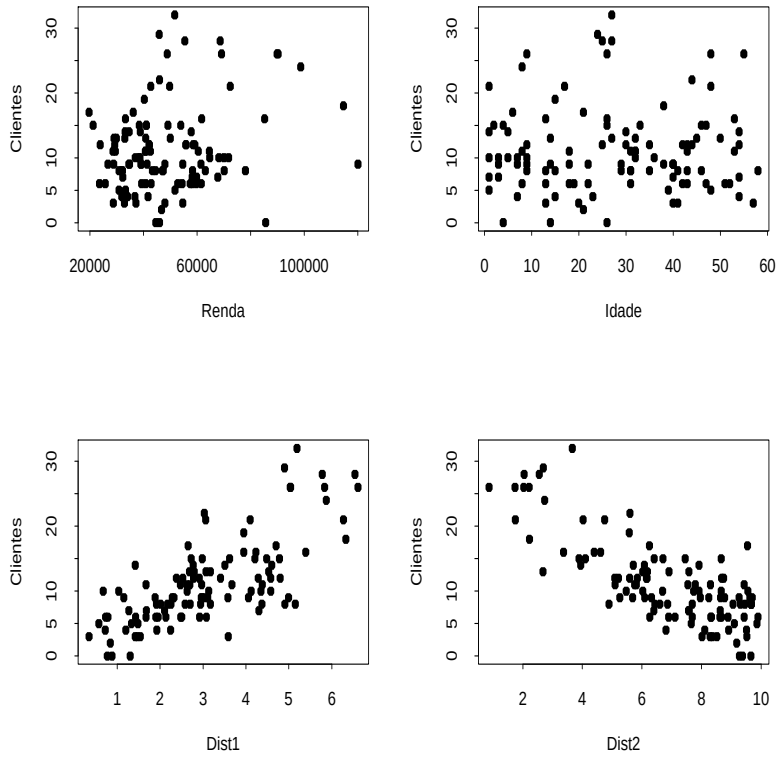

Figura 5.2: Diagramas de dispersão entre o número de clientes que visitaram a loja e algumas variáveis explicativas.

Tabela 5.5
Estimativas dos parâmetros do modelo log-linear de Poisson ajustado aos dados sobre perfil de clientes.

Efeito	Parâmetro	Estimativa	E/E.Padrão
Constante	α	2,942	14,21
Domicílio	β_1	0,606	4,27
Renda	β_2	-0,012	-5,54
Idade	β_3	-0,004	-2,09
Dist1	β_4	0,168	6,54
Dist2	β_5	-0,129	-7,95

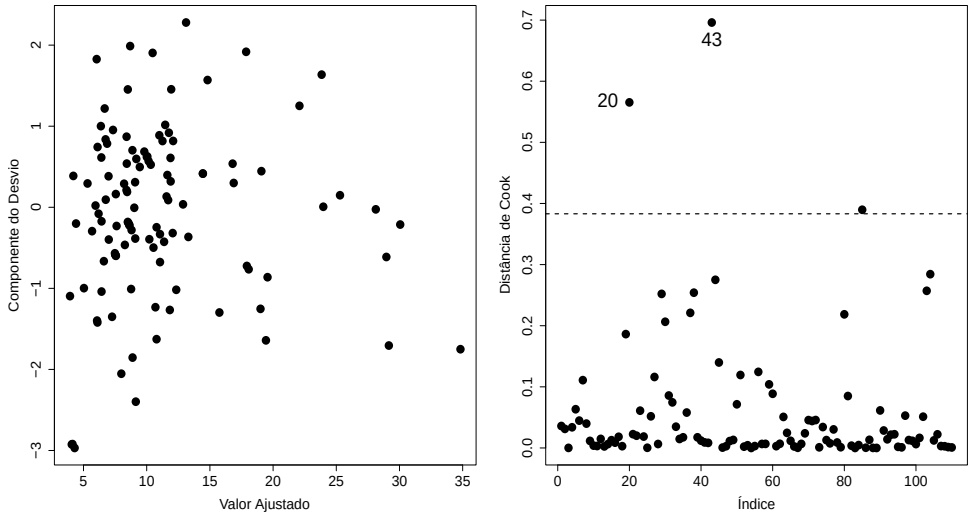

Figura 5.3: Gráficos de diagnóstico referentes ao modelo log-linear de Poisson ajustado aos dados sobre perfil de clientes.

Pode-se fazer algumas interpretações. Por exemplo, aumentando-se em 1

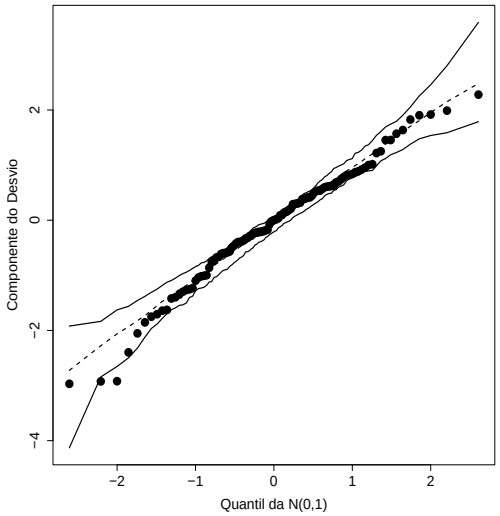

Figura 5.4: Gráfico normal de probabilidades referente ao modelo log-linear de Poisson ajustado aos dados sobre perfil de clientes.

mil USD a renda média dos domicílios de uma determinada área espera-se aumento relativo no número de clientes que irão à loja de $\exp(-0,012) = 0,988$. Ou seja, decrescimento de 1,2%, com estimativa intervalar de 95% dada por [0,8%,1,2%]. Por outro lado, se a distância ao concorrente mais próximo aumentar em uma milha espera-se aumento relativo no número de clientes de $\exp(0,168) = 1,183$. Ou seja, aumento de 18,3% com estimativa intervalar de 95% de [15%, 20%]. Pela Figura 5.3 nota-se que os resíduos estão bem comportados com o valor ajustado, sugerindo que a variabilidade foi controlada. A distância de Cook destaca as áreas #20 e #43, que apresentam algumas variações desproporcionais nas estimativas dos parâmetros, porém sem ocorrência de mudança inferencial. O gráfico normal de probabilidades (Figura 5.4) não apresenta indicações de afastamentos da suposição de distribuição de Poisson para o número de clientes que visitaram a loja no período.

5.4 Modelos com resposta binomial negativa

5.4.1 Distribuição binomial negativa

O fenômeno de sobredispersão, similarmente ao caso de dados com resposta binária discutido na Seção 4.9, ocorre quando é esperada uma distribuição de Poisson para a resposta, porém a variância é maior do que a resposta média. Uma causa provável desse fenômeno é a heterogeneidade das unidades amostrais que pode ser devido à variabilidades interunidades experimentais. Isso pode ser visto, por exemplo, supondo que para um conjunto fixo $\mathbf{x} = (x_1, \ldots, x_p)^\top$ de valores de variáveis explicativas, $Y|z$ tem média z e variância z, no entanto Z, que é não observável, varia nas unidades amostrais com $\mathbf{x}$ fixo, de modo que $\mathrm{E}(Z) = \mu$. Então,

$$\mathrm{E}(Y) = \mathrm{E}[\mathrm{E}(Y|Z)] = \mathrm{E}[Z] = \mu \quad \text{e}$$

$$\begin{aligned} \mathrm{Var}(Y) &= \mathrm{E}[\mathrm{Var}(Y|Z)] + \mathrm{Var}[\mathrm{E}(Y|Z)] \\ &= \mu + \mathrm{Var}(Z). \end{aligned}$$

Pode-se, adicionalmente, supor que $Y|z$ tem distribuição de Poisson com média z e função de probabilidade denotada por $f(y|z)$ e que Z segue uma distribuição gama de média μ e parâmetro de dispersão $k = \phi\mu$ cuja função de densidade será denotada por $g(z; \mu, k)$.

Tem-se $\mathrm{E}(Z) = \mu$ e $\mathrm{Var}(Z) = \mu^2/k$ de modo que $\mathrm{E}(Y) = \mu$ e $\mathrm{Var}(Y) = \mu + \mu^2/k = \mu(1 + \phi)/\phi$. Assim, as funções densidades $f(y|z)$ e $g(z; \mu, k)$ assumem as seguintes formas:

$$f(y|z) = \frac{e^{-z} z^y}{y!} \quad \text{e} \quad g(z; \mu, k) = \frac{1}{\Gamma(k)} \left(\frac{zk}{\mu}\right)^k e^{-\frac{kz}{\mu}} \frac{1}{z}.$$

Logo, Y tem função de probabilidade dada por

$$\begin{aligned} Pr\{Y=y\} &= \int_0^\infty f(y|z)g(z;\mu,k)dz \\ &= \frac{1}{y!\Gamma(k)}\left(\frac{k}{\mu}\right)^k \int_0^\infty e^{-z(1+k/\mu)}z^{k+y-1}dz. \end{aligned}$$

Fazendo a transformação de variável $t = z(1+\frac{k}{\mu})$ tem-se que $\frac{dz}{dt} = (1+\frac{k}{\mu})^{-1}$. Então,

$$\begin{aligned} Pr\{Y=y\} &= \frac{1}{y!\Gamma(k)}\left(\frac{k}{\mu}\right)^k \left(1+\frac{k}{\mu}\right)^{-(k+y)} \int_0^\infty e^{-t}t^{k+y-1}dt \\ &= \frac{\Gamma(y+k)\phi^k}{\Gamma(y+1)\Gamma(k)(1+\phi)^{y+k}} \\ &= \frac{\Gamma(y+k)}{\Gamma(y+1)\Gamma(k)}\left(\frac{\phi}{1+\phi}\right)^k\left(\frac{1}{1+\phi}\right)^y \\ &= \frac{\Gamma(y+k)}{\Gamma(y+1)\Gamma(k)}(1-\pi)^k\pi^y, \quad y=0,1,2,\ldots, \end{aligned}$$

em que $\pi = 1/(1+\phi)$. Portanto, Y tem distribuição binomial negativa de média μ e parâmetro de forma k.

Pode-se, similarmente, supor que $Y|z \sim \mathrm{P}(z)$ e que $Z \sim \mathrm{G}(\mu,\phi)$, em que ϕ não depende de μ. Nesse caso $\mathrm{E}(Z) = \mu$ e $\mathrm{Var}(Z) = \mu^2/\phi$ de onde segue que $\mathrm{E}(Y) = \mu$ e $\mathrm{Var}(Y) = \mu + \mu^2/\phi$. Tem-se então que

$$f(y|z) = \frac{e^{-z}z^y}{y!} \quad \text{e} \quad g(z;\mu,\phi) = \frac{1}{\Gamma(\phi)}\left(\frac{z\phi}{\mu}\right)^\phi e^{-\frac{\phi z}{\mu}}\frac{1}{z}.$$

A função de probabilidade de Y fica dada por

$$\begin{aligned} Pr\{Y=y\} &= \int_0^\infty f(y|z)g(z;\mu,\phi)dz \\ &= \frac{1}{y!\phi}\left(\frac{\phi}{\mu}\right)^\phi \int_0^\infty e^{-z(1+\phi/\mu)}z^{\phi+y-1}dz. \end{aligned}$$

Fazendo a transformação de variável $t = z(1+\frac{\phi}{\mu})$ tem-se que $\frac{dz}{dt} = (1+\frac{\phi}{\mu})^{-1}$. Daí segue que

$$\begin{aligned} Pr\{Y=y\} &= \frac{1}{y!\Gamma(\phi)}\left(\frac{\phi}{\mu}\right)^{\phi}\left(1+\frac{\phi}{\mu}\right)^{-(\phi+y)}\int_0^{\infty} e^{-t}t^{\phi+y-1}dt \\ &= \frac{\Gamma(\phi+y)\mu^y\phi^{\phi}}{\Gamma(\phi)\Gamma(y+1)(\mu+\phi)^{\phi+y}} \\ &= \frac{\Gamma(\phi+y)}{\Gamma(y+1)\Gamma(\phi)}\left(\frac{\mu}{\mu+\phi}\right)^{y}\left(\frac{\phi}{\mu+\phi}\right)^{\phi} \\ &= \frac{\Gamma(\phi+y)}{\Gamma(y+1)\Gamma(\phi)}(1-\pi)^{\phi}\pi^{y}, \quad y=0,1,2,\ldots, \end{aligned}$$

com $\pi = \mu/(\mu+\phi)$. Portanto, neste caso Y também segue distribuição binomial negativa de média μ e parâmetro de forma ϕ. Será denotado $Y \sim \text{BN}(\mu,\phi)$. Pode-se mostrar (ver, por exemplo, Jørgensen, 1997, p. 96) que

$$\frac{1}{\sqrt{\phi}}(Y-\mu) \ \to_d \ \text{N}(0,\pi/(1-\pi)^2), \quad \text{quando} \ \ \phi\to\infty.$$

Pode-se obter também aproximações da binomial negativa para a Poisson e gama.

5.4.2 Modelos de regressão com resposta binomial negativa

Supor então que $Y_1,\ldots,Y_n$ são variáveis aleatórias independentes tais que $Y_i \sim \text{BN}(\mu_i,\phi)$. A função de probabilidade de Y_i fica dada por

$$f(y_i;\mu_i,\phi) = \frac{\Gamma(\phi+y_i)}{\Gamma(y_i+1)\Gamma(\phi)}\left(\frac{\mu_i}{\mu_i+\phi}\right)^{y_i}\left(\frac{\phi}{\mu_i+\phi}\right)^{\phi}, \quad y_i=0,1,2,\ldots.$$

Tem-se que $\text{E}(Y_i)=\mu_i$ e $\text{Var}(Y_i)=\mu_i+\mu_i^2/\phi$. Similarmente aos MLGs a parte sistemática será denotada por $g(\mu_i)=\eta_i=\mathbf{x}_i^{\top}\boldsymbol{\beta}$, em que $\mathbf{x}_i=(x_{i1},\ldots,x_{ip})^{\top}$ contém valores de variáveis explicativas, $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_p)^{\top}$ é um vetor de parâmetros desconhecidos e $g(\cdot)$ é a função de ligação. Como nos modelos

de Poisson as ligações mais utilizados são logarítmica ($g(\mu_i) = \log(\mu_i)$, raiz quadrada ($g(\mu_i) = \sqrt{\mu_i}$) e identidade ($g(\mu_i) = \mu_i$).

Definindo $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$ o logaritmo da função de verossimilhança fica dado por

$$\mathrm{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\log\left\{ \frac{\Gamma(\phi + y_i)}{\Gamma(y_i + 1)\Gamma(\phi)} \right\} + \phi \log(\phi) + y_i \log(\mu_i) - (\phi + y_i) \log(\mu_i + \phi) \right],$$

em que $\mu_i = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$. A fim de obter-se a função escore para $\boldsymbol{\beta}$ obtém-se inicialmente as derivadas

$$\begin{aligned}
\partial \mathrm{L}(\boldsymbol{\theta})/\partial \beta_j &= \sum_{i=1}^{n} \left\{ \frac{y_i}{\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \right\} \\
&= \sum_{i=1}^{n} \left\{ \frac{y_i}{\mu_i} \frac{d\mu_i}{d\eta_i} x_{ij} - \frac{(\phi + y_i)}{(\phi + \mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij} \right\} \\
&= \sum_{i=1}^{n} \left\{ \frac{\phi (d\mu_i/d\eta_i)}{\mu_i(\phi + \mu_i)} (y_i - \mu_i) x_{ij} \right\} \\
&= \sum_{i=1}^{n} \omega_i f_i^{-1} (y_i - \mu_i) x_{ij},
\end{aligned}$$

em que $\omega_i = (d\mu_i/d\eta_i)^2/(\mu_i^2 \phi^{-1} + \mu_i)$ e $f_i = d\mu_i/d\eta_i$. Logo, pode-se espressar a função escore na forma matricial

$$\mathbf{U}_\beta(\boldsymbol{\theta}) = \mathbf{X}^\top \mathbf{W} \mathbf{F}^{-1} (\mathbf{y} - \boldsymbol{\mu}), \tag{5.5}$$

em que $\mathbf{X}$ é a matriz modelo com linhas $\mathbf{x}_i^\top$, $i = 1, \ldots, n$, $\mathbf{W} = \mathrm{diag}\{\omega_1, \ldots, \omega_n\}$, $\mathbf{F} = \mathrm{diag}\{f_1, \ldots, f_n\}$, $\mathbf{y} = (y_1, \ldots, y_n)^\top$ e $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$. De forma similar a função escore para ϕ fica dada por

$$\mathrm{U}_\phi(\boldsymbol{\theta}) = \sum_{i=1}^{n} [\psi(\phi + y_i) - \psi(\phi) - (y_i + \phi)/(\phi + \mu_i) + \log\{\phi/(\phi + \mu_i)\} + 1], \tag{5.6}$$

em que $\psi(\cdot)$ é a função digama.

Para obter-se a matriz de informação de Fisher calcula-se as derivadas

$$\begin{aligned}\partial^2 \mathrm{L}(\boldsymbol{\theta})/\partial\beta_j\partial\beta_\ell &= -\sum_{i=1}^{n}\left\{\frac{(\phi+y_i)}{(\phi+\mu_i)^2}-\frac{y_i}{\mu_i^2}\right\}\left(\frac{d\mu_i}{d\eta_i}\right)^2 x_{ij}x_{i\ell} \\ &\quad +\sum_{i=1}^{n}\left\{\frac{y_i}{\mu_i}-\frac{(\phi+y_i)}{(\phi+\mu_i)}\right\}\frac{d^2\mu_i}{d\eta_i^2}x_{ij}x_{i\ell},\end{aligned}$$

cujos valores esperados ficam dados por

$$\begin{aligned}\mathrm{E}\{\partial^2 \mathrm{L}(\boldsymbol{\theta})/\partial\beta_j\partial\beta_\ell\} &= -\sum_{i=1}^{n}\frac{\phi(d\mu_i/d\eta_i)^2}{(\phi+\mu_i)}x_{ij}x_{i\ell} \\ &= -\sum_{i=1}^{n}\omega_i x_{ij}x_{i\ell}.\end{aligned}$$

Logo, pode-se expressar a informação de Fisher para $\boldsymbol{\beta}$ em forma matricial

$$\mathbf{K}_{\beta\beta}(\boldsymbol{\theta}) = \mathrm{E}\left\{-\frac{\partial^2 \mathrm{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right\} = \mathbf{X}^\top\mathbf{W}\mathbf{X}.$$

Lawless (1987) mostra que a informação de Fisher para ϕ pode ser expressa na forma

$$\mathrm{K}_{\phi\phi}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\{\sum_{j=0}^{\infty}(\phi+j)^{-2}Pr(Y_i\geq j)-\phi^{-1}\mu_i/(\mu_i+\phi)\},$$

e que $\boldsymbol{\beta}$ e ϕ são parâmetros ortogonais. Assim , a matriz de informação de Fisher para $\boldsymbol{\theta}$ assume a forma bloco diagonal

$$\mathbf{K}_{\theta\theta} = \begin{bmatrix}\mathbf{K}_{\beta\beta} & \mathbf{0}\\ \mathbf{0} & \mathrm{K}_{\phi\phi}\end{bmatrix}.$$

As estimativas de máxima verossimilhaça para $\boldsymbol{\beta}$ e ϕ podem ser obtidas através de um algoritmo de mínimos quadrados reponderados, aplicando o método escore de Fisher, a partir de (5.5) e do método de Newton-Raphson para obter $\hat{\phi}$ desenvolvido a partir de (5.6), os quais são descritos abaixo

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top\mathbf{W}^{(m)}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{W}^{(m)}\mathbf{y}^{*(m)}$$

e

$$\phi^{(m+1)} = \phi^{(m)} - \{\mathrm{U}_{\phi}^{(m)}/\ddot{\mathrm{L}}_{\phi\phi}^{(m)}\},$$

para $m = 0, 1, 2, \ldots$, em que

$$\mathbf{y}^{*} = \mathbf{X}\boldsymbol{\beta} + \mathbf{F}^{-1}(\mathbf{y} - \boldsymbol{\mu})$$

é uma variável dependente modificada e

$$\ddot{\mathrm{L}}_{\phi\phi} = \sum_{i=1}^{n}\{\psi'(\phi + y_i) + (y_i - 2\mu_i - \phi)/(\phi + \mu_i)^2\} + n\phi^{-1}\{1 - \phi\psi'(\phi)\}.$$

Tabela 5.5
Quantidades ω_i e f_i para algumas ligações.

Ligação	ω_i	f_i
$\log(\mu_i) = \eta_i$	$\mu_i/(\mu_i\phi^{-1} + 1)$	μ_i
$\mu_i = \eta_i$	$(\mu_i^2\phi^{-1} + \mu_i)^{-1}$	1
$\sqrt{\mu_i} = \eta_i$	$4/(\mu_i\phi^{-1} + 1)$	$2\sqrt{\mu_i}$

Os dois procedimentos são aplicados simultaneamente até a convergência. Pode-se encontrar as estimativas de máxima verossimilhança $(\hat{\boldsymbol{\beta}}^{\top}, \hat{\phi})^{\top}$ pela aplicação do comando `library(MASS)` do `R`. Como ilustração, supor um modelo log-linear com resposta binomial negativa `resp` e covariáveis `cov1` e `cov2`. Deve-se acionar os seguintes comandos no `R`:

```
library(MASS)
fit.bn = glm.nb( resp ~ cov1 + cov2).
```

No objeto `fit.bn` estarão os resultados do ajuste. Outras ligações, além da ligação logarítmica, podem ser usadas com a distribuição binomial negativa. Por exemplo, para o ajuste de um modelo com resposta binomial negativa e ligação identidade se `resp` é considerada resposta e `cov1` e `cov2` são consideradas variáveis explicativas, deve-se fazer o seguinte:

```
library(MASS)
```

`fit.bn = glm.nb( resp ~ cov1 + cov2, link=identity).`

A Tabela 5.5 apresenta as expressões para ω_i e f_i para algumas ligações usuais em modelos com resposta binomial negativa.

Usando os mesmos argumentos da Seção 2.6 tem-se que para n grande $\hat{\boldsymbol{\beta}}$ segue distribuição aproximadamente normal p-variada de média $\boldsymbol{\beta}$ e matriz de variância-covariância $\mathbf{K}_{\beta\beta}^{-1}$, ou seja , para n grande $\hat{\boldsymbol{\beta}} \sim \mathrm{N}_p(\boldsymbol{\beta}, \mathbf{K}_{\beta\beta}^{-1})$. Similarmente para n grande $\hat{\phi} \sim \mathrm{N}(\phi, \mathrm{K}_{\phi\phi}^{-1})$. Além disso, $\hat{\boldsymbol{\beta}}$ e $\hat{\phi}$ são assintoticamente independentes.

5.4.3 Qualidade do ajuste

A função desvio assumindo ϕ fixo fica dada por

$$\mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\sum_{i=1}^{n}\left[\phi \log\left\{\frac{\hat{\mu}_i + \phi}{y_i + \phi}\right\} + y_i \log\left\{\frac{y_i(\hat{\mu}_i + \phi)}{\hat{\mu}_i(y_i + \phi)}\right\}\right],$$

em que $\hat{\mu}_i = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$. Quando $y_i = 0$ o i-ésimo componente da função desvio $\mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ fica dado por

$$\begin{aligned} d^{*2}(y_i, \hat{\mu}_i) &= 2[\log\{f(0; y_i, \phi)\} - \log\{f(0; \hat{\mu}_i, \phi)\}] \\ &= 2\phi \log\{\phi/(y_i + \phi)\} - 2\phi \log\{\phi/(\hat{\mu}_i + \phi)\} \\ &= 2\phi \log\{(\mu_i + \phi)/(\hat{y}_i + \phi)\} \\ &= 2\phi \log\{(\hat{\mu}_i + \phi)/\phi\}. \end{aligned}$$

Portanto, os componentes do desvio no caso binomial negativo assumem as seguintes formas:

$$d^{*2}(y_i; \hat{\mu}_i) = \begin{cases} 2\left[\phi \log\left\{\frac{(\hat{\mu}_i+\phi)}{(y_i+\phi)}\right\} + y_i \log\left\{\frac{y_i(\hat{\mu}_i+\phi)}{\hat{\mu}_i(y_i+\phi)}\right\}\right] & \text{se } \ y_i > 0; \\ 2\phi \log\left\{\frac{(\hat{\mu}_i+\phi)}{\phi}\right\} & \text{se } \ y_i = 0. \end{cases}$$

Sob a hipótese de que o modelo adotado está correto $\mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ segue para ϕ grande e μ_i grande, $\forall i$, distribuição qui-quadrado com $(n - p)$ graus de liberdade.

Supor agora a partição $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ em que $\boldsymbol{\beta}_1$ é um vetor q-dimensional enquanto $\boldsymbol{\beta}_2$ tem dimensão $p - q$ e que ϕ é fixo ou conhecido. O teste da razão de verossimilhanças para testar $\mathrm{H}_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ contra $\mathrm{H}_1 : \boldsymbol{\beta}_1 \neq \mathbf{0}$ reduz, neste caso, à diferença entre dois desvios

$$\xi_{RV} = \mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - \mathrm{D}^*(\mathbf{y}; \hat{\boldsymbol{\mu}}),$$

em que $\hat{\boldsymbol{\mu}}^0$ e $\hat{\boldsymbol{\mu}}$ são, respectivamente, as estimativas de $\boldsymbol{\mu}$ sob H_0 e H_1. Para ϕ desconhecido o teste da razão de verossimilhanças fica expresso na seguinte forma:

$$\begin{aligned}\xi_{RV} &= 2\sum_{i=1}^{n}[\log\{\Gamma(\hat{\phi}+y_i)\Gamma(\hat{\phi}^0)/\Gamma(\hat{\phi}^0+y_i)\Gamma(\hat{\phi})\} + \hat{\phi}\log\{\hat{\phi}/(\hat{\phi}+\hat{\mu}_i)\} \\ &\quad -\hat{\phi}^0\log\{\hat{\phi}^0/(\hat{\phi}^0+\hat{\mu}_i^0)\} + y_i\log\{\hat{\mu}_i(\hat{\phi}^0+\hat{\mu}_i^0)/\hat{\mu}_i^0(\hat{\phi}+\hat{\mu}_i)\}],\end{aligned}$$

em que $\hat{\phi}^0$ e $\hat{\phi}$ são as estimativas de máxima verossimilhança de ϕ sob H_0 e H_1, respectivamente. Para n grande e sob H_0 tem-se que $\xi_{RV} \sim \chi^2_q$.

5.4.4 Técnicas de diagnóstico

Fazendo uma analogia com os MLGs a matriz de projeção $\mathbf{H}$ assume aqui a seguinte forma:

$$\mathbf{H} = \mathbf{W}^{\frac{1}{2}}\mathbf{X}(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{W}^{\frac{1}{2}}.$$

O i-ésimo elemento da diagonal principal de $\mathbf{H}$ fica dado por

$$h_{ii} = \frac{(d\mu_i/d\eta_i)^2}{(\mu_i\phi^{-1}+\mu_i)}\mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{x}_i.$$

Em particular, para os modelos log-lineares h_{ii} fica dado por

$$h_{ii} = \frac{\phi\mu_i}{(\phi+\mu_i)}\mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{x}_i,$$

em que $\omega_i = \phi\mu_i/(\phi+\mu_i)$. Como $\hat{h}_{ii}$ deverá depender de $\hat{\mu}_i$, gráficos de $\hat{h}_{ii}$ contra os valores ajustados são mais informativos do que os gráficos de $\hat{h}_{ii}$ contra a ordem das observações.

Estudos de Monte Carlo desenvolvidos por Svetliza (2002) indicam boa concordância entre o resíduo componente do desvio

$$t_{D_i} = \frac{d^*(y_i;\hat{\mu}_i)}{\sqrt{1-\hat{h}_{ii}}}$$

com a distribuição normal padrão, em que

$$d^*(y_i;\hat{\mu}_i) = \begin{cases} \pm\sqrt{2}\left[\phi\log\left\{\frac{\hat{\mu}_i+\phi}{y_i+\phi}\right\} + y_i\log\left\{\frac{y_i(\hat{\mu}_i+\phi)}{\hat{\mu}_i(y_i+\phi)}\right\}\right]^{\frac{1}{2}} & \text{se } \ y_i > 0; \\ \pm\sqrt{2}\left[\phi\log\left\{\frac{(\hat{\mu}_i+\phi)}{\phi}\right\}\right] & \text{se } \ y_i = 0. \end{cases}$$

Para extrair a quantidade $d_i^*(y_i;\hat{\mu}_i)$ do objeto `fit.bn` deve-se fazer o seguinte:

`d = resid(fit.bn, type= ‘‘deviance").`

Uma versão da distância de Cook aproximada é dada por

$$\mathrm{LD}_i = \frac{\hat{h}_{ii}}{(1-\hat{h}_{ii})^2}\hat{r}_{P_i}^2,$$

em que $r_{P_i} = (y_i-\mu_i)/\sqrt{\mathrm{Var}(Y_i)}$ e $\mathrm{Var}(Y_i) = \mu_i + \mu_i^2/\phi$. A quantidade r_{P_i} é obtida no `R` através do comando

`rp = resid(fit.bn, type=‘‘pearson").`

O gráfico de LD_i contra as observações ou valores ajustados pode revelar pontos influentes nas estimativas $\hat{\boldsymbol{\beta}}$ e $\hat{\phi}$. Svetliza (2002) desenvolveu as expressões matriciais para a obtenção de $\boldsymbol{\ell}_{max}$ para $\hat{\boldsymbol{\beta}}$ e $\hat{\phi}$.

5.4.5 Seleção de modelos

Similarmente aos modelos lineares generalizados, pelo critério de Akaike deve-se encontrar um submodelo para o qual a quantidade abaixo seja minimizada

$$\mathrm{AIC} = \mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) + 2p.$$

Deve-se acionar no `R` os seguintes comandos:

```
library(MASS)
stepAIC(fit.model).
```

Aqui `fit.model` denota o objeto com o modelo binomial negativo ajustado.

5.4.6 Aplicações

Estudantes australianos

Venables e Ripley(1999, Caps. 6 e 7) apresentam os resultados de um estudo sociológico desenvolvido na Austrália com 146 estudantes de 8ª série e ensino médio com o objetivo de comparar a ausência na escola segundo os seguintes fatores: ano que o estudante está cursando (1: 8ª série, 2: 1º ano do ensino médio, 3: 2º ano do ensino médio, 4: 3º ano do ensino médio), etnia (0: aborígine, 1: não aborígine), desempenho escolar (0: insuficiente, 1: suficiente) e sexo (0: masculino, 1: feminino). Para obter esses dados no `R` deve-se acionar o comando `library(MASS)` e em seguida `quine`. Uma cópia desses dados está disponível no arquivo **quine.txt**.

Denota-se por $Y_{ijk\ell m}$ o número de faltas num determinado período referentes ao m-ésimo aluno, cursando o i-ésimo ano, de etnia j, com desempenho escolar k e pertencente ao ℓ-ésimo sexo, em que $i = 1, 2, 3, 4$, $j, k, \ell = 1, 2$ e $m = 1, \ldots, 144$. Supor que $Y_{ijk\ell m} \overset{\text{ind}}{\sim} \text{BN}(\mu_{ijk\ell}, \phi)$, em que

$$\log(\mu_{ijk\ell}) = \alpha + \beta_i + \gamma_j + \delta_k + \theta_\ell,$$

com $\beta_1 = 0$, $\gamma_1 = 0$, $\delta_1 = 0$ e $\theta_1 = 0$. Assim, tem-se um modelo casela de referência com β_2, β_3 e β_4 denotando os incrementos do primeiro, segundo e terceiro anos do ensino médio, respectivamente, em relação à 8ª série, γ_2 é a diferença entre os efeitos do grupo não aborígine com relação ao grupo aborígine, δ_2 denota a diferença entre os efeitos dos grupos com desempenho

suficiente e insuficiente e θ é a diferença entre os efeitos do sexo feminino e masculino.

Tabela 5.6
Estimativas de máxima verossimilhança referentes ao modelo log-linear binomial negativo ajustado aos dados sobre ausência escolar de estudantes australianos.

Efeito	Modelo 1	E/E.Padrão	Modelo 2	E/E.Padrão
Intercepto	2,895	12,70	2,628	10,55
Etnia	-0,569	-3,72	0,131	0,38
Sexo	0,082	0,51		
Ano2	-0,448	-1,87	0,178	0,56
Ano3	0,088	0,37	0,827	2,61
Ano4	0,357	1,44	0,371	1,11
Desemp	0,292	1,57		
Etn*Ano2			-0,991	-2,26
Etn*Ano3			-1,239	-2,78
Etn*Ano4			-0,176	-0,38
ϕ	1,275	7,92	1,357	7,80

Na Tabela 5.6 tem-se as estimativas de máxima verossimilhança com os respectivos erros padrão aproximados. O desvio do modelo ajustado (modelo 1) foi de $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = 167,95$ (139 graus de liberdade). Nota-se que os fatores sexo e desempenho escolar não são significativos marginalmente ao nível de 10%. Após testar a ausência de efeito conjunto desses fatores, tem-se que ambos são conjuntamente não significativos sendo portanto retirados do modelo. Contudo, nota-se a necessidade de inclusão da interação entre etnia e ano no modelo. O valor da estatística do teste da razão de verossimilhanças nesse caso é de $\xi_{RV} = 11,16$ (P$= 0,0109$). As novas estimativas são também apresentadas na Tabela 5.6. O desvio do novo modelo (modelo 2) foi de $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = 167,84$ (138 graus de liberdade).

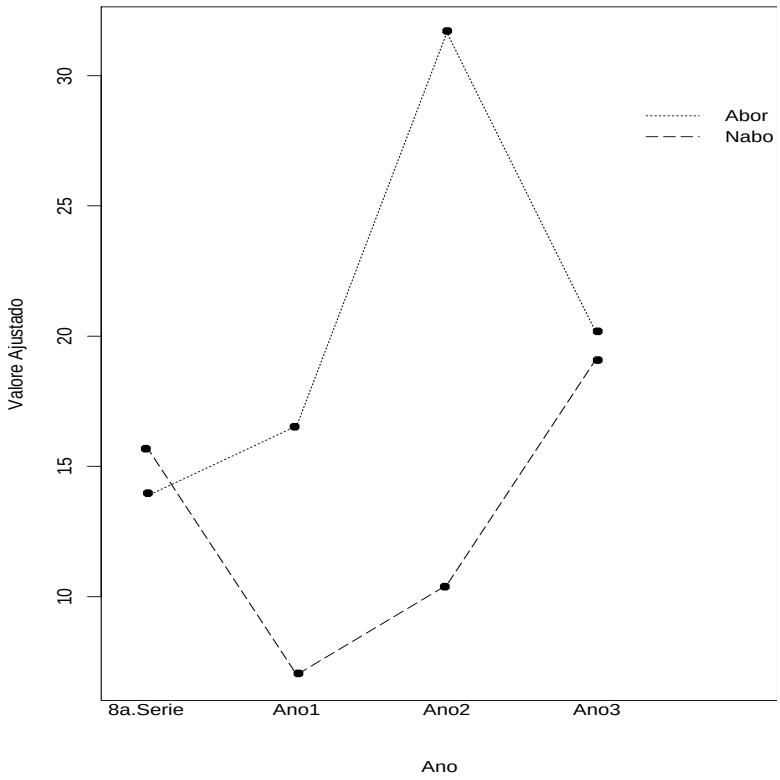

Figura 5.5: Valores médios estimados pelo modelo log-linear binomial negativo ajustado aos dados sobre ausência escolar de estudantes australianos.

A Figura 5.5 apresenta as médias ajustadas do modelo final com resposta binomial negativa. Pode-se notar que o grupo não aborígine tem em geral um nº médio menor de dias ausentes. A maior média é observada para estudantes do grupo aborígine cursando o 2º ano do ensino médio e o menor valor médio é observado para estudantes do grupo não aborígine cursando o 1º ano do ensino médio. Embora a interação entre etnia e ano seja significativa, não implica que para cada ano a diferença entre o número médio de faltas nos grupos aborígene e não aborígene seja significativa. Isso poderia ser avaliado através de testes de contrastes. A presença de interação significa que pelo menos uma das diferenças médias entre os dois grupos é significativa.

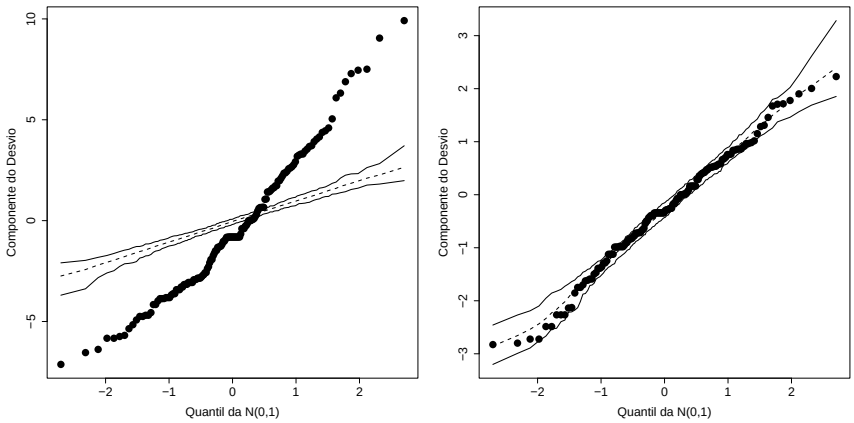

Figura 5.6: Gráficos normais de probabilidade referentes ao modelo log-linear de Poisson (esquerda) e ao modelo log-linear binomial negativo (direita) ajustados aos dados sobre ausência escolar de estudantes australianos.

Verifica-se também, neste estudo, como fica o ajuste através de um modelo log-linear de Poisson. Tem-se nas Figura 5.6 os gráficos normais de probabilidade para os dois ajustes e nota-se uma clara superioridade do modelo log-linear com resposta binomial negativa. O modelo log-linear de Poisson apresenta fortes indícios de sobredispersão com os resíduos cruzando o envelope gerado. Isso é justificado pelo valor do desvio $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = 1597,11$ (138 graus de liberdade).

Na Figura 5.7 são apresentados gráficos de diagnóstico referentes ao ajuste do modelo log-linear binomial negative. Nota-se que o resíduo componente do desvio se comporta de forma aleatória com o valor ajustado, indicando que a variabilidade foi controlada. Pelo gráfico da distância de Cook nota-se três pontos com mais destaque como possivelmente influentes em $\hat{\boldsymbol{\beta}}$, são os alunos #72, #104 e #36. Os três alunos têm vários dias ausentes, respectivamente,

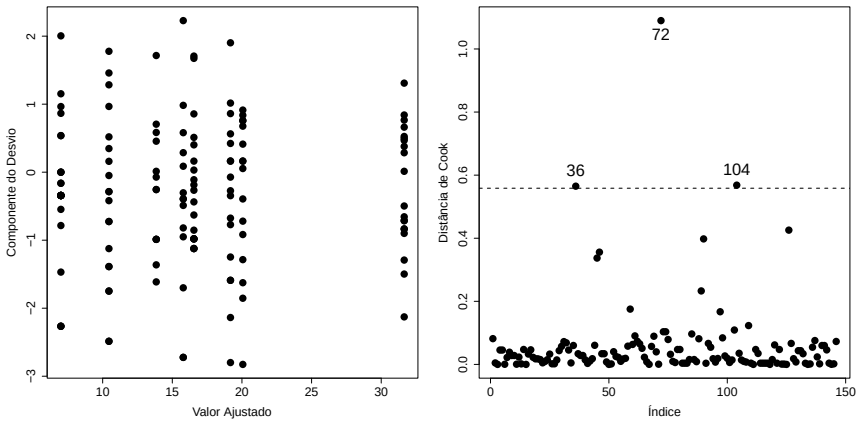

Figura 5.7: Gráficos de diagnóstico referentes ao modelo binomial negativo ajustado aos dados sobre ausência escolar de estudantes australianos.

67, 69 e 45. O aluno #72 é não aborígine e estava cursando a 8ª série. O aluno #104 é também não aborígine, porém estava cursando o 3º ano, enquanto o aluno #36 é aborígine e estava também cursando a 8ª série. A retirada desses pontos causa aumentos desproporcionais nas estimativas, mas não altera os resultados inferenciais.

Demanda de TV a cabo

Na Tabela 5.7 é apresentado um conjunto de dados sobre a demanda de TVs a cabo em 40 áreas metropolitanas dos EUA (Ramanathan, 1993). Esses dados estão também disponíveis no arquivo **tvcabo.txt**. Foram observadas, para cada área, o número de assinantes (em milhares) de TV a cabo (nass), o número de domicílios (em milhares) na área (domic), a porcentagem de domicílios com TV a cabo (perc), a renda per capita (em mil USD) por domicílio com TV a cabo (percap), a taxa de instalação de TV a cabo (taxa) em USD, o custo médio mensal de manutenção de TV a cabo (custo) em

Tabela 5.7

Demanda de TV a cabo em 40 áreas metropolitanas dos EUA.

Nass	Domic	Perc	Percap	Taxa	Custo	Ncabo	Ntv
105	350	30,000	9,839	14,95	10	16	13
90	255,631	35,207	10,606	15	7,5	15	11
14	31	45,161	10,455	15	7	11	9
11,7	34,840	33,582	8,958	10	7	22	10
46	153,434	29,980	11,741	25	10	20	12
11,217	26,621	42,136	9,378	15	7,66	18	8
12	18	66,667	10,433	15	7,5	12	8
6,428	9,324	68,940	10,167	15	7	17	7
20,1	32	62,813	9,218	10	5,6	10	8
8,5	28	30,357	10,519	15	6,5	6	6
1,6	8	20,000	10,025	17,5	7,5	8	6
1,1	5	22,000	9,714	15	8,95	9	9
4,355	15,204	28,644	9,294	10	7	7	7
78,910	97,889	80,612	9,784	24,95	9,49	12	7
19,6	93	21,075	8,173	20	7,5	9	7
1	3	33,333	8,967	9,95	10	13	6
1,65	2,6	63,462	10,133	25	7,55	6	5
13,4	18,284	73,288	9,361	15,5	6,3	11	5
18,708	55	34,015	9,085	15	7	16	6
1,352	1,7	79,529	10,067	20	5,6	6	6
170	270	62,963	8,908	15	8,75	15	5
15,388	46,540	33,064	9,632	15	8,73	9	6
6,555	20,417	32,106	8,995	5,95	5,95	10	6
40	120	33,333	7,787	25	6,5	10	5
19,9	46,39	42,897	8,890	15	7,5	9	7
2,45	14,5	16,897	8,041	9,95	6,25	6	4
3,762	9,5	39,600	8,605	20	6,5	6	5
24,882	81,98	30,351	8,639	18	7,5	8	4
21,187	39,7	53,368	8,781	20	6	9	4
3,487	4,113	84,780	8,551	10	6,85	11	4
3	8	37,500	9,306	10	7,95	9	6
42,1	99,750	42,206	8,346	9,95	5,73	8	5
20,350	33,379	60,966	8,803	15	7,5	8	4
23,15	35,5	65,211	8,942	17,5	6,5	8	5
9,866	34,775	28,371	8,591	15	8,25	11	4
42,608	64,840	65,713	9,163	10	6	11	6
10,371	30,556	33,941	7,683	20	7,5	8	6
5,164	16,5	31,297	7,924	14,95	6,95	8	5
31,150	70,515	44,175	8,454	9,95	7	10	4
18,350	42,040	43,649	8,429	20	7	6	4

USD, o número de canais a cabo disponíveis na área (ncabo) e o número de canais não pagos com sinal de boa qualidade disponíveis na área (ntv). Como são dados de contagem pode-se pensar inicialmente num modelo de Poisson em que nass_i denota o número de assinantes na i-ésima região, $\text{nass}_i \overset{\text{ind}}{\sim} \text{P}(\mu_i)$, e componente sistemático dado por $\log(\mu_i) = \alpha + \beta_1 \text{domic}_i + \beta_2 \text{percap}_i + \beta_3 \text{taxa}_i + \beta_4 \text{custo}_i + \beta_5 \text{ncabo}_i + \beta_6 \text{ntv}_i$, para $i = 1, \ldots, 40$. No entanto, o ajuste do modelo forneceu desvio $\text{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 225$ para 33 graus de liberdade indicando fortes indícios de sobredispersão, que é confirmado pelo gráfico normal de probabilidades da Figura 5.8. Então um modelo log-linear com resposta binomial negativa foi ajustado, em que $\text{nass}_i \overset{\text{ind}}{\sim} \text{BN}(\mu_i, \phi)$.

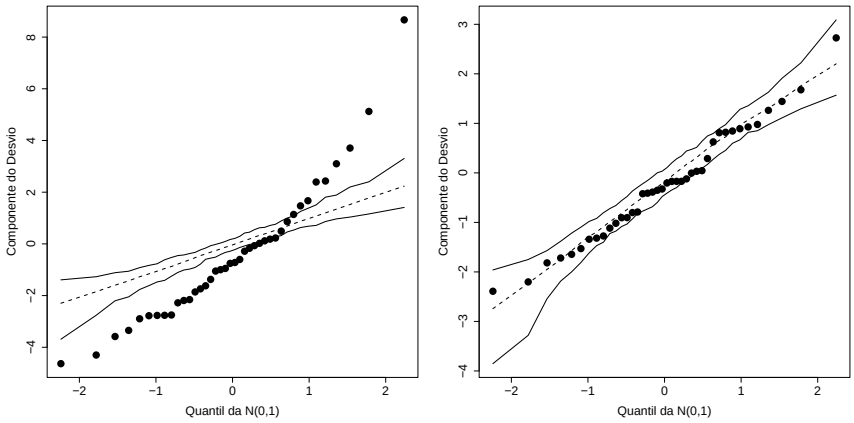

Figura 5.8: Gráficos normais de probabilidade referentes aos modelos log-linear de Poisson (esquerda) e log-linear binomial negativo (direita) ajustados aos dados sobre demanda de TV a cabo.

O gráfico normal de probabilidades (Figura 5.8) bem como o desvio

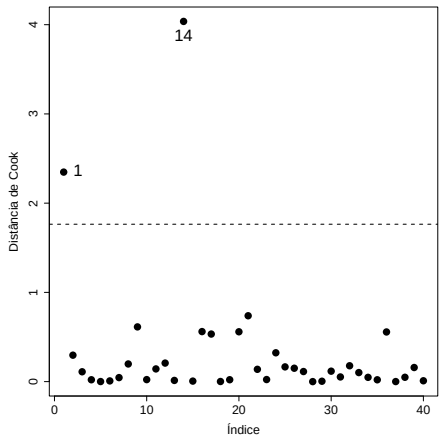

Figura 5.9: Gráfico de diagnóstico referente ao modelo log-linear binomial negativo ajustado aos dados sobre demanda de TV a cabo.

$\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = 42,35$ fornecem indícios de ajuste adequado do modelo log-linear com resposta binomial negativa. No entanto, pela Figura 5.9, nota-se uma área suspeita de ser altamente influente (observação #14) nas estimativas dos coeficientes e outra área com suspeita de moderada influência (observação #1). A área #14 apresenta custos altos de instalação e manutenção de TV a cabo, porém um alto índice de assinantes. A área #1 tem um baixo índice de assinantes com grande oferta de canais a cabo e canais não pagos de boa qualidade. As estimativas dos coeficientes com todos os pontos e eliminando as observações mais discrepantes (1 e 14) são apresentadas na Tabela 5.8. Como pode-se observar há indícios de que quatro coeficientes (percap, taxa, ncabo e ntv) são marginalmente não significativos a 10%. Aplica-se então o teste da razão de verossimilhanças para testar $\mathrm{H}_0 : \beta_2 = \beta_3 = \beta_5 = \beta_6 = 0$ contra pelo menos um parâmetro diferente de zero que forneceu o valor $\xi_{RV} = 2,50$ para 4 graus de liberdade (P=0,64), indicando pela não rejeição da hipótese nula. Isso significa que as duas observações discrepantes são responsáveis

pela significância de três desses coeficientes que aparecem significativos marginalmente com todos os pontos, bem como pelo aumento da sobredispersão uma vez que a estimativa de ϕ cresce com a eliminação das duas áreas. Uma maneira de reduzir a influência dessas duas áreas seria através da atribuição de pesos para as mesmas, por exemplo aplicando-se procedimentos robustos em que os pesos são obtidos de forma iterativa. Modelos alternativos também poderiam ser aplicados no sentido de reduzir a influência dessas observações, tais como modelos de quase-verossimilhança ou modelos com resposta beta, em que a resposta seria a porcentagem de domicílios com TV a cabo.

Tabela 5.8
Estimativas de máxima verossimilhança referentes do modelo log-linear binomial negativo ajustado aos dados sobre demanda de TV a cabo.

Efeito	Todos pontos	E/E.Padrão	Sem 1 e 14	E/E.Padrão
Intercepto	2,437	1,99	3,608	3,34
Domic	0,013	8,24	0,014	9,69
Percap	0,065	0,42	-0,002	-0,02
Taxa	0,041	1,84	0,010	0,50
Custo	-0,207	1,95	-0,266	-2,69
Ncabo	0,067	2,01	0,050	1,63
Ntv	-0,135	1,84	-0,071	-1,02
ϕ	3,311	3,49	5,060	2,89

5.4.7 Sobredispersão e quase-verossimilhança

De uma forma geral o fenômeno de sobredispersão sugere que a variância de Y seja dada por $\text{Var}(Y) = \sigma^2\mu$, em que $\sigma^2 > 1$. Uma maneira mais simples de resolver o problema é ajustar um modelo log-linear de Poisson aos dados e estimar σ^2 separadamente (método de quase-verossimilhança), por exemplo, usando a estimativa proposta por Wedderburn (1974), dada por

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} / (n - p), \tag{5.7}$$

em que $\hat{\mu}_i = \exp(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}})$. Algumas quantidades, tais como a matriz de variância-covariância assintótica de $\hat{\boldsymbol{\beta}}$, o desvio, resíduos etc, deverão ser corrigidos de maneira similar ao caso tratado na Seção 4.9. Finalmente, pode-se pensar na aplicação de modelos mais gerais de quase-verossimilhança que serão discutidos no Capítulo 6.

Aplicação

Como ilustração, considere os dados descritos na Tabela 5.9 (McCullagh e Nelder, 1989, Seção 6.3.2) e também no arquivo **navios.txt** em que avarias causadas por ondas em navios de carga são classificadas segundo o tipo do navio (A-E), ano da fabricação (1:1960-64, 2:1965-69, 3:1970-74 e 4:1975-79) e período de operação (1:1960-74 e 2:1975-79).

Foi também considerado o tempo em que cada navio ficou em operação (em meses). Inicialmente, um modelo log-linear de Poisson com *offset*, dado por log(meses), e efeitos principais é ajustado aos dados. Assim, denotando por Y_{ijk} o número de avarias observadas para o navio do tipo i, construído no ano j que operou no período k e supondo que $Y_{ijk} \overset{\text{ind}}{\sim} \text{P}(\lambda_{ijk} t_{ijk})$, em que t_{ijk} é o total de meses de operação e λ_{ijk} o número médio esperado de avarias por unidade de tempo. A parte sistemática do modelo é dada por

$$\log(\lambda_{ijk}) = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{3(k)},$$

com as restrições $\beta_{1(1)} = \beta_{2(1)} = \beta_{3(1)} = 0$, para $i = 1, \ldots, 5$; $j = 1, \ldots, 4$ e $k = 1, 2$, com β_1, β_2 e β_3 denotando, respectivamente, o efeito de tipo, de ano de construção e período de operação.

O desvio do modelo foi de $\text{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 38,69$ (25 graus de liberdade) que corresponde a um nível descritivo P$= 0,040$, indicando que o ajuste não está satisfatório. Pelo gráfico normal de probabilidades, descrito na Figura 5.10,

Tabela 5.9

Distribuição de avarias em navios de carga segundo o tipo do navio, ano de fabricação período de operação e total de meses em operação.

Tipo	Ano	Período	Meses	Avarias
A	1	1	127	0
A	1	2	63	0
A	2	1	1095	3
A	2	2	1095	4
A	3	1	1512	6
A	3	2	3353	18
A	4	2	2244	11
B	1	1	44882	39
B	1	2	17176	29
B	2	1	28609	58
B	2	2	20370	53
B	3	1	7064	12
B	3	2	13099	44
B	4	2	7117	18
C	1	1	1179	1
C	1	2	552	1
C	2	1	781	0
C	2	2	676	1
C	3	1	783	6
C	3	2	1948	2
C	4	2	274	1
D	1	1	251	0
D	1	2	105	0
D	2	1	288	0
D	2	2	192	0
D	3	1	349	2
D	3	2	1208	11
D	4	2	2051	4
E	1	1	45	0
E	2	1	789	7
E	2	2	437	7
E	3	1	1157	5
E	3	2	2161	12
E	4	2	542	1

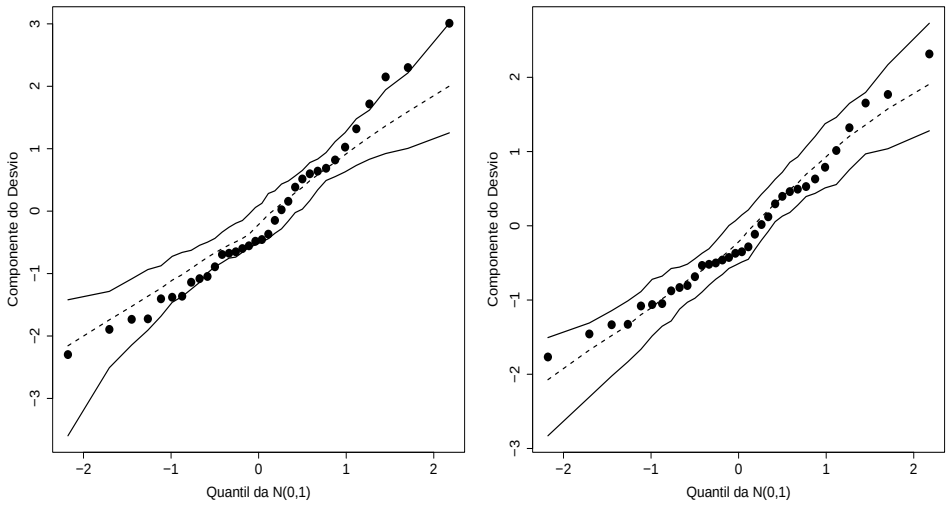

Figura 5.10: Gráficos normais de probabilidades referentes ao modelo log-linear de Poisson (esquerda) e log-linear de quase-verossimilhança (direita) ajustados aos dados sobre avarias em navios de carga.

nota-se a maioria dos resíduos próximos dos limites superior e inferior do envelope gerado, sugerindo sobredispersão que neste caso deve ser devido ao fato de um mesmo navio ter sido observado mais de uma vez. Usando (5.7) obtém-se $\hat{\sigma}^2 = 1,69$, e corrigindo o componente do desvio padronizado de modo que

$$t^*_{D_i} = \pm d_i/\hat{\sigma}\sqrt{1-\hat{h}_{ii}},$$

obtém-se um novo gráfico normal de probabilidades descrito na Figura 5.10, em que os resíduos estão melhor distribuídos dentro do envelope gerado.

O desvio corrigido fica dado por $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}}) = \mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}})/\hat{\sigma}^2$= 38,69/1,69 = 22,89 (25 graus de liberdade), indicando um ajuste adequado. Deve-se observar que tanto o resíduo $t^*_{D_i}$ como o desvio $\mathrm{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}})$ devem ser olhados de maneira meramente descritiva uma vez que em modelos de quase-verossimilhança a distribuição da resposta é em geral desconhecida. As esti-

mativas de máxima verossimilhança e os valores padronizados pelos respectivos erros padrão aproximados, já multiplicados pelo fator $\hat{\sigma}$, são apresentadas na Tabela 5.10. Williams (1987) mostra que o problema de sobredispersão neste exemplo é causado particularmente por duas observações discrepantes e sugere a inclusão da interação tipo*ano com pelo menos uma dessas observações excluídas. Pela Tabela 5.10 nota-se que os navios de tipos B e C são aqueles com uma incidência menor de avarias por unidade de tempo. Por outro lado, os navios fabricados de 65 a 74 como também aqueles que operaram de 75 a 79 apresentam uma inicidência maior de avarias por unidade de tempo do que os demais.

Tabela 5.10
Estimativas dos parâmetros referentes ao modelo log-linear de quase-verossimilhança ajustado aos dados sobre avarias em navios de carga.

Efeito		Estimativa	E/E.Padrão
Constante		-6,406	-22,69
Tipo			
	A	0,000	-
	B	-0,543	-2,36
	C	-0,687	-1,61
	D	-0,076	0,20
	E	0,326	1,06
Ano			
	60-64	0,000	-
	65-69	0,697	3,59
	70-74	0,818	3,71
	75-79	0,453	1,50
Período			
	60-74	0,000	-
	75-79	0,384	2,50

5.5 Relação entre a multinomial e a Poisson

Supor agora que todas as unidades amostrais são acompanhadas durante o mesmo período e que são classificadas segundo s níveis de exposição e r grupos, conforme descrito abaixo.

	Exposição				
Grupo	E1	E2	E3	$\cdots$	Es
G1	y_{11}	y_{12}	y_{13}	$\cdots$	y_{1s}
G2	y_{21}	y_{22}	y_{23}	$\cdots$	y_{2s}
		$\cdots$			
Gr	y_{r1}	y_{r2}	y_{r3}	$\cdots$	y_{rs}

Supondo que $Y_{ij} \overset{\text{ind}}{\sim} \text{P}(\mu_{ij})$, $i = 1, \ldots, r$ e $j = 1, \ldots, s$, tem-se que

$$Pr\{\mathbf{Y} = \mathbf{a} | \sum_{i,j} Y_{ij} = n\} = \frac{n!}{\Pi_{i,j} a_{ij}!} \Pi_{i,j} \pi_{ij}^{a_{ij}},$$

em que $\pi_{ij} = \mu_{ij}/\mu_{++}$, $\mu_{++} = \sum_{i,j} \mu_{ij}$, $\mathbf{Y} = (Y_{11}, \ldots, Y_{rs})^\top$ e $\mathbf{a} = (a_{11}, \ldots, a_{rs})^\top$. Considere o modelo log-linear de Poisson com parte sistemática dada por $\log(\mu_{ij}) = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)}$ e restrições $\beta_{1(1)} = \beta_{2(1)} = \beta_{12(1j)} = \beta_{12(i1)} = 0$, para $i = 1, \ldots, r$ e $j = 1, \ldots, s$. Segue que

$$\begin{aligned} \tau = \mu_{++} &= \sum_{i=1}^{r} \sum_{j=1}^{s} \exp\{\alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)}\} \\ &= e^{\alpha} \sum_{i=1}^{r} \sum_{j=1}^{s} \exp\{\beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)}\}, \end{aligned}$$

e pode-se definir as probabilidades

$$\pi_{ij} = \frac{\exp\{\beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)}\}}{\sum_{i=1}^{r} \sum_{j=1}^{s} \exp\{\beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)}\}},$$

em que o total do denominador é invariante com a parametrização utilizada no modelo. Tem-se que as probabilidades $\pi'_{ij}s$ não dependem do parâmetro

α. Como será mostrado a seguir, a estimativa de máxima verossimilhança do vetor $\boldsymbol{\beta}$ correspondente ao modelo multinomial coincide com a estimativa de máxima verossimilhança para $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\beta}_{12}^\top)^\top$ referente ao modelo log-linear de Poisson.

Por outro lado, se for ajustado um modelo multinomial do tipo log-linear aos dados tal que

$$\log(\pi_{ij}) = \alpha^* + \beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)},$$

tem-se, devido à imposição $\sum_{i,j} \pi_{ij} = 1$, que $\exp(\alpha^*) = 1/\sum_i \sum_j \exp\{\beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)}\}$, ou seja, $\alpha^* = \alpha - \log(\tau)$. O que muda é a estimativa do intercepto, embora na prática sempre seja possível obter α^* através de α e vice-versa. Para mostrar a equivalência das estimativas considere a relação abaixo

$$Pr\{\mathbf{Y} = \mathbf{a}|n\} = \frac{Pr\{\mathbf{Y} = \mathbf{a}; Y_{++} = n\}}{Pr\{Y_{++} = n\}},$$

em que $Y_{++} = \sum_{i,j} Y_{ij}$. Denotando $\mathrm{L}_{\mathbf{y}|n}(\boldsymbol{\beta}) = \log\{Pr(\mathbf{Y} = \mathbf{a}|n)\}$, $\mathrm{L}_{\mathbf{y}}(\tau, \boldsymbol{\beta}) = \log\{Pr(\mathbf{Y} = \mathbf{a}; Y_{++} = n)\}$ e $\mathrm{L}_{y_{++}}(\tau) = \log\{Pr(Y_{++} = n)\}$ tem-se que

$$\mathrm{L}_{\mathbf{y}}(\tau, \boldsymbol{\beta}) = \mathrm{L}_{y_{++}}(\tau) + \mathrm{L}_{\mathbf{y}|n}(\boldsymbol{\beta}), \tag{5.8}$$

em que

$$\mathrm{L}_{y_{++}}(\tau) = -\tau + y_{++}\log(\tau) - \log(y_{++}!)$$

e

$$\mathrm{L}_{\mathbf{y}|n}(\boldsymbol{\beta}) = \log(n!) + \sum_{i,j} a_{ij}\log(\pi_{ij}) - \sum_{i,j}\log(a_{ij}!).$$

Portanto, maximizar $\mathrm{L}_{\mathbf{y}}(\tau, \boldsymbol{\beta})$ com relação a $\boldsymbol{\beta}$ é equivalente a maximizar $\mathrm{L}_{\mathbf{y}|n}(\boldsymbol{\beta})$ com relação a $\boldsymbol{\beta}$. Isso quer dizer que as estimativas de máxima verossimilhança para o vetor $\boldsymbol{\beta}$ são as mesmas sob o modelo log-linear multinomial com probabilidades $\pi_{11}, \ldots, \pi_{rs}$ e sob o modelo log-linear de Poisson

de médias $\mu_{11}, \ldots, \mu_{rs}$. As matrizes de segundas derivadas com relação a $\boldsymbol{\beta}$, para os dois modelos, são tais que

$$\frac{\partial^2 \mathrm{L}_{\mathbf{y}}(\tau, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \frac{\partial^2 \mathrm{L}_{\mathbf{y}|n}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}.$$

Devido à linearidade em (5.8) segue que a matriz de informação observada para $(\tau, \boldsymbol{\beta}^\top)^\top$ é bloco-diagonal com elementos dados por $-\partial^2 \mathrm{L}_{\mathbf{y}}(\tau, \boldsymbol{\beta})/\partial \tau^2$ e $-\partial^2 \mathrm{L}_{\mathbf{y}}(\tau, \boldsymbol{\beta})/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top$, respectivamente. Segue, portanto, que a matriz de informação de Fisher será também bloco-diagonal com os valores esperados das quantidades acima,

$$\mathbf{K}_{\tau\beta} = \begin{bmatrix} \mathrm{E}_{\mathbf{y}} \left\{ -\frac{\partial^2 \mathrm{L}_{\mathbf{y}}(\tau,\beta)}{\partial \tau^2} \right\} & \mathbf{0} \\ \mathbf{0} & \mathrm{E}_{\mathbf{y}} \left\{ -\frac{\partial^2 \mathrm{L}_{\mathbf{y}}(\tau,\beta)}{\partial \beta \partial \beta^\top} \right\} \end{bmatrix}.$$

A variância assintótica de $\hat{\boldsymbol{\beta}}$ fica então dada por

$$\mathrm{Var}_{\mathbf{y}}(\hat{\boldsymbol{\beta}}) = [\mathrm{E}_{\mathbf{y}}\{-\partial^2 \mathrm{L}_{\mathbf{y}}(\tau, \boldsymbol{\beta})/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top\}]^{-1}.$$

Palmgren (1981) mostra que $\mathbf{K}_{\tau\beta}$ coincide com a matriz de informação observada sob a restrição $\tau = n$.

Esses resultados podem ser generalizados para quaisquer dimensões de tabelas bem como sob a presença de variáveis explicativas contínuas. A variância assintótica de $\hat{\boldsymbol{\beta}}$ fica no modelo multinomial dada por

$$\mathrm{Var}_{\mathbf{y}|n}(\hat{\boldsymbol{\beta}}) = \left[\mathrm{E}_{\mathbf{y}|n} \left\{ -\frac{\partial^2 \mathrm{L}_{\mathbf{y}|n}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right\} \right]^{-1},$$

coincidindo com a variância assintótica do modelo não condicional sob a restrição $\tau = n$. Contudo, do ponto de vista prático, as variâncias assintóticas de $\hat{\boldsymbol{\beta}}$ devem coincidir uma vez que a estimativa de máxima verossimilhança de τ é dada por $\hat{\tau} = n$.

5.5.1 Modelos log-lineares hierárquicos

Um modelo log-linear é dito hierárquico se dado que uma interação está no modelo, todas as interações de ordem menor como também os efeitos principais correspondentes deverão estar também no modelo. A utilização de tais modelos tem a vantagem de permitir uma interpretação das interações nulas como probabilidades condicionais. Em muitos casos estimativas dos valores médios podem ser expressas em forma fechada, evitando assim a utilização de processos iterativos.

Como ilustração, supor o modelo log-linear apresentado na seção anterior. Pode-se mostrar que a hipótese $\mathrm{H}_0 : \beta_{12(ij)} = 0$, $\forall ij$, é equivalente à hipótese de independência na tabela, isto é $\mathrm{H}_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$, $\forall ij$. Dado que não há interação, testar a ausência de efeito de exposição, isto é testar $\mathrm{H}_0 : \beta_{1(i)} = 0$, $i = 1, \ldots, r$, é equivalente a testar $\mathrm{H}_0 : \pi_{1+} = \cdots = \pi_{r+} = 1/r$. Finalmente, dado que não há interação, testar a ausência de efeito de grupo, isto é testar $\mathrm{H}_0 : \beta_{2(j)} = 0$, $j = 1, \ldots, s$, é equivalente a testar $\mathrm{H}_0 : \pi_{+1} = \cdots = \pi_{+s} = 1/s$.

Supor agora um modelo log-linear de Poisson com três fatores de r, s e t níveis, respectivamente. Pode-se representar a parte sistemática do modelo saturado da seguinte forma:

$$\log(\mu_{ijk}) = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{3(k)} + \beta_{12(ij)} + \beta_{13(ik)} + \beta_{23(jk)} + \beta_{123(ijk)}, \quad (5.9)$$

com as restrições $\beta_{1(1)} = \beta_{2(1)} = \beta_{3(1)} = 0$, $\beta_{12(1j)} = \beta_{12(i1)} = 0$, $\beta_{13(1k)} = \beta_{13(i1)} = 0$, $\beta_{23(1k)} = \beta_{23(j1)} = 0$, $\beta_{123(1jk)} = \beta_{123(i1k)} = \beta_{123(ij1)} = 0$, para $i = 1, \ldots, r$; $j = 1, \ldots, s$ e $k = 1, \ldots, t$. Há várias classes de modelos hierárquicos que correspondem a situações de interesse na tabela de contingência formada. Uma primeira classe corresponde à hipótese de ausência de interação de segunda ordem, representada por $\mathrm{H}_0 : \beta_{123(ijk)} = 0, \forall ijk$, sendo equivalente à hipótese de associação entre dois fatores quaisquer ser

constante nos níveis do terceiro fator. Isso quer dizer, em outras palavras, que a razão de produtos cruzados $\pi_{ijk}\pi_{i'j'k}/\pi_{ij'k}\pi_{i'jk}$, representando a associação entre os níveis (i,j) e (i',j') dos dois primeiros fatores, é constante nos níveis do terceiro fator. Se for omitido no modelo (5.9) a interação de segunda ordem mais uma interação de primeira ordem, os dois fatores omitidos correspondentes à interação de primeira ordem são independentes do terceiro fator. Por exemplo, se for omitido $\beta_{123(ijk)}$ e $\beta_{23(jk)}$, $\forall ijk$, ficando o modelo com a parte sistemática

$$\log(\mu_{ijk}) = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{3(k)} + \beta_{12(ij)} + \beta_{13(ik)},$$

os fatores 2 e 3 são independentes nos níveis do primeiro fator, ou equivalentemente, tem-se que

$$\pi_{ijk} = \pi_{ij+}\pi_{i+k}/\pi_{i++}, \quad \forall ijk.$$

Se agora for omitido além de $\beta_{123(ijk)}$ e $\beta_{23(jk)}$ também $\beta_{13(ik)}$, $\forall ijk$, ficando a parte sistemática dada por

$$\log(\mu_{ijk}) = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{3(k)} + \beta_{12(ij)},$$

o terceiro fator é independente dos dois primeiros, ou equivalentemente, tem-se que

$$\pi_{ijk} = \pi_{ij+}\pi_{++k}, \quad \forall ijk.$$

O modelo apenas com os efeitos principais, cuja parte sistemática é dada por

$$\log(\mu_{ijk}) = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{3(k)},$$

equivale à hipótese de independência entre os três fatores, isto é, tem-se que

$$\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}, \quad \forall ijk.$$

A Tabela 5.11 resume as três situações de independência para o modelo (5.9).

Tabela 5.11
Algumas interações em modelos log-lineares de Poisson.

Forma para π_{ijk}	Interação	Interpretação
$\pi_{i++}\pi_{+j+}\pi_{++k}$	nenhuma	fatores mutuamente independentes
$\pi_{ij+}\pi_{++k}$	$\beta_{12(ij)}$	fatores 1 e 2 independentes do fator 3
$\pi_{ij+}\pi_{i+k}/\pi_{i++}$	$\beta_{12(ij)} + \beta_{13(ik)}$	fatores 2 e 3 independentes nos níveis do fator 1

Em muitos desses casos é possível expressar as estimativas das probabilidades π_{ijk}'s em forma fechada. Uma análise mais completa de modelos hierárquicos pode ser encontrada, por exemplo, Agresti (2013).

5.5.2 Aplicações

Associação entre renda e satisfação no emprego

A Tabela 5.12 apresenta o resultado de uma pesquisa com 901 indivíduos (Agresti, 1990, pgs. 20-21) classificados segundo a renda anual e o grau de satisfação no emprego. Denote por Y_{ij} o número de indivíduos pertencentes à classe de renda i com grau de satisfação j. Esses dados estão disponíveis no arquivo **emprego.txt**.

Tabela 5.12
Classificação de indivíduos segundo a renda e o grau de satisfação no emprego.

	Grau de Satisfação			
Renda (US$)	Alto	Bom	Médio	Baixo
<6000	20	24	80	82
6000-15000	22	38	104	125
15000-25000	13	28	81	113
>25000	7	18	54	92

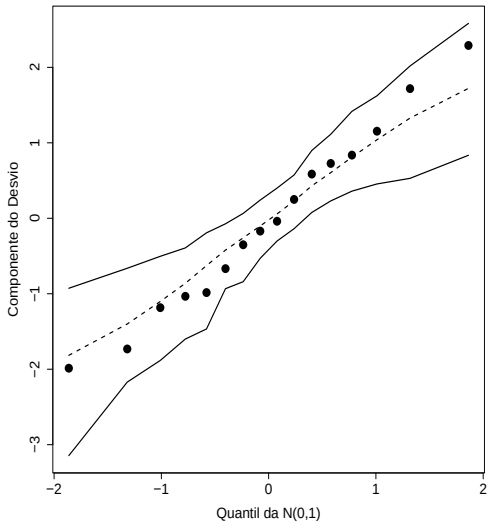

Figura 5.11: Gráfico normal de probabilidades referente ao modelo log-linear de Poisson ajustado aos dados sobre renda e satisfação no emprego.

Tabela 5.13
Estimativas dos parâmetros do modelo log-linear de Poisson ajustado ao dados sobre renda e satisfação no emprego.

Efeito	Parâmetro	Estimativa	E/E.Padrão
Constante	α	2,651	18,80
Renda 2	$\beta_{1(2)}$	0,338	3,71
Renda 3	$\beta_{1(3)}$	0,132	1,389
Renda 4	$\beta_{1(4)}$	-0,186	-1,81
Grau 2	$\beta_{2(2)}$	0,555	3,49
Grau 3	$\beta_{2(3)}$	1,638	11,87
Grau 4	$\beta_{2(4)}$	1,894	13,93

Supor que $Y_{ij} \sim \mathrm{P}(\mu_{ij})$ com parte sistemática inicialmente dada por (modelo saturado)

$$\log(\mu_{ij}) = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{12(ij)},$$

em que μ_{ij} denota o número esperado de indivíduos pertencentes à classe de renda i com grau de satisfação j, $\beta_{1(i)}$ denota o efeito renda, $\beta_{2(j)}$ denota o efeito satisfação e $\beta_{12(ij)}$ denota a interação. Tem-se as restrições $\beta_{1(1)} = \beta_{2(1)} = 0$. O teste da razão de verossimilhanças para testar H_0 : $\beta_{12(ij)} = 0$, $\forall ij$ (ausência de interação) fornece o valor $\xi_{RV} = 12,04$ com nível descritivo P= $0,21$, indicando pela ausência de interação ou independência entre os dois fatores. Denotando por π_{ij} a proporção de indivíduos na classe de renda i e grau de satisfação j, não rejeitar H_0 é equivalente a escrever $\pi_{ij} = \pi_{i+}\pi_{+j}$, $\forall ij$, em que π_{i+} denota a proporção de indivíduos na classe de renda i e π_{+j} denota a proporção de indivíduos com grau de satisfação j. Ou seja, tem-se independência entre renda e satisfação no emprego. Isso significa que a distribuição do grau de satisfação no emprego é mesma em todos as faixas de renda.

A Tabela 5.13 apresenta as estimativas dos parâmetros do modelo com efeitos principais. Os fatores renda e grau de satisfação são altamente significativos. Nota-se pelas estimativas dos parâmetros que há uma proporção maior de indivíduos na classe de renda 2 (6000-15000) e uma proporção menor na classe de renda 4 (>25000). Por outro lado, nota-se que a proporção de indivíduos cresce com o aumento do grau de satisfação. O desvio do modelo foi de $\mathrm{D}(\mathbf{y};\hat{\boldsymbol{\mu}}) = 12,04$ (9 graus de liberdade) com nível descritivo de P= $0,21$, indicando um ajuste adequado.

Pelo gráfico normal de probabilidades com o resíduo componente do desvio t_{D_i}, descrito na Figura 5.11, não há indícios fortes de que o modelo adotado seja incorreto, embora o fato dos resíduos negativos estarem abaixo da reta mediana e os resíduos positivos ligeiramente acima seja uma indício de sobredispersão nos dados. Assim, um modelo log-linear com resposta binimoail negativa poderia levar a um ajuste mais adequado;

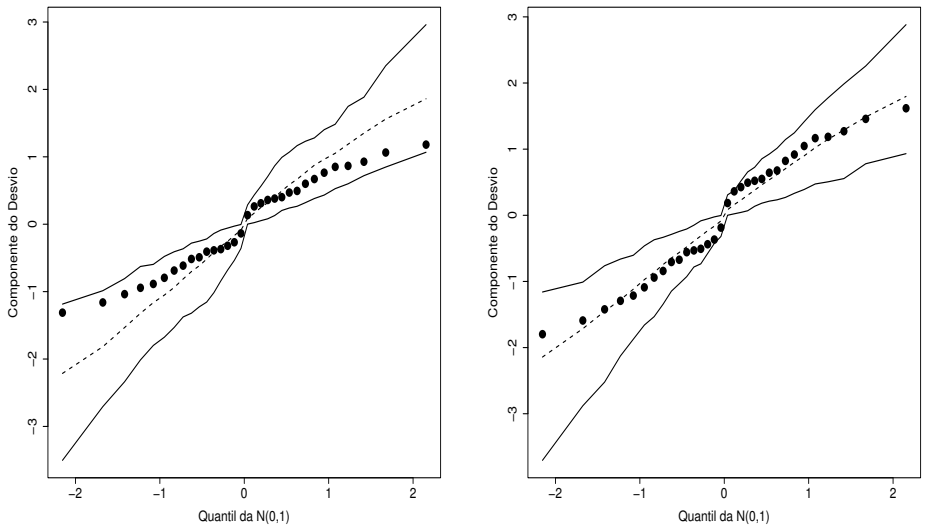

Figura 5.12: Gráficos normais de probabilidades referentes ao modelo log-linear de Poisson (esquerda) e log-linear de quase-verossimilhança (direita) ajustados aos dados sobre doença das coronárias.

Doença das coronárias

Considere agora os dados da Tabela 5.14 (Everitt, 1977) referente à classificação de 1330 pacientes segundo três fatores: doença das coronárias (sim ou não), nível de colesterol (1: menor do que 200 mg/100 cc, 2: 200-219, 3: 220-259 e 4: 260 ou +) e pressão arterial (1: menor do que 127 mm Hg, 2: 127-146, 3: 147-166 e 4: 167 ou +). Os dados estão também descritos no arquivo **heart.txt**. Denote por Y_{ijk} o número de pacientes nos níveis (i, j, k) dos três fatores: doença das coronárias, nível de colesterol e pressão arterial, respectivamente. Supor que $Y_{ijk} \sim \mathrm{P}(\mu_{ijk})$ com parte sistemática inicialmente dada por (modelo saturado)

$$\log(\mu_{ijk}) = \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{3(k)} + \beta_{12(ij)} + \beta_{13(ik)} + \beta_{23(jk)} + \beta_{123(ijk)},$$

Tabela 5.14
Distribuição de 1330 pacientes segundo ocorrência de doença das coronárias, nível de colesterol e pressão arterial.

Doença das coronárias	Nível de colesterol	Pressão arterial 1	2	3	4
Sim	1	2	3	3	4
	2	3	2	1	3
	3	8	11	6	6
	4	7	12	11	11
Não	1	117	121	47	22
	2	85	98	43	20
	3	119	209	68	43
	4	67	99	46	33

em que μ_{ijk} denota o número esperado de indivíduos pertencentes aos níveis (i, j, k), respectivamente, $\beta_{1(i)}$ denota o efeito doença das coronárias, $\beta_{2(j)}$ denota o efeito nível de colesterol, $\beta_{3(k)}$ denota o efeito pressão arterial e $\beta_{12(ij)}$, $\beta_{13(ik)}$, $\beta_{23(jk)}$ e $\beta_{123(ijk)}$ são as interações de $1^{\underline{a}}$ e $2^{\underline{a}}$ ordens, respectivamente, com as restrições dadas na Seção 5.4.1.

Tabela 5.15
Resumo do ANODEV referente ao modelo log-linear de Poisson ajustado aos dados sobre doença das coronárias.
(D:doença, C:colesterol e P:pressão)

Efeito	Desvio	g.l.	Diferença	g.l.
D+C+P	78,96	24	-	-
+ D.C	48,51	21	30,45	3
+ D.P	24,40	18	24,10	3
+ C.P	4,77	9	19,63	9

Pela Tabela 5.15 nota-se que, segundo o princípio hierárquico, apenas a interação de segunda ordem pode ser eliminada. A inclusão dos efeitos

principais é altamente significativa. Dado que os efeitos principais estão no modelo, a inclusão da interação doença*colesterol ($\beta_{12(ij)}$) leva a $\xi_{RV} = 30,45$ (3 graus de liberdade) com P= $0,00$. Dado que essa interação está no modelo, a inclusão da interação doença*pressão ($\beta_{13(ik)}$) fornece $\xi_{RV} = 24,10$ (3 graus de liberdade) com P= $0,00$. Finalmente, dadas as duas interações de primeira ordem, a inclusão da interação remanescente, colesterol*pressão, leva a $\xi_{RV} = 19,62$ (9 graus de liberdade) com P= $0,02$. O desvio do modelo (5.9) sem a interação de segunda ordem é de $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 4,77$ (9 graus de liberdade) para um nível descritivo de P= $0,853$, indicando um ajuste adequado.

A ausência de interação de segunda ordem neste exemplo significa que as razões de chances (entre os níveis de colesterol ou entre os níveis de pressão arterial) são as mesmas nos grupos de doentes e não doentes. Contudo, o gráfico normal de probabilidades descrito na Figura 5.12 (esquerda) indica que os resíduos negativos estão acima da média esperada, ocorrendo o contrário com os resíduos positivos, embora todos sejam em geral pequenos. Isso é um indício modesto de subdispersão, fenômeno que também pode ocorrer em modelos de Poisson. Um modelo de quase-verossimilhança similar ao que foi usado no exemplo da Seção 5.2.6 leva à estimativa $\hat{\sigma}^2 = 0,5326$. Na Figura 5.12 (direita) tem-se o gráfico normal de probabilidades com o resíduo componente do desvio corrigido pela estimativa de dispersão. Nota-se que os resíduos estão melhor distribuídos dentro do envelope gerado.

Gestantes fumantes

Na Tabela 5.15, extraída de Agresti (2013, Tabela 10.9), uma amostra de $n = 6851$ gestantes fumantes foi classificada segundo os seguintes fatores: idade (=0 < 30, =1 30+), número de cigarros consumidos por dia (=0 < 5, =1 5+), tempo de gestação (=0 ≤ 260 dias, =1 > 260 dias) e situação da

criança (=0 não sobreviveu, =1 sobreviveu). Esses dados estão descritos no arquivo **gestantesc.txt**.

Tabela 5.15
Distribuição de 6851 gestantes fumantes segundo a idade, o nível de exposição ao cigarro, o tempo da gestação e a sobrevivência da criança.

		Duração da	Sobrevivência	
Idade	No. cigarros	Gestação	Não	Sim
< 30	< 5	≤ 260	50	315
		> 260	24	4012
	5+	≤ 260	9	40
		> 260	6	459
30+	< 5	≤ 260	41	147
		> 260	14	1594
	5+	≤ 260	4	11
		> 260	1	124

Como todos os efeitos são binários, é possível construir 6 tabelas 2×2 e avaliar marginalmente a associação em cada tabela através de razões de chances. A seguir são resumidas as tabelas com as respectivas interpretações.

	No. cigarros		
Idade	< 5	5+	Total
< 30	4401	514	4915
30+	1796	140	1936
Total	6197	654	6851
$\hat{\psi}$	$\frac{514 \times 1796}{4401 \times 140}$	1,50	

	Gestação		
Idade	≤ 260	> 260	Total
< 30	414	4501	4915
30+	203	1733	1936
Total	617	6234	6851
$\hat{\psi}$	$\frac{4501 \times 203}{414 \times 1733}$	1,27	

Portanto, das duas tabelas acima, tem-se que gestantes que fumam 5+ cigarros por dia têm chance maior de ser mais jovem, enquanto gestantes

com gestação maior de 260 dias têm chance ligeiramente maior de ser mais jovem.

	Sobrevivência		
Idade	Não	Sim	Total
< 30	89	4826	4915
30+	60	1876	1936
Total	149	6702	6851
$\hat{\psi}$	$\frac{4826\times60}{89\times1876}$	1,73	

	Gestação		
No. Cigarros	≤ 260	> 260	Total
< 5	553	5644	6197
5+	64	590	654
Total	617	6234	6851
$\hat{\psi}$	$\frac{5644\times64}{553\times590}$	1,10	

Asim, das duas tabelas acima, gestantes com criança que sobreviveu têm chance maior de ser mais jovem, enquanto gestantes com gestação maior têm chance ligeiramente maior de fumar < 5 cigarros por dia.

	Sobrevivência		
No. Cigarros	Não	Sim	Total
< 5	129	6068	6197
5+	20	634	654
Total	149	6702	6851
$\hat{\psi}$	$\frac{6068\times20}{129\times634}$	1,48	

	Sobrevivência		
Gestação	Não	Sim	Total
≤ 260	104	513	617
> 260	45	6189	6234
Total	149	6702	6851
$\hat{\psi}$	$\frac{104\times6189}{513\times45}$	27,88	

Logo, pelas tabelas acima, gestantes com criança que sobreviveu têm chance maior de fumar < 5 cigarros, enquanto gestantes com criança que não sobreviveu têm chance muito maior de gestação menor.

Todos os resultados acima, além de marginais são descritivos e precisam ser validados inferencialmente. Nesse sentido, denote por Y_{ijkl} o número de gestantes fumantes no i-ésimo grupo de idade, j-ésimo grupo de consumo diário de cigarros, k-ésimo grupo de gestação e l-ésima condição da criança, para $i, j, k, l = 0, 1$. Supor $Y_{ijkl} \overset{\text{ind}}{\sim} \text{P}(\mu_{ijkl})$ com componente sistemático

envolvendo apenas interações até 1ª ordem

$$\begin{aligned}\log(\mu_{ijkl}) &= \alpha + \beta_{1(i)} + \beta_{2(j)} + \beta_{3(k)} + \beta_{4(l)} + \beta_{12(ij)} + \beta_{13(ik)} + \beta_{14(il)} + \\ &\quad \beta_{23(jk)} + \beta_{24(jl)} + \beta_{34(kl)},\end{aligned}$$

em que μ_{ijkl} denota o número médio de gestantes fumantes na condição (i, j, k, l). Na parametrização acima tem-se um modelo casela de referência.

Aplicando o método de Akaike apenas três interações permanecem no modelo. As estimativas do modelo ajustado estão descritas na Tabela 5.16.

Tabela 5.16
Estimativas dos parâmetros referentes ao modelo log-linear de Poisson ajustado aos dados sobre gestantes fumantes.

Efeito	Parâmetro	Estimativa	E/E.Padrão
Constante	α	4,019	33,77
Idade30+	$\beta_{1(1)}$	-0,359	-2,15
Cigarros5+	$\beta_{2(1)}$	-2,147	-46,07
Gestação>260	$\beta_{3(1)}$	-0,838	-4,70
SobrevivênciaSim	$\beta_{4(1)}$	1,783	14,03
Idade30+*Cigar5+	$\beta_{12(11)}$	-0,404	-4,07
Idade30+*SobrevSim	$\beta_{14(11)}$	-0,551	-3,25
Gest>260*SobrevSim	$\beta_{34(11)}$	3,328	18,06

Nota-se que apenas as interações idade*cigarros, idade*sobrevivência e gestação*sobrevivência são significativas. O desvio do modelo $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 7,72$ (8 graus de liberdade) com P=0,46 indica modelo bem ajustado. O gráfico normal de probabilidades com o resíduo componente do desvio descrito na Figura 5.12 confirma a qualidade do ajuste.

As interações podem ser interpretadas conjuntamente através de gráficos de perfis dos valores ajustados, ou através de tabelas de contingência tipo 2×2 envolvendo apenas as interações significativas. Nessas tabelas de contingência

ao invés dos valores observados tem-se os valores ajustados, sendo possível também encontrar as estimativas pontuais das razões de chances correspondentes com as respectivas estimativas intervalares. As tabelas ajustadas são descritas a seguir.

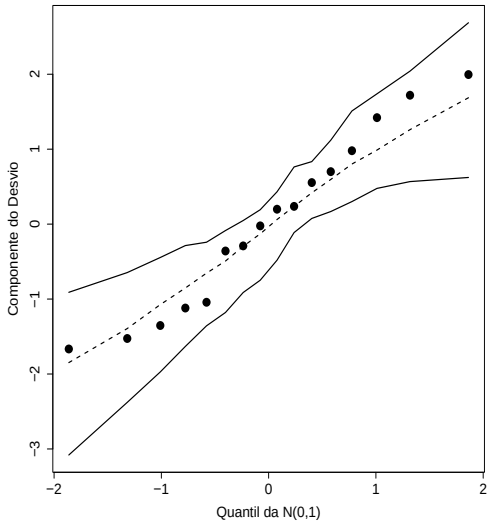

Figura 5.13: Gráfico normal de probabilidades referente ao modelo log-linear de Poisson ajustado aos dados sobre gestantes fumantes.

	No. cigarros	
Idade	< 5	$5+$
< 30	$\hat{\mu}_{00kl}$	$\hat{\mu}_{01kl}$
$30+$	$\hat{\mu}_{10kl}$	$\hat{\mu}_{11kl}$

Razão de chances estimada $\hat{\psi} = \frac{\hat{\mu}_{01kl} \times \hat{\mu}_{10kl}}{\hat{\mu}_{00kl} \times \hat{\mu}_{11kl}} = e^{-\hat{\beta}_{12(11)}} = e^{0,404} = 1,50$ com estimativa intervalar de 95% $\exp[0,404 \pm 1,96 * 0,099] = [1,23; 1,82]$. Portanto,

fixando os demais níveis, uma gestante que fuma 5+ cigarros em média por dia tem $1,50$ a chance de ter < 30 anos de uma gestante que fuma em média < 5 cigarros por dia.

	Sobrevivência	
Idade	Não	Sim
< 30	$\hat{\mu}_{0jk0}$	$\hat{\mu}_{0jk1}$
30+	$\hat{\mu}_{1jk0}$	$\hat{\mu}_{1jk1}$

Razão de chances estimada $\hat{\psi} = \frac{\hat{\mu}_{0jk1} \times \hat{\mu}_{1jk0}}{\hat{\mu}_{0jk0} \times \hat{\mu}_{1jk1}} = e^{-\hat{\beta}_{14(11)}} = e^{0,551} = 1,73$ com estimativa intervalar de 95% $\exp[0,551 \pm 1,96 * 0,169] = [1,25; 2,42]$. Logo, fixando os demais níveis, uma gestante que a criança sobreviveu tem $1,73$ a chance de ter < 30 anos de uma gestante que a criança não sobreviveu.

	Sobrevivência	
Gestação	Não	Sim
≤ 260	$\hat{\mu}_{ij00}$	$\hat{\mu}_{ij01}$
> 260	$\hat{\mu}_{ij10}$	$\hat{\mu}_{ij11}$

Razão de chances estimada $\hat{\psi} = \frac{\hat{\mu}_{ij00} \times \hat{\mu}_{ij11}}{\hat{\mu}_{ij01} \times \hat{\mu}_{ij10}} = e^{\hat{\beta}_{34(11)}} = e^{3,328} = 27,88$ com estimativa intervalar de 95% $\exp[3,328 \pm 1,96 * 0,184] = [19,44; 39,99]$. Assim, fixando os demais níveis, uma gestante que a criança nao sobreviveu tem $27,88$ a chance de ter uma gestação ≤ 260 dias de uma gestante que a criança sobreviveu.

A conclusão é que as estimativas ajustadas para as médias e consequentemente para as razões de chances são muito próximas, reforçando a qualidade preditiva do modelo. Contudo, as estimativas ajustadas são mais precisas uma vez que levam em conta as informações de todas as caselas da tabela.

5.6 Modelos com excesso de zeros

5.6.1 Modelos ajustados em zero

Os modelos de contagem ajustados em zero são também conhecidos como modelos de barreira (ver, por exemplo, Mullaby, 1986). Para formalizá-los vamos supor que Z é uma variável aleatória com função de probabilidades dada por

$$P\{Z=z\} = \begin{cases} \pi & \text{se} \quad z=0, \\ (1-\pi)\frac{f_Y(z)}{\{1-f_Y(0)\}} & \text{se} \quad z=1,2,\ldots, \end{cases}$$

em que $0<\pi<1$ e $f_Y(z)$ denota a função de probabilidades de uma variável aleatória Y de contagem, por exemplo, Poisson ou binomial negativa. Portanto, desde que $\sum_{z=1}^{\infty} f_Y(z) = 1 - f_Y(0)$, segue que

$$\begin{aligned} P\{Z \geq 1\} &= (1-\pi)\sum_{z=1}^{\infty} f_Y(z)/\{1-f_Y(0)\} \\ &= (1-\pi)\{1-f_Y(0)\}/\{1-f_Y(0)\} \\ &= 1-\pi. \end{aligned}$$

Logo, $\sum_{y=0}^{\infty} P\{Z=z\} = \pi + (1-\pi) = 1$. Um exemplo poderia ser Z denotando o número de dias que pacientes dependendes de álcool que estão fazendo tratamento consumiram a bebida. O zero representa os pacientes que ficaram em abstinência no período mas que poderiam ter consumido alcool. Um outro exemplo poderia ser estudar o número de vezes que um idoso visita um médico no período de 1 ano. Os zeros são aqueles idosos que naquele ano não precisaram ir ao médico. Os fatores que explicam a probabilidade de zero podem ser diferentes daqueles que explicam a probabilidade de ocorrência do evento.

Os dois primeiros momentos de Z ficam dados por

$$\begin{aligned}
\mathrm{E}(Z) &= \sum_{z=1}^{\infty} z(1-\pi)\frac{f_Y(z)}{\{1-f_Y(0)\}} \\
&= \frac{(1-\pi)}{\{1-f_Y(0)\}}\sum_{z=1}^{\infty} z f_Y(z) \\
&= \frac{\mathrm{E}(Y)(1-\pi)}{\{1-f_Y(0)\}}
\end{aligned}$$

e

$$\begin{aligned}
\mathrm{E}(Z^2) &= \sum_{z=1}^{\infty} z^2(1-\pi)\frac{f_Y(z)}{\{1-f_Y(0)\}} \\
&= \frac{(1-\pi)}{\{1-f_Y(0)\}}\sum_{z=1}^{\infty} z^2 f_Y(z) \\
&= \frac{\mathrm{E}(Y^2)(1-\pi)}{\{1-f_Y(0)\}}.
\end{aligned}$$

Daí segue que

$$\mathrm{Var}(Z) = \frac{(1-\pi)}{\{1-f_Y(0)\}}\left[\mathrm{E}(Y^2) - \frac{\mathrm{E}^2(Y)(1-\pi)}{\{1-f_Y(0)\}}\right].$$

Iremos denotar $Z \sim \mathrm{ZAP}(\lambda, \pi)$ para o modelo de Poisson ajustado em zero e $Z \sim \mathrm{ZANB}(\lambda, \phi, \pi)$ para o modelo binomial negativo ajustado em zero. Logo, se $Y \sim \mathrm{P}(\lambda)$ então $f_Y(y) = e^{-\lambda}\lambda^y/y!$ e em particular $f_Y(0) = e^{-\lambda}$. Para $Y \sim \mathrm{BN}(\lambda, \phi)$ temos que

$$f_Y(y) = \frac{\Gamma(\phi+y)}{\Gamma(y+1)\Gamma(\phi)}\left(\frac{\lambda}{\lambda+\phi}\right)^y\left(\frac{\phi}{\lambda+\phi}\right)^{\phi},$$

em particular $f_Y(0) = \phi^\phi/(\lambda+\phi)^\phi$.

5.6.2 Modelos de regressão ajustados em zero

Vamos supor agora que $Z_1, \ldots, Z_n$ são variáveis aleatórias independentes com distribuição de Poisson ou binomial negativa ajustadas em zero. Então,

$$P\{Z_i = z_i\} = \begin{cases} \pi_i & \text{se} \quad z_i = 0, \\ (1-\pi_i)\frac{f_{Y_i}(z_i)}{\{1-f_{Y_i}(0)\}} & \text{se} \quad z_i = 1, 2, \ldots, \end{cases}$$

para $i = 1, \ldots, n$. O logaritmo da função de verossimilhança fica dado por L$= \sum_{i=1}^{n} \log f_{Z_i}(z_i)$, em que $\log f_{Z_i}(0) = \log \pi_i$ e $\log f_{Z_i}(z_i) = \log(1-\pi_i) + \log f_{Y_i}(z_i) - \log\{1 - f_{Y_i}(0)\}$ para $z_i = 1, 2, \ldots$.

Por exemplo, se assumimos que $Z_i \sim \text{ZAP}(\lambda_i, \pi)$ em que $\lambda_i = e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$ e $\mu_i = \text{E}(Z_i)$, então segue que

$$\begin{aligned} \mu_i &= \frac{\text{E}(Y_i)(1-\pi)}{\{1 - f_Y(0)\}} \\ &= \frac{\lambda_i(1-\pi)}{\{1 - e^{-\lambda_i}\}} \\ &= \frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}(1-\pi)}{[1 - \exp\{-\exp(\mathbf{x}_i^\top \boldsymbol{\beta})\}]}. \end{aligned}$$

Portanto,

$$\log \mu_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \log(1-\pi) - \log[1 - \exp\{-\exp(\mathbf{x}_i^\top \boldsymbol{\beta})\}].$$

5.6.3 Modelos inflacionados de zeros

Os modelos de contagem inflacionados de zeros (ver, por exemplo, Lambert, 1992) são caracterizados pela ocorrência de zeros em duas situações: (i) zeros que ocorrem segundo uma distribuição de contagem ou (ii) zeros inflados que podem ser zeros estruturais. Por exemplo, podemos estar interessados em estudar o número de dias num período que um indivíduo consumiu um determinado produto. Aqueles indivíduos que não consomem o produto por

alguma razão, por exemplo intolerância ao produto, serão tratadas como zeros estruturais e aqueles indivíduos que não consumiram o produto no período, mas podem vir a consumí-lo, como zeros eventuais que serão associados a alguma distribuição de contagem Y. Um outro exemplo poderia ser o estudo do número de vezes que um indíviduo contraiu um tipo de infecção num determinado período. Aqueles indivíduos imunes à infecção seriam tratados como zeros inflados. Nesses casos, a variável aleatória Z tem função de probabilidades expressa na seguinte forma:

$$P\{Z=z\} = \begin{cases} \pi + (1-\pi)f_Y(0) & \text{se} \quad z=0, \\ (1-\pi)f_Y(z) & \text{se} \quad z=1,2,\ldots, \end{cases}$$

em que $0 < \pi < 1$ e $f_Y(z)$ denota a função de probabilidades de uma variável aleatória Y, por exemplo, Poisson ou binomial negativa. Desde que $\sum_{z=1}^{\infty} f_Y(z) = 1 - f_Y(0)$ obtemos $\sum_{z=0}^{\infty} P\{Z=z\} = \pi + (1-\pi)f_Y(0)\} + (1-\pi)\{1-f_Y(0)\} = \pi + (1-\pi) = 1$.

Os dois primeiros momentos de Y ficam dados por

$$\begin{aligned} \mathrm{E}(Z) &= \sum_{z=1}^{\infty} z(1-\pi)f_Y(z) \\ &= (1-\pi)\sum_{z=1}^{\infty} z f_Y(z) \\ &= (1-\pi)\mathrm{E}(Y) \end{aligned}$$

e

$$\begin{aligned} \mathrm{E}(Z^2) &= \sum_{z=1}^{\infty} z^2(1-\pi)f_Y(z) \\ &= (1-\pi)\sum_{z=1}^{\infty} z^2 f_Y(z) \\ &= (1-\pi)\mathrm{E}(Y^2). \end{aligned}$$

Assim,

$$\begin{aligned}\mathrm{Var}(Z) &= \mathrm{E}(Z^2) - \mathrm{E}^2(Z) \\ &= (1-\pi)\mathrm{E}(Y^2) - (1-\pi)^2\mathrm{E}^2(Y) \\ &= (1-\pi)\{\mathrm{E}(Y^2) - (1-\pi)\mathrm{E}^2(Y)\}.\end{aligned}$$

Iremos denotar $Z \sim \mathrm{ZIP}(\lambda, \pi)$ para a distribuição de Poisson inflacionada de zeros e por $Z \sim \mathrm{ZINB}(\lambda, \phi, \pi)$ para a distribuição binomial negativa inflacionada de zeros.

5.6.4 Modelos de regressão inflacionados de zeros

Vamos supor agora que $Z_1, \ldots, Z_n$ são variáveis aleatórias independentes com distribuição de Poisson ou binomial negativa inflacionadas de zeros. Então,

$$P\{Z_i = z_i\} = \begin{cases} \pi_i + (1-\pi_i)f_Y(0) & \text{se} \quad z_i = 0, \\ (1-\pi_i)f_Y(z_i) & \text{se} \quad z_i = 1, 2, \ldots, \end{cases}$$

para $i = 1, \ldots, n$. O logaritmo da função de verossimilhança fica dado por $\mathrm{L} = \sum_{i=1}^{n} \log f_{Z_i}(z_i)$, em que $\log f_{Z_i}(0) = \log\{\pi_i + (1-\pi_i)f_{Y_i}(0)\}$ e $\log f_{Z_i}(z_i) = \log(1-\pi_i) + \log f_{Y_i}(z_i)$ para $z_i = 1, 2, \ldots$.

Por exemplo, podemos supor que $Y_i \sim \mathrm{P}(\lambda_i)$ com $\lambda_i = e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$ e $\log\{\pi_i/(1-\pi_i)\} = \mathbf{u}_i^\top \boldsymbol{\gamma}$. Dessa forma segue que $\mu_i = \mathrm{E}(Z_i)$ fica expresso como

$$\begin{aligned}\mu_i &= (1-\pi_i)\mathrm{E}(Y_i) \\ &= (1-\pi_i)\lambda_i \\ &= \left\{1 - \frac{e^{\mathbf{u}_i^\top \boldsymbol{\gamma}}}{1 + e^{\mathbf{u}_i^\top \boldsymbol{\gamma}}}\right\} e^{\mathbf{x}_i^\top \boldsymbol{\beta}} \\ &= \frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{\{1 + e^{\mathbf{u}_i^\top \boldsymbol{\gamma}}\}}.\end{aligned}$$

Isto é,

$$\log \mu_i = \mathbf{x}_i^\top \boldsymbol{\beta} - \log\{1 + e^{\mathbf{u}_i^\top \boldsymbol{\gamma}}\}.$$

Os modelos de contagem ajustados em zero e inflacionados de zeros podem ajustados, por exemplo, pela bibilioteca `pscl` (Zeileis et al. 2008) do `R`.

5.7 Exercícios

1. Seja Y uma variável aleatória com distribuição binomial negativa, isto é, Y é o número de ensaios até a ocorrência do r-ésimo sucesso, em que π é a probabilidade de sucesso em cada ensaio. Mostre que a função de probabilidades de Y pode ser expressa na forma exponencial. Calcule μ e $V(\mu)$. Use a forma abaixo para a função de probabilidades de Y

$$f(y;\pi,r) = \binom{y-1}{r-1}\pi^r(1-\pi)^{(y-r)},$$

em que $y = r, r+1, \ldots$. Obtenha a função desvio supondo uma amostra de n variáveis aleatórias independentes de probabilidades de sucesso π_i.

2. Seja Y variável aleatória com distribuição binomial negativa biparamétrica de média μ e parâmetro de forma ν, cuja função de probabilidades é dada por

$$f(y;\mu,\nu) = \frac{\Gamma(\nu+y)}{\Gamma(y+1)\Gamma(\nu)}\left(\frac{\mu}{\mu+\nu}\right)^y\left(\frac{\nu}{\mu+\nu}\right)^\nu,$$

em que $\mu > 0$, $\nu > 0$ e $y = 0, 1, 2, \ldots$. Mostre que para ν conhecido a distribuição de Y pertence à família exponencial de distribuições. Encontrar a função de variância. Obtenha a função desvio supondo uma amostra de n variáveis aleatórias independentes de médias μ_i e parâmetro de dispersão ν.

3. Sejam Y_1 e Y_2 variáveis aleatórias independentes tais que $Y_i \sim \mathrm{P}(\lambda_i)$, $i = 1, 2$. Considere a razão de taxas $\psi = \lambda_1/\lambda_2$. Encontre a variância assintótica de $\tilde{\psi}$, $\mathrm{Var}_A(\tilde{\psi})$.

4. A tabela abaixo (Breslow e Day, 1987) apresenta o número de mortes por câncer respiratório e o número de pessoas-anos de observação entre trabalhadores de indústrias siderúrgicas do estado de Montana (EUA) segundo o nível de exposição ao arsênico.

	Nível de Exposição	
	Alto	Baixo
Casos	68	47
Pessoas-Anos	9018	13783

Sejam Y_1 e Y_2 o número de casos observados para o nível alto e baixo de arsênico, respectivamente. Suponha que $Y_i \sim \mathrm{P}(\lambda_i t_i)$, em que t_i denota o número de pessoas-anos, $i = 1, 2$. Considere a razão de taxas $\psi = \lambda_1/\lambda_2$. Encontre $\tilde{\psi}$ e um intervalo de confiança exato de 95% para ψ. Com base neste intervalo qual sua conclusão sobre a hipótese $\mathrm{H}_0 : \psi = 1$? Informações úteis: $F_{136,96}(0,025) = 0,694$ e $F_{138,94}(0,975) = 1,461$.

5. Os dados do arquivo **nasal.txt** (Breslow e Day, 1987, pgs. 140-142) são provenientes de um estudo de seguimento para estudar a associação entre a taxa anual de câncer nasal em trabalhadores de uma refinaria de níquel no País de Gales e algumas variáveis explicativas: idade no primeiro emprego (4 níveis), ano do primeiro emprego (4 níveis) e tempo decorrido desde o primeiro emprego (5 níveis). São também apresentados o número de casos de câncer nasal e o total de pessoas-anos para cada combinação desses três fatores.

   Para ler o arquivo no R use os comandos

   ```
   nasal = read.table("nasal.txt", header=TRUE)
   ```

   Informar que as variáveis explicativas idade, ano e tempo são fatores

   ```
   nasal$idade = factor(nasal$idade)
   ```

```
levels(nasal$idade) = c("<20", "20-27,4", "27,5-34,9",
"35-54,4")

nasal$tempo = factor(nasal$tempo)

levels(nasal$tempo) = c("0-19,9", "20-29,9", "30-39,9",
"40-49,9", "50+")

nasal$ano = factor(nasal$ano)

levels(nasal$ano) = c("1902-1910", "1910-1914", "1915-1919",
"1920-1924")

summary(nasal)

attach(nasal)
```

Fazer inicialmente uma análise descritiva dos dados. Como o número de casos depende do tempo de observação é preciso criar uma taxa de câncer nasal por unidade de tempo e a partir daí gerar boxplots robustos e comentar.

```
require(robustbase)

tnasal = (casos/panos)*100

adjbox(split(tnasal,idade), ylab="Taxa (por 100 panos)",
xlab="Idade", col="gray",cex=2, cex.axis=1.5, cex.lab=1.5,
names=c("<20", "20-27,4", "27,5-34,9", "35-54,4")

adjbox(split(tnasal,tempo), ylab="Taxa (por 100 panos)",
xlab="Tempo", col="gray", cex=2, cex.axis=1.5, cex.lab=1.5,
names=c("0-19,9", "20-29,9", "30-39,9", "40-49,9", "50+")

adjbox(split(tnasal,ano), ylab="Taxa (por 100 panos)", xlab="Ano",
col="gray", cex=2, cex.axis=1.5, cex.lab=1.5, names=c("1902-1910",
"1910-1914", "1915-1919", "1920-1924")
```

Ajustar um modelo log-linear de Poisson no GAMLSS com pessoas-anos como offset

```
require(gamlss)
```

```
fit1.nasal = gamlss(casos ∼ idade + ano + tempo +
offset(log(panos)), family=PO)
```

```
summary(fit1.nasal)
```

Verificar através de análise de resíduos se o modelo está bem ajustado.

```
plot(fit1.nasal)
```

```
rqres.plot(fit1.nasal, howmany=50, plot="all")
```

Se o modelo estiver bem ajustado verifique se é possível fazer agrupamentos dos níveis de alguns dos fatores. Segue abaixo exemplo de possíveis agrupamentos dos fatores tempo e ano.

```
ntempo = tempo
```

```
levels(ntempo) = c("0-39,9", "0-39,9","0-39,9", "40-49,9",
"50+")
```

```
nano = ano
```

```
levels(nano) = c("1902-1919", "1902-1919", "1902-1919",
"1920-1924")
```

Ajustar novamente o modelo no GAMLSS com os fatores agrupados

```
fit2.nasal = gamlss(casos ∼ idade + nano + ntempo +
offset(log(panos)), family=PO)
```

```
summary(fit2.nasal)
```

Verificar através de análise de resíduos se o modelo está bem ajustado.

```
plot(fit2.nasal)
```

```
rqres.plot(fit2.nasal, howmany=50, plot="all")
```

Interpretar os resultados através do `term.plot` e das estimativas pontuais e intervalares de 95% para as razões de taxas entre os níveis de cada novo fator.

```
term.plot(fit2.nasal, terms=1)

term.plot(fit2.nasal, terms=2)

term.plot(fit2.nasal, terms=3).
```

6. No arquivo **geriatra.txt** (Neter et al., 1996, p. 623) estão descritos os dados de um estudo prospectivo com 100 indivíduos de pelo menos 65 anos de idade em boas condições físicas. O objetivo do estudo é tentar relacionar o número médio de quedas num período de seis meses com algumas variáveis explicativas. Os dados estão descritos na seguinte ordem: `quedas` (número de quedas no período), `interv` (intervenção, =0 educação somente, =1 educação + exercícios físicos), `sexo` (=0 feminino, =1 masculino), `balanco` (escore do balanço) e `forca` (escore da força). Para as duas últimas variáveis quanto maior o valor maior o balanço e a força do indivíduo, respectivamente.

   Para ler o arquivo no R use os comandos

   ```
   geriatra = read.table("geriatra.txt"', header=TRUE)

   attach(geriatra)

   summary(geriatra)
   ```

   Informar o sistema as variáveis que são fatores

   ```
   geriatra$interv = factor(geriatra$interv)

   levels(geriatra$interv) = c("EDUC","EDUC+AF")
   ```

```
geriatra$sexo = factor(geriatra$sexo)

levels(geriatra$sexo) = c("FEM","MASC")
```

Fazer inicialmente uma análise descritiva com boxplots robustos de cada variável quantitativa e de cada variável quantitativa com sexo e intervenção e também diagramas de dispersão entre o número de quedas e cada uma das variáveis explicativas contínuas. Por exemplo

```
require(robustbase)

adjbox(quedas,ylab="Número de Quedas", cex=2, cex.axis=1.5,
cex.lab=1.5, col="gray")

adjbox(split(quedas,sexo),ylab="Número de Quedas", xlab="Sexo",
names=c("Feminino","Masculino"),

cex=2, cex.axis=1.5, cex.lab=1.5, col="gray")

adjbox(split(quedas,interv), ylab="Número de Quedas",
xlab="Intervenção", col="gray",

names=c("Educação","Educação + Ativ.Físicas"), cex=2,
cex.axis=1.5, cex.lab=1.5)
```

Ajustar inicialmente um modelo log-linear de Poisson no R apenas com os efeitos principais

```
fit1.geriatra = glm(quedas ~ sexo + interv + força + balanço,
family=poisson)

summary(fit1.geriatra)
```

Fazer uma seleção dos efeitos principais através do comando `stepAIC`

```
require(MASS)

fit2.geriatra = stepAIC(fit1.geriatra)
```

```
summary(fit2.geriatra)
```

Com o modelo final fazer uma análise de diagnóstico

```
fit.model = fit2.geriatra

source("envel_pois.txt")

source("diag_cook_pois.txt")
```

Para as observações mais influentes avaliar o impacto no ajuste. Interpretar os coeficientes do modelo final.

Avaliar a inferência assintótica do modelo log-linear de Poisson através do *bootstrap*

```
require(car)

set.seed(12345)

geriatra.boot = Boot(fit2.geriatra, R=1000)

summary(geriatra.boot)

Confint(geriatra.boot, level=.95, type="bca")

hist(geriatra.boot, legend="none").
```

Comparar os intervalos BCa *bootstrap* com os intervalos assintóticos do estimador de máxima verossimilhança.

7. No arquivo **rolos.txt** (Hinde, 1982) são apresentados os dados referentes à produção de peças de tecido numa determinada fábrica. Na primeira coluna tem-se o comprimento da peça (em metros) e na segunda coluna o número de falhas. Faça inicialmente um gráfico do número de falhas contra o comprimento da peça. Ajuste um modelo log-linear de Poisson apropriado. Faça uma análise de resíduos e verifique se há indícios de sobredispersão. Em caso afirmativo ajuste

um modelo de quase-verossimilhança e um modelo log-linear com distribuição binomial negativa. Interprete os resultados pelas razões de médias $\mu(x+1)/\mu(x)$, em que x denota o comprimento da peça.

8. Sejam $Y_1, \ldots, Y_n$ variáveis aleatórias independentes tais que $Y_i \sim \mathrm{P}(\mu_i)$ e parte sistemática dada por $\eta_i = \sum_{j=1}^{p} x_{ij}\beta_j$ tal que $\sum_{i=1}^{n} x_{ij}x_{i\ell} = 0$, para $j \neq \ell$. Ou seja, as colunas da matriz modelo $\mathbf{X}$ são ortogonais. Encontre a ligação que faz com que $\mathrm{Corr}(\hat{\beta}_j, \hat{\beta}_\ell) = 0$, para $j \neq \ell$. Comente sobre as vantagens desse resultado.

9. Considere um experimento em que duas máquinas, M1 e M2, são observadas durante o mesmo período sendo computados para cada uma o número de peças defeituosas produzidas, conforme descrito pelo esquema abaixo.

	M1	M2
P. Defeituosas	y_1	y_2

Suponha que $Y_1 \sim \mathrm{P}(\lambda_1)$ e $Y_2 \sim \mathrm{P}(\lambda_2)$ e considere o modelo log-linear $\log(\lambda_1) = \alpha$ e $\log(\lambda_2) = \alpha + \beta$. Obtenha a variância assintótica de $\hat{\beta}$, $\mathrm{Var}_y(\hat{\beta})$, expressando-a em função de α e β. Proponha agora um modelo binomial condicional, dado $Y_1 + Y_2 = m$. Expresse a probabilidade de sucesso π em função de β. Interprete π e encontre a variância assintótica de $\hat{\beta}$, $\mathrm{Var}_{y|m}(\hat{\beta})$. Mostre que as duas variâncias assintóticas estimadas coincidem e são dadas por

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{\beta}) = \frac{(1+e^{\hat{\beta}})^2}{me^{\hat{\beta}}},$$

em que $\hat{\beta}$ é o estimador de máxima verossimilhança de β. Comente.

10. Supor $Y_1, \ldots, Y_n$ variáveis aleatórias independents tais que $Y_i \sim \mathrm{P}(\mu_i)$ e seja $\sqrt{\mu_i} = \alpha + \beta(x_i - \bar{x})$, em que $\bar{x}$ é a média amostral de $x_1, \ldots, x_n$. (i) Obtenha a matriz modelo $\mathbf{X}$. (ii) Calcule as variâncias assintóticas $\mathrm{Var}(\hat{\alpha})$ e $\mathrm{Var}(\hat{\beta})$. (iii) Mostre também que $\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = 0$ e comente. (iv) Como fica o teste de escore para testar $\mathrm{H}_0 : \beta = 0$ contra $\mathrm{H}_1 : \beta \neq 0$? Qual a distribuição nula assintótica da estatística do teste?

11. Sejam Y_{ij} variáveis aleatórias mutuamente independentes tais que $Y_{ij} \sim \mathrm{BN}(\mu_i, \nu)$ para $i = 1, 2$ e $j = 1, \ldots, m$ com parte sistemática dada por $\mu_1 = \alpha - \beta$ e $\mu_2 = \alpha + \beta$. (i) Como fica a matriz modelo $\mathbf{X}$? (ii) Calcule $\mathrm{Var}(\hat{\beta})$ e (iii) mostre que a estatística de escore para testar $\mathrm{H}_0 : \beta = 0$ contra $\mathrm{H}_1 : \beta \neq 0$ pode ser expressa na forma

$$\xi_{SR} = \frac{m\hat{\nu}^0}{2\bar{y}} \frac{(\bar{y}_2 - \bar{y}_1)^2}{(\bar{y} + \hat{\nu}^0)},$$

em que $\bar{y} = (\bar{y}_1 + \bar{y}_2)/2$ e $\hat{\nu}^0$ denota a estimativa de ν sob H_0.

12. Sejam $Y_1, \ldots, Y_n$ variáveis aleatórias independentes tais que $Y_i \sim \mathrm{BN}(\mu_i, \nu)$ com parte sistemática dada por $\log(\mu_i) = \alpha + \beta(x_i - \bar{x})$ em que $\bar{x} = \frac{\sum x_i}{n}$. (i) Como fica a matriz modelo X? (ii) Obtenha $\mathrm{Var}(\hat{\beta})$. (iii) Como fica o teste de escore para testar $\mathrm{H}_0 : \beta = 0$ contra $\mathrm{H}_1 : \beta \neq 0$? Qual a distribuição nula assintótica da estatística do teste?

13. Sejam $Y_1, \ldots, Y_n$ variáveis i.i.d. tais que $Y_i \sim \mathrm{BN}(\mu, \phi)$, $i = 1, \ldots, n$. Como fica o teste da razão de verossimilhanças para testar $\mathrm{H}_0 : \phi = 1$ contra $\mathrm{H}_1 : \phi \neq 1$? Qual a distribuição nula assintótica da estatística do teste? Como fica a estimativa de μ sob as duas hipóteses?

14. Na tabela abaixo uma amostra de 174 alunos de Estatística Básica no IME-USP foi classificada segundo o curso e o desempenho na disciplina.

	Resultado da Avaliação		
Curso	Aprovado	Reprovado	Reavaliação
Pedagogia	32	16	3
Geografia	32	18	10
Física	35	14	14

Ajustar um modelo log-linear de Poisson para explicar π_{ij}, a proporção de alunos do curso i com resultado j, em que $i, j = 1, 2, 3$. Interprete os resultados e faça uma análise de diagnóstico.

15. Supor, por um lado, o modelo log-linear de Poisson em que $Y_i \sim \mathrm{P}(\mu_i)$, $i = 1, 2, 3$, em que $\log(\mu_1) = \alpha$, $\log(\mu_2) = \alpha + \beta_2$ e $\log(\mu_3) = \alpha + \beta_3$. Fazendo $\tau = \mu_1 + \mu_2 + \mu_3$ expresse o logaritmo da função de verossimilhança desse modelo em função de (τ, β_2, β_3). Mostre que a matriz de informação de Fisher é bloco diagonal $\mathbf{K}_{\tau\beta} = \mathrm{diag}\{\mathrm{K}_\tau, \mathbf{K}_\beta\}$, em que $\boldsymbol{\beta} = (\beta_2, \beta_3)^\top$. Por outro lado, sabe-se que a distribuição condicional $\mathbf{Y} = \mathbf{a}|Y_1 + Y_2 + Y_3 = n$, em que $\mathbf{Y} = (Y_1, Y_2, Y_3)^\top$ e $\mathbf{a} = (a_1, a_2, a_3)^\top$, é multinomial $M(a_1, a_2, a_3; \pi_1, \pi_2, \pi_3)$. Supor o modelo log-linear $\log(\pi_1) = \alpha^*$, $\log(\pi_2) = \alpha^* + \beta_2$ e $\log(\pi_3) = \alpha^* + \beta_3$, em que $\alpha^* = -\log(1 + e^{\beta_2} + e^{\beta_3})$ devido à restrição $\pi_1 + \pi_2 + \pi_3 = 1$. Encontre a matriz de informação de Fisher $\mathbf{K}^*_\beta$ para $\boldsymbol{\beta} = (\beta_2, \beta_3)^\top$ no modelo multinomial. Mostre que as estimativas de máxima verossimilhança para $\boldsymbol{\beta}$ coincidem nos dois modelos log-lineares. Mostre também que $\mathbf{K}_\beta = \mathbf{K}^*_\beta$ quando $\tau = n$, comente.

16. Supor que $Y_{ij} \sim \mathrm{P}(\mu_{ij})$, para $i = 1, \ldots, r$ e $j = 1, \ldots, c$, com parte sistemática dada por

$$\log(\mu_{ij}) = \alpha + \beta_i + \gamma_j,$$

em que $\beta_1 = \gamma_1 = 0$. Supor ainda que os β_i's referem-se aos efeitos

do fator A e os γ_j's aos efeitos do fator B. Defina um modelo multinomial equivalente e mostre que a representação acima corresponde à independência (no sentido probabilístico) entre os fatores A e B.

17. A tabela abaixo (Bishop et al., 1975, p. 143) apresenta o resultado de uma pesquisa em que 1008 pessoas receberam duas marcas de detergente, X e M, e posteriormente responderam às seguintes perguntas: maciez da água (leve, média ou forte); uso anterior do detergente M (sim ou não); temperatura da água (alta ou baixa); preferência (marca X ou marca M). Esses dados estão descritos no arquivo **detergente.txt**.

			Maciez		
Temperatura	Uso de M	Preferência	Leve	Média	Forte
Alta	Sim	X	19	23	24
		M	29	47	43
	Não	X	29	33	42
		M	27	23	30
Baixa	Sim	X	57	47	37
		M	49	55	52
	Não	X	63	66	68
		M	53	50	42

Ajustar um modelo log-linear de Poisson para explicar $\pi_{ijk\ell}$, a proporção de indivíduos que responderam, respectivamente, nível de temperatura (i=1 alta, i=2 baixa), uso prévio de M (j=1 sim, j=2 não), preferência (k=1 X, k=2 M) e nível de maciez ($\ell = 1$ leve, $\ell = 2$ médio, $\ell = 3$ forte). Selecionar através do método `AIC` os efeitos principais significativos. Depois incluir apenas as interações significativas de primeira ordem. Interpretar os resultados e fazer uma análise de diagnóstico.

18. Seja o modelo trinomial em que $\pi_0 = Pr(Y = 0)$, $\pi_1 = Pr(Y = 1)$ e $\pi_2 = Pr(Y = 2)$ com a restrição $\pi_0 + \pi_1 + \pi_2 = 1$. Suponha que $Y = 0$ se $(Z_0 = 1, Z_1 = 0, Z_2 = 0)$, $Y = 1$ se $(Z_0 = 0, Z_1 = 1, Z_2 = 0)$ e $Y = 2$ se $(Z_0 = 0, Z_1 = 0, Z_2 = 1)$. Note que $Z_0 + Z_1 + Z_2 = 1$. Portanto, a função de probabilidades de (Z_0, Z_1, Z_2) fica dada por

$$g(z_0, z_1, z_2; \pi_0, \pi_1, \pi_2) = \pi_0^{z_0}\pi_1^{z_1}\pi_2^{z_2}.$$

Logo, para uma amostra aleatória de tamanho n a função de probabilidades de $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ pode ser expressa na forma

$$g(\mathbf{y}; \boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2) = \Pi_{i=1}^n \pi_{0i}^{z_{0i}}\pi_{1i}^{z_{1i}}\pi_{2i}^{z_{2i}}.$$

É usual considerar a parte sistemática

$$\log\left\{\frac{\pi_{1i}}{\pi_{0i}}\right\} = \eta_{1i} = \mathbf{x}_i^\top\boldsymbol{\beta}_1 \quad \text{e} \quad \log\left\{\frac{\pi_{2i}}{\pi_{0i}}\right\} = \eta_{2i} = \mathbf{x}_i^\top\boldsymbol{\beta}_2$$

sendo que $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$, $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{1p})^\top$ e $\boldsymbol{\beta}_2 = (\beta_{21}, \ldots, \beta_{2p})^\top$. Responda aos itens abaixo:

(a) Verifique que $\pi_{0i} = \{1 + e^{\eta_{1i}} + e^{\eta_{2i}}\}^{-1}$, $\pi_{1i} = e^{\eta_{1i}}/\{1 + e^{\eta_{1i}} + e^{\eta_{2i}}\}$ e $\pi_{2i} = e^{\eta_{2i}}/\{1 + e^{\eta_{1i}} + e^{\eta_{2i}}\}$.

(b) Encontre as funções escore $\mathbf{U}_{\beta_1}$ e $\mathbf{U}_{\beta_2}$ de $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$, respectivamente.

(c) Encontre a matriz de informação de Fisher para $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$.

(d) Desenvolva um processo iterativo para obter a estimativa de máxima verossimilhança de $\boldsymbol{\beta}$. Deixe o processo iterativo em forma matricial. Como iniciá-lo?

(e) Como fica o desvio do modelo? E o resíduo componente do desvio?

19. Supor que $Z_i \overset{\text{iid}}{\sim} \text{ZAP}(\mu, \pi)$, para $i = 1, \ldots, n$. Obtenha as estimativas de máxima verossimilhança $\hat{\mu}$ e $\hat{\pi}$ e a matriz de informação de Fisher

para (μ, π). Sugestão: supor que o total de zeros na amostra segue uma $\mathrm{B}(n, \pi)$.

20. Supor que $Z_i \overset{\text{iid}}{\sim} \text{ZANBI}(\mu, \nu, \pi)$, para $i = 1, \ldots, n$, em que a função de probabilidades de z_i fica dada por

$$f_z(z_i; \mu, \nu, \pi) = \begin{cases} \pi & \text{se} \quad z_i = 0 \\ (1-\pi)\frac{f_y(z_i;\mu,\nu)}{1-f_y(0;\mu,\nu)} & \text{se} \quad z_i = 1, 2, \ldots, \end{cases}$$

em que $f_y(y_i; \mu, \nu)$ denota a função de probabilidades de uma $\text{BN}(\mu, \nu)$. Supondo $\nu = 1$ obter a estatística da razão de verossimilhanças para testar H: $\mu = 1$ contra A: $\mu \neq 1$?

21. No arquivo **visitas.txt** são descritas as seguintes variáveis observadas numa amostra aleatória de 4380 indivíduos com mais de 65 anos e atendidos através de um programa de saúde pública durante os anos de 1987-88: (i) `nvis` (número de visitas ao médico), (ii) `hosp` (número de internações hospitalares), (iii) `altacond` (0:não, 1:sim), (iv) `baixacond` (0:não, 1:sim), (v) `nucron` (número de condições crônicas), (vi) `gênero` (0:feminino, 1:masculino), (vii) `escol` (escolaridade em anos de estudo) e (viii) `seguro` (seguro particular, 0:não, 1:sim). Quando `altacond=0` e `baixacond=0` tem-se condição média (casela de referência).

    O objetivo do estudo é explicar a demanda por serviços médicos através de modelos de regressão em que a resposta é o número de visitas ao médico e as demais variáveis como explicativas. Compare os modelos com resposta Poisson e com resposta binomial negativa. Para ler esse arquivo no R faça o seguinte:

    ```
    visitas = read.table("visitas.txt", header=TRUE)

    attach(visitas).
    ```

Fazer inicialmente uma análise descritiva dos dados, por exemplo histograma de `nvis`, boxplots robustos de `nvis` segundo os níveis das variáveis categóricas e diagrama de dispersão (com tendência) entre `nvis` e `escol`.

A distribuição do número de visitas ao médico por ser obtida através do comando

```
plot(table(nvis), xlab="Visitas", ylab="Frequência").
```

Para ajustar o modelo de Poisson com todas as variáveis explicativas use o comando

```
require(gamlss)
```

```
fit1.visitas = gamlss(nvis ∼., family=PO, data=visitas).
```

Use o comando `fit2.visitas = stepGAIC(fit1.visitas)` para selecionar um submodelo. As análises de resíduos podem ser realizadas através dos comandos

```
plot(fit2.visitas)
```

```
rqres.plot(fit2.visitas, howmany=8, ylim.all=1)
```

```
rqres.plot(fit2,visitas, howmany=40, plot="all").
```

Para ajustar o modelo com resposta binomial negativa use os comandos

```
fit3.visitas = gamlss(nvis ∼.,family=NBI, data=visitas)
```

```
fit4.vistias = stepGAIC(fit3.visitas,direction="both").
```

Qual modelo ajusta melhor os dados? Interpretar as estimativas das razões de médias do modelo selecionado apresentando estimativas intervalares de 95%. Interpretar também o `term.plot(fit4.visitas,pages=1)`.

Seria possível melhorar o ajuste modelando também o parâmetro de dispersão do modelo binomial negativo? Verifique através dos comandos

```
fit5.visitas = gamlss(nvis ~ selecionadas, ~ .,family=NBI,
data=visitas)

fit6.visitas = stepGAIC(fit5.visitas, what="sigma").
```

22. No arquivo **nitrofen.txt** (Lang et al., 1994) estão descritos os dados de um experimento com uma amostra de 50 `C.dubia` (pequeno animal invertebrado aquático de água doce), que foram submetidos a dosagens diferentes do herbicida `Nitrofen`: 0, 80, 160, 235 e 310 mg/ℓ. Para cada nível de `Nitrofen` 10 animais ficaram expostos e foi observado o total de ovos eclodidos após 3 ninhadas. Faça inicialmente uma análise descritiva dos dados, por exemplo um diagrama de dispersão entre o número de ovos eclodidos (`tovos`) contra o nível de exposição do herbicida (`dose`). Compare os ajustes de alguns modelos com resposta de Poisson para explicar o total de ovos eclodidos dado o nível de exposição. Escolha o melhor ajuste através de métodos de diagnóstico. Para o modelo selecionado faça uma interpretação dos coeficientes estimados.

23. Supor que $Z_i \overset{\text{iid}}{\sim} \text{ZAP}(\mu, \pi)$, em que $\pi = e^\alpha/(1+e^\alpha)$, para $i = 1, \ldots, n$. Obtenha a estimativa de máxima verossimilhança $\hat{\alpha}$ bem como $\text{Var}(\hat{\alpha})$. Como fica a estatística do teste da razão de verossimilhanças para testar $\text{H}_0 : \alpha = 0$ contra $\text{H}_1 : \alpha \neq 0$?

24. Supor que Y_i são variáveis aleatórias iid Poisson truncada em zero com

função de probabilidades dada por

$$f(y_i;\lambda) = \frac{e^{-\lambda}\lambda^{y_i}}{y_i!(1-e^{-\lambda})},$$

para $y_i = 1, 2, \ldots$, em que $\lambda > 0$ e $i = 1, \ldots, n$. Mostre que $\mathrm{U}_\lambda = \frac{n}{\lambda}\{\bar{y} - \lambda e^{\lambda}/(e^{\lambda}-1)\}$, obtenha $\mathrm{K}_{\lambda\lambda}$ e apresente o procedimento iterativo escore de Fisher para obter $\hat{\mu}$.

25. No arquivo **bioChemists** da biblioteca **pscl** do R são descritas as seguintes variáveis observadas numa amostra de 915 doutores formados na área de Bioquímica: (i) `art` (número de artigos publicados nos últimos 3 anos pelo doutor), (ii) `fem` (fator indicando gênero, masculino ou feminino), (iii) `mar` (fator indicando o estado civil, casado ou solteiro), (iv) `kids5` (número de crianças com até 5 anos), (v) `phd` (prestígio do departamento onde o aluno fez o doutorado) e (vi) `ment` (número de artigos publicados pelo orientador). As variáveis `fem` e `mar` são fatores, a variável `phd` é contínua e as variàveis `kids5` e `ment` são discretas.

    O objetivo do estudo é explicar o número médio de artigos publicados nos últimos 3 anos pelo doutor dadas as demais variáveis como explicativas. Para disponibilizar o arquivo faça o seguinte:

    ```
    require(gamlss)

    require(pscl)

    summary(bioChemists)

    attach(bioChemists).
    ```

    Fazer inicialmente uma análise descritiva dos dados, por exemplo tabelas de contingência entre o número de artigos publicados e os fatores gênero e estado civil, boxplot robusto da variável resposta e diagramas

de dispersão (com tendência) entre `art` e as variáveis quantitativas. Comente.

Ajustar inicialmente um modelo binomial negativo ajustado em zero apenas com variáveis explicativas no componente da localização

`fit1.bio = gamlss(art ~ ·, family=ZANBI, data=bioChemists).`

Através do procedimento `stepGAIC` selecionar um subconjunto de variáveis explicativas. Dadas as variáveis explicativas selecionadas no componente da localização, repetir o procedimento para o componente da probabilidade de zero. Após os dois procedimentos verificar se é possível remover variáveis explicativas não significativas nos dois componentes ao nível de 10%.

Fazer análises de resíduos para o modelo selecionado e interpretar os coeficientes estimados nos dois componentes.

Capítulo 6

Modelos de Quase-Verossimilhança

6.1 Introdução

Wedderburn (1974) propôs uma função biparamétrica, denominada função de quase-verossimilhança, que engloba algumas funções de verossimilhança da família exponencial. Todavia, na maioria das situações não é possível através da função de quase-verossimilhança recuperar a verdadeira distribuição da variável resposta. Se Y é a variável aleatória de interesse o logaritmo da função de quase-verossimilhnaça é definido por

$$\mathrm{Q}(\mu; y) = \frac{1}{\sigma^2} \int_y^\mu \frac{y-t}{V(t)} dt,$$

em que $V(t)$ é uma função positiva e conhecida, $-\infty < y, \mu < \infty$ e $\sigma^2 > 0$ é um parâmetro de dispersão. Como temos acima uma integral definida, segue que

$$\begin{aligned} \frac{\partial \mathrm{Q}(\mu; y)}{\partial \mu} &= \frac{y-t}{\sigma^2 V(t)} \Big|_y^\mu \\ &= \frac{y-\mu}{\sigma^2 V(\mu)}. \end{aligned}$$

Aplicando as condições abaixo de regularidade

(i) $\mathrm{E}\left\{\frac{\partial \mathrm{Q}(\mu;Y)}{\partial \mu}\right\} = 0$ e

(ii) $\mathrm{E}\left[\left\{\frac{\partial \mathrm{Q}(\mu;Y)}{\partial \mu}\right\}^2\right] = -\mathrm{E}\left\{\frac{\partial^2 \mathrm{Q}(\mu;Y)}{\partial \mu^2}\right\}$,

mostra-se facilmente que $\mathrm{E}(Y) = \mu$ e $\mathrm{Var}(Y) = \sigma^2 V(\mu)$. Ou seja, μ é a média da variável resposta e a variância de Y é proporcional a $V(\mu)$, como nos MLGs, embora nem sempre $V(\mu)$ seja uma função de variância. Uma terceira propriedade mostrada por Wedderburn (1974) é a seguinte:

(iii) $-\mathrm{E}\left\{\frac{\partial^2 \mathrm{Q}(\mu;Y)}{\partial \mu^2}\right\} \leq -\mathrm{E}\left\{\frac{\partial^2 \mathrm{L}(\mu;Y)}{\partial \mu^2}\right\}$.

Essa relação mostra que a informação a respeito de μ quando se conhece apenas a relação entre a variância e a média é menor do que a informação a respeito de μ quando se conhece a distribuição da resposta (informação de Fisher). Assim, a quantidade $\mathrm{E}\{\partial^2(\mathrm{Q}-\mathrm{L})/\partial\mu^2\}$ pode ser interpretada como o ganho quando acrescenta-se ao conhecimento da relação média-variância também o conhecimento da distribuição da resposta.

Dependendo das especificações de σ^2 e $V(\mu)$ poderemos recuperar a distribuição de Y. Abaixo são apresentados alguns exemplos.

Exemplos

Normal

Vamos supor $V(t) = 1$ e $-\infty < t, y < \infty$. Logo, o logaritmo da função de quase-verossimilhança fica dado por

$$\mathrm{Q}(\mu; y) = \int_y^\mu \frac{y-t}{\sigma^2} dt = -\frac{(y-t)^2}{2\sigma^2}|_y^\mu = -\frac{(y-\mu)^2}{2\sigma^2},$$

que é proporcional ao logaritmo da função de verossimilhança de uma $N(\mu, \sigma^2)$ para σ^2 conhecido.

Poisson

Vamos supor $V(t) = t$ e $y \geq 0, t > 0$. Logo, obtemos

$$\begin{aligned} \mathrm{Q}(\mu; y) &= \int_y^\mu \frac{y-t}{\sigma^2 t} dt \\ &= \frac{1}{\sigma^2}(y\log t - t)|_y^\mu \\ &= \frac{1}{\sigma^2}\{y\log\mu - \mu - y\log y + y\}. \end{aligned}$$

Se assumirmos $\sigma^2 = 1$ e $y > 0$ temos que $\mathrm{Q}(\mu; y)$ é proporcional ao logaritmo da função de verossimilhança de uma $\mathrm{P}(\mu)$.

Para $y = 0$ obtemos

$$\mathrm{Q}(\mu; y) = \int_0^\mu \frac{-t}{\sigma^2 t} dt = \frac{-t}{\sigma^2}|_0^\mu = -\frac{\mu}{\sigma^2},$$

que coincide quando $\sigma^2 = 1$ com $\log P(Y = 0)$, em que $Y \sim \mathrm{P}(\mu)$.

Binomial

Supor a função $V(t) = t(1-t)$, $0 \leq y \leq 1$ e $0 < t < 1$. O logaritmo da função de quase-verossimilhança fica nesse caso dado por

$$\begin{aligned} \mathrm{Q}(\mu; y) &= \int_y^\mu \frac{y-t}{\sigma^2 t(1-t)} dt \\ &= \frac{y}{\sigma^2}\int_y^\mu \frac{1}{t(1-t)} dt - \frac{1}{\sigma^2}\int_y^\mu \frac{1}{(1-t)} dt \\ &= \frac{y}{\sigma^2}\log\left(\frac{t}{1-t}\right)|_y^\mu + \frac{1}{\sigma^2}\log(1-t)|_y^\mu \\ &= \frac{y}{\sigma^2}[\log\{\mu(1-\mu) - \log\{y/(1-y)\}] + \frac{1}{\sigma^2}\{\log(1-\mu) - \log(1-y)\}, \end{aligned}$$

para $0 < y, \mu < 1$.

Para $y = 0$ temos que

$$\begin{aligned} \mathrm{Q}(\mu; y) &= \int_0^\mu \frac{-t}{\sigma^2 t(1-t)} dt \\ &= -\frac{1}{\sigma^2} \int_0^\mu \frac{t}{t(1-t)} dt \\ &= \frac{1}{\sigma^2} \log(1-t)|_0^\mu \\ &= \frac{1}{\sigma^2} \log(1-\mu), \end{aligned}$$

que para $\sigma^2 = 1$ coincide com $\log P(Y = 0)$, em que $Y \sim Be(\mu)$.

Quando $y = 1$ segue que

$$\begin{aligned} \mathrm{Q}(\mu; y) &= \int_1^\mu \frac{(1-t)}{\sigma^2 t(1-t)} dt \\ &= \frac{1}{\sigma^2} \int_1^\mu \frac{1}{t} \\ &= \frac{1}{\sigma^2} \log t|_1^\mu \\ &= \frac{1}{\sigma^2} \log \mu, \end{aligned}$$

que para $\sigma^2 = 1$ coincide com $\log P(Y = 1)$, em que $Y \sim Be(\mu)$.

Gama

Supor a função $V(t) = t^2$ e $y, t > 0$. O logaritmo da função de quase-verossimilhança fica nesse caso dado por

$$\begin{aligned} \mathrm{Q}(\mu; y) &= \int_y^\mu \frac{y-t}{\sigma^2 t^2} dt \\ &= \frac{1}{\sigma^2} (-y/t - \log t)|_y^\mu \\ &= \frac{1}{\sigma^2} \{-y/\mu - \log \mu + 1 + \log y\}. \end{aligned}$$

Para σ^2 conhecido temos que $\mathrm{Q}(y; \mu)$ é proporcional ao logaritmo da função de verossimilhança de uma $\mathrm{G}(\mu, \phi)$, em que $\phi = 1/\sigma^2$

FUNÇÃO $V(t) = t^2(1-t)^2$

Suponha $0 < t < 1$ e $0 \leq y \leq 1$. Nesse caso o logaritmo da função de quase-verossimilhança fica dada por

$$\begin{aligned} \mathrm{Q}(\mu; y) &= \frac{1}{\sigma^2} \int_y^\mu \frac{y-t}{t^2(1-t)^2} dt \\ &\propto \frac{1}{\sigma^2}[(2y-1)\log\{\mu/(1-\mu)\} - y/\mu - (1-y)/(1-\mu)]. \end{aligned}$$

A função $\mathrm{Q}(\mu; y)$ obtida acima não corresponde a nenhuma função com verossimilhança conhecida. Portanto, apenas para algumas funções de quase-verossimilhança tem-se uma função de verossimilhança correspondente.

Em particular, para as funções $V(t) = t^3$, $t > 0$, $V(t) = t(1+t)$, $t > 0$ e $V(t) = e^{-t}$, $t \in I\!R$, é possível recuperar distribuições da família exponencial uniparamétrica, bem como definir novos modelos de quae-verossimilhança.

6.2 Respostas independentes

Vamos supor que $Y_1, \ldots, Y_n$ são variáveis aleatórias independentes com logaritmo da função de quase-verossimilhança $\mathrm{Q}(\mu_i; y_i)$, $i = 1, \ldots, n$. O logaritmo da função de quase-verossimilhança correspondente à distribuição conjunta fica dado por

$$\mathrm{Q}(\boldsymbol{\mu}; \mathbf{y}) = \sum_{i=1}^{n} \mathrm{Q}(\mu_i; y_i). \tag{6.1}$$

Vamos supor ainda que

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}, \tag{6.2}$$

em que $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^T$ contém valores de variáveis explicativas, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ e $g(\cdot)$ é uma função de ligação. Note que os MLGs são um caso particular de (6.1)-(6.2).

6.2.1 Estimação

Denotando $Q(\boldsymbol{\beta}) = Q(\boldsymbol{\mu}(\boldsymbol{\beta}); \mathbf{y})$, podemos mostrar que a função quase-escore para $\boldsymbol{\beta}$ fica expressa na forma

$$\mathbf{U}_\beta = \frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2}\mathbf{D}^T\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}),$$

em que $\mathbf{D} = \partial\boldsymbol{\mu}/\partial\boldsymbol{\beta} = \mathbf{W}^{1/2}\mathbf{V}^{1/2}\mathbf{X}$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$, $\mathbf{y} = (y_1, \ldots, y_n)^T$, $\mathbf{V} = \text{diag}\{V_1, \ldots, V_n\}$, $\mathbf{W} = \text{diag}\{\omega_1, \ldots, \omega_n\}$ com $\omega_i = (d\mu/d\eta)_i^2/V_i$ e $\mathbf{X}$ é uma matriz $n \times p$ de linhas $\mathbf{x}_i^T$, $i = 1, \ldots, n$. A matriz de quase-informação para $\boldsymbol{\beta}$ fica dada por

$$\mathbf{K}_{\beta\beta} = -\text{E}\left\{\frac{\partial^2 Q(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right\} = \frac{1}{\sigma^2}\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D}.$$

A estimativa de quase-verossimilhança para $\boldsymbol{\beta}$ sai da solução da equação $\mathbf{U}_\beta = \mathbf{0}$ que pode ser resolvida pelo método escore de Fisher resultando no seguinte processo iterativo:

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + \{\mathbf{D}^{(m)T}\mathbf{V}^{-(m)}\mathbf{D}^{(m)}\}^{-1}\mathbf{D}^{(m)T}\mathbf{V}^{-(m)}\{\mathbf{y} - \boldsymbol{\mu}^{(m)}\}, \qquad (6.3)$$

$m = 0, 1, 2, \ldots$. Note que o processo iterativo (6.3) não depende de σ^2, no entanto, precisa ser iniciado numa quantidade $\boldsymbol{\beta}^{(0)}$. Mostra-se, sob certas condições de regularidade (vide, por exemplo, McCullagh e Nelder, 1989, p. 333), que $\hat{\boldsymbol{\beta}}$ é consistente e assintoticamente normal com matriz de variância-covariância dada por $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}$. O parâmetro de dispersão σ^2 deve ser estimado separadamente.

6.2.2 Estimador de momentos

Podemos verificar facilmente que

$$\text{Var}\left\{\frac{(Y_i - \mu_i)}{\sigma\sqrt{V(\mu_i)}}\right\} = 1,$$

e daí segue

$$\text{Var}\left\{\frac{(Y_i-\mu_i)}{\sqrt{V(\mu_i)}}\right\}=\sigma^2,$$

e, portanto, um estimador de momentos para σ^2 fica dado por

$$\hat{\sigma}^2=\frac{1}{(n-p)}\sum_{i=1}^{n}\frac{(y_i-\hat{\mu}_i)^2}{V(\hat{\mu}_i)}.$$

6.2.3 Função quase-desvio

É possível definir uma função tipo desvio para os modelos de quase-verossimilhança de forma similar aos MLGs. Sejam $\text{Q}(\mathbf{y};\mathbf{y})$ e $\text{Q}(\hat{\boldsymbol{\mu}};\mathbf{y})$, respectivamente, as funções de quase-verossimilhança do modelo saturado e do modelo sob investigação. A função quase-desvio não escalonada é definida por

$$\begin{aligned}\text{D}(\mathbf{y};\hat{\boldsymbol{\mu}}) &= 2\sigma^2\{\text{Q}(\mathbf{y};\mathbf{y})-\text{Q}(\hat{\boldsymbol{\mu}};\mathbf{y})\}\\ &= -2\sigma^2\text{Q}(\hat{\boldsymbol{\mu}};\mathbf{y})=-2\sigma^2\sum_{i=1}^{n}\text{Q}(\hat{\mu}_i;y_i)\\ &= 2\sum_{i=1}^{n}\int_{\hat{\mu}_i}^{y_i}\frac{y_i-t}{V(t)}dt,\end{aligned}$$

que não depende de σ^2. É natural que se compare a função quase-desvio escalonada $\text{D}^*(\mathbf{y};\hat{\boldsymbol{\mu}})=\sigma^{-2}\text{D}(\mathbf{y};\hat{\boldsymbol{\mu}})$ com os percentis da distribuição $\chi^2_{(n-p)}$, embora não seja em geral conhecida a distribuição nula de $\sigma^{-2}\text{D}(\mathbf{y};\hat{\boldsymbol{\mu}})$. Apresentamos abaixo a função quase-desvio não escalonada para alguns casos particulares supondo uma única observação.

$V(\mu)$	Componente de $\text{D}(\mathbf{y};\boldsymbol{\mu})$
μ	$-2\{y\log\mu-\mu-y\log y+y\},\ y,\mu>0$
$\mu(1-\mu)$	$-2[y\log\{\mu/(1-\mu)\}+\log(1-\mu)-\log y],\ 0<y,\mu<1$
μ^2	$-2\{1-y/\mu-\log\mu+\log y\},\ y,\mu>0$

6.2.4 Teste de hipóteses

Seja o vetor paramétrico $\boldsymbol{\beta}$ particionado tal que $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$, $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$ são subvetores de dimensão q e $p-q$, respectivamente. Suponha que temos interesse em testar $\mathrm{H}_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ contra $\mathrm{H}_1 : \boldsymbol{\beta}_1 \neq \mathbf{0}$. McCullagh (1983) mostra que também no caso de quase-verossimilhança a diferença entre duas funções quase-desvio funciona como um teste da razão de verossimilhanças. Ou seja, se denotarmos por $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}^0)$ a função quase-desvio sob H_0 e por $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})$ a função quase-desvio sob H_1, para n grande e sob H_0, temos que

$$\frac{1}{\sigma^2}\left\{\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - \mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}})\right\} \sim \chi^2_q,$$

para σ^2 fixo que pode ser estimado consistemente, como ocorre com os MLGs. Testes tipo Wald e tipo escore são também possíveis de serem desenvolvidos. Usando resultados do Capítulo 1 podemos mostrar que

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_1) = \{\mathbf{D}_1^T \mathbf{V}^{1/2} \mathbf{M}_2 \mathbf{V}^{1/2} \mathbf{D}_1\}^{-1},$$

em que $\mathbf{M}_2 = \mathbf{I} - \mathbf{H}_2$, $\mathbf{H}_2 = \mathbf{V}^{1/2}\mathbf{D}_2(\mathbf{D}_2^T\mathbf{V}\mathbf{D}_2)^{-1}\mathbf{D}_2^T\mathbf{V}^{1/2}$, $\mathbf{D}_1 = \mathbf{W}^{1/2}\mathbf{V}^{1/2}\mathbf{X}_1$ e $\mathbf{D}_2 = \mathbf{W}^{1/2}\mathbf{V}^{1/2}\mathbf{X}_2$. Assim, um teste tipo Wald fica dado por

$$\xi_W = \hat{\boldsymbol{\beta}}_1^T \hat{\mathrm{V}}\mathrm{ar}^{-1}(\hat{\boldsymbol{\beta}}_1)\hat{\boldsymbol{\beta}}_1,$$

em que $\hat{\mathrm{V}}\mathrm{ar}(\hat{\boldsymbol{\beta}}_1)$ denota que a variância está sendo avaliada em $\hat{\boldsymbol{\beta}}$. Já o teste quase-escore para testar $\mathrm{H}_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ contra $\mathrm{H}_1 : \boldsymbol{\beta}_1 \neq \mathbf{0}$ fica dado por

$$\xi_{SR} = \mathbf{U}_{\beta_1}(\hat{\boldsymbol{\beta}}^0)^T \hat{\mathrm{V}}\mathrm{ar}_0(\hat{\boldsymbol{\beta}}_1)\mathbf{U}_{\beta_1}(\hat{\boldsymbol{\beta}}^0),$$

em que

$$\begin{aligned}
\mathbf{U}_{\beta_1} &= \frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1} \\
&= \frac{1}{\sigma^2}\mathbf{D}_1^T\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}),
\end{aligned}$$

com todas as quantidades sendo avaliadas em $\hat{\boldsymbol{\beta}}^0 = (\mathbf{0}^T, \hat{\boldsymbol{\beta}}_2^{0T})^T$ e $\hat{\boldsymbol{\beta}}_2^0$ sendo a estimativa de $\boldsymbol{\beta}_2$ sob H_0. Sob H_0 e sob condições usuais de regularidade temos que, para $n \to \infty$, $\xi_W, \xi_{SR} \sim \chi_q^2$.

6.2.5 Resíduos

O não conhecimento da verdadeira função de verossimilhança de $\boldsymbol{\beta}$ dificulta o desenvolvimento de alguns métodos de diagnóstico. Tanto o estudo de resíduos como de medidas de influência dependem em geral do conhecimento de $L(\boldsymbol{\beta})$. O que tem sido proposto em modelos de quase-verossimilhança no sentido de avaliar a qualidade do ajuste são gráficos de resíduos. Uma sugestão (vide McCullagh e Nelder, 1989, Cap. 9) é o gráfico do resíduo de Pearson

$$\hat{r}_{P_i} = \frac{y_i - \hat{\mu}_i}{\hat{\sigma}\sqrt{V(\hat{\mu}_i)}}$$

contra alguma função dos valores ajustados, como por exemplo contra $g(\hat{\mu}_i)$, em que $g(\cdot)$ é a função de ligação. Espera-se uma distribuição aleatória dos resíduos em torno do eixo zero. Tendências diferentes, como por exemplo aumento da variabilidade, podem indicar que a função $V(\mu_i)$ não é adequada. Um outro resíduo que pode também ser utilizado, embora de forma descritiva, é dado por

$$t_{D_i} = \frac{\pm d_(y_i; \hat{\mu}_i)}{\hat{\sigma}\sqrt{1 - \hat{h}_{ii}}},$$

em que $d(y_i; \hat{\mu}_i)$ é a raiz quadrada com sinal de $y_i - \hat{\mu}_i$ do i-ésimo componente do quase-desvio $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$, enquanto h_{ii} é o i-ésimo elemento da diagonal principal da matriz

$$\mathbf{H} = \mathbf{V}^{-1/2}\mathbf{D}(\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}\mathbf{D}^T\mathbf{V}^{-1/2}.$$

6.2.6 Influência

Uma versão da distância de Cook para os modelos de quase-verossimilhança fica dada por

$$\mathrm{LD}_i = \frac{\hat{h}_{ii}}{(1-\hat{h}_{ii})^2}\hat{r}^2_{P_i},$$

em que $\hat{r}_{P_i}$ é o resíduo de Pearson e $\hat{h}_{ii}$ denota o i-ésimo elemento da diagonal principal da matriz $\hat{\mathbf{H}}$. Gráficos de LD_i contra a ordem das observações ou contra os valores ajustados podem revelar pontos possivelmente influentes nos parâmetros do preditor linear.

6.2.7 Seleção de Modelos

Uma extensão natural do método de Akaike para os modelos de quase-verossimilhança (ver, por exemplo, Pan, 2001) é considerar

$$\mathrm{AIC} = -2Q(\hat{\boldsymbol{\beta}}) + 2p,$$

em que $Q(\hat{\boldsymbol{\beta}})$ á a função de quase-verossimilhança avalianda em $\hat{\boldsymbol{\beta}}$.

6.2.8 Aplicações

Mosca do chifre

No arquivo **mosca.txt** é apresentado parte dos dados de um experimento desenvolvido para estudar a distribuição do número de ácaros em placas de esterco de gado bovino no estado de S. Paulo (Paula e Tavares, 1992). Essas placas são depósitos de ovos da mosca do chifre (*Haematobia irritans*), uma das pragas mais importantes da pecuária brasileira. Os ácaros são inimigos naturais da mosca do chifre uma vez que se alimentam de ovos e larvas dessas moscas. No arquivo **mosca.txt** tem-se a distribuição do número de ácaros de quatro espécies segundo algumas variáveis de interesse: (i) `N`, número de

partes da posição da placa onde foram coletados os ácaros, (ii) `Posição`, posição na placa onde foram coletados os ácaros (1: lateral, 0: central), (iii) `Região`, região onde a placa foi coletada (1: São Roque, 2: Pindamonhangaba, 3: Nova Odessa e 4: Ribeirão Preto) e (iv) `Temp`, temperatura no local da coleta (em ^oC).

Tabela 6.1
Estimativas dos parâmetros do modelo de quase-verossimilhança com função $V(\mu) = \mu^2$ ajustado aos dados sobre a mosca do chifre.

	Com todos os pontos		Sem pontos aberrantes	
Efeito	Estimativa	E/E.Padrão	Estimativa	E/E.Padrão
Constante	-0,828	-0,74	-2,575	-2,13
Posição	-0,288	-0,64	0.380	0,78
Pinda	-0,424	-0,66	-0,910	-1,31
N. Odessa	-1,224	-1,71	-1,836	-2,36
R. Preto	-2,052	-2,98	-2,589	-3,46
Temp.	0,029	0,67	0,087	1,84
σ^2	5,129		5,913	

Pensou-se inicialmente, como trata-se de dados de contagem, num modelo log-linear de Poisson para explicar o número médio de ácaros segundo as variáveis explicativas. Denotando por Y_{ijk} o número de ácaros coletados na i-ésima posição da k-ésima placa e j-ésima região, vamos supor que $Y_{ijk} \sim \mathrm{P}(\mu_{ijk})$, $\mu_{ijk} = N_{ijk}\lambda_{ijk}$, $i = 1, 2$ e $j = 1, \ldots, 6$, com N_{ijk} denotando o número de partes na i-ésima posição da k-ésima placa coletada na j-ésima região. A parte sistemática do modelo fica dada por

$$\log(\mu_{ijk}) = \log N_{ijk} + \log\lambda_{ijk}, \tag{6.4}$$

em que

$$\log(\lambda_{ijk}) = \alpha + \beta_i + \gamma_j + \delta \mathrm{Temp}_{jk}, \tag{6.5}$$

$\log N_{ijk}$ desempenha papel de *offset*, β_i denota o efeito da posição, γ_j o efeito da região e $\texttt{Temp}_{jk}$ a temperatura na j-ésima região no momento da coleta da k-ésima placa. Temos as restrições $\beta_1 = \gamma_1 = 0$. O desvio do modelo ajustado para a espécie 6 foi de $\mathrm{D}(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 318,69$ (96 graus de liberdade) indicando fortes indícios de sobredispersão. Propomos então um modelo de quase-verossimilhança com função dada por $V(\mu_{ijk}) = \mu_{ijk}$. Esse modelo parece também inadequado pelo gráfico de resíduos de Pearson $\hat{r}_{P_{ijk}} = (y_{ijk} - \hat{\mu}_{ijk})/\hat{\sigma}\sqrt{\hat{\mu}_{ijk}}$ contra $\log\hat{\mu}_{ijk}$ (Figura 6.1).

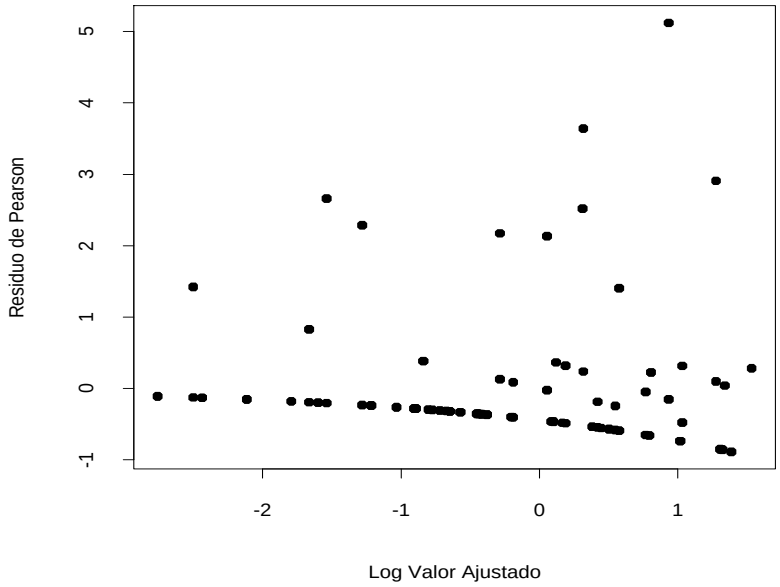

Figura 6.1: Gráfico do resíduo de Pearson contra $\log\hat{\mu}$ para o modelo ajustado com função $V(\mu) = \mu$ aos dados sobre a mosca do chifre.

Nota-se um aumento da variabilidade com o aumento do logaritmo das médias ajustadas, indício de que a variabilidade não foi totalmente controlada. Para ajustar o modelo no `R`, vamos supor que as variáveis `Posição`, `Região` e `Temp` sejam colocadas em `posicao`, `regiao` e `temp`, respectivamente,

e que logN denota o logaritmo do número de partes da placa. O número de ácaros será denotado por `acaros`. A sequência de comandos é dada abaixo

```
regiao = factor(regiao)
fit1.mosca = glm(acaros ~ posicao + regiao + temp +
offset(logN), family=quasi(link=log, variance= "mu")).
```

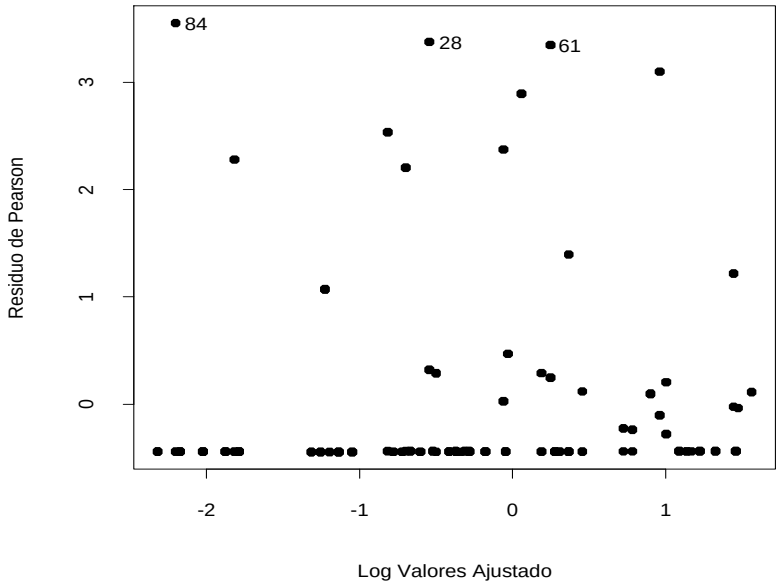

Figura 6.2: Gráfico do resíduo de Pearson contra $\log\hat{\mu}$ para o modelo ajustado com função $V(\mu) = \mu^2$ aos dados sobre a mosca do chifre.

Se colocarmos em `phi` a estimativa do parâmetro de dispersão, o resíduo de Pearson padronizado será obtido pelo comando

```
phi = summary(fit1.mosca)$dispersion
rp = resid(fit.mosca, type = "pearson")/sqrt(phi).
```

No objeto `fit.mosca` estão os principais resultados do ajuste. Propomos agora, a fim de controlar a variabilidade, um modelo de quase-verossimilhança

com função quadrática $V(\mu_{ijk}) = \mu_{ijk}^2$ e parte sistemática dada por (6.4)-(6.5). O gráfico do resíduo de Pearson contra o logaritmo das médias ajustadas (Figura 6.2) parece bastante razoável, embora apareçam 9 placas com valores para $\hat{r}_{P_{ijk}}$ acima de 2. Na Tabela 6.1 apresentamos as estimativas dos parâmetros com todas as placas e também eliminando as placas com resíduos mais aberrantes, #28, #61 e #84.

Os comandos no `R` para ajustar os dois modelos são dados abaixo

```
fit1.mosca = glm(acaros ~ posicao + regiao + temp +
offset(logN), family=quasi(link=log, variance= "mu^ 2"), maxit=50)
fit2.mosca = glm(acaros ~ posicao + regiao + temp +
offset(logN), family=quasi(link=log, variance= "mu^ 2 "), subset
= -c(28,61,84), maxit=50).
```

Nota-se pelas estimativas dos dois modelos ajustados que Nova Odessa e Ribeirão Preto apresentam um número médio de ácaros bem menor do que as outras duas regiões. Não há indícios de efeito de posição, porém a eliminação das três placas com valores mais aberrantes faz com que o efeito de temperatura fique mais acentuado, havendo indícios de que o número médio de ácaros cresce com o aumento da temperatura.

As placas #28, #61 e #84 têm em comum o fato de apresentarem um número médio de ácaros (por parte de placa) pelo menos duas vezes acima da média em temperaturas relativamente baixas. Essas placas foram coletadas nas regiões de Pindamonhangaba, Nova Odessa e Ribeirão Preto, respectivamente. Assim, é esperado que a eliminação dessas placas reduza o valor das estimativas dos efeitos dessas regiões como também aumente a estimativa do coeficiente da temperatura. A fim de que as 9 placas com resíduos mais aberrantes possam ser melhor ajustadas pode-se tentar outras formas para a função $V(\mu)$, por exemplo $V(\mu) = \mu^2(1+\mu)^2$ (vide Paula e Tavares, 1992).

Demanda de TV a cabo

Vamos reanalisar nesta seção o exemplo sobre demanda de TV a cabo discutido no Capítulo 5 sob um enfoque de modelo log-linear com resposta binomial negativa. Proporemos aqui um modelo um pouco diferente. Ao invés de ser ajustado o número médio esperado de assinantes de TV a cabo será ajustada a proporção esperada de assinantes de TV a cabo em cada área. A proporção observada é dada por `Razao = Nass/Domic`. Como $0 \leq$ `Razao` ≤ 1, propomos o seguinte modelo de quase-verossimilhança:

$$\begin{aligned} \mathrm{E}(\mathrm{Razao}_i) &= \pi_i \ \text{ e} \\ \mathrm{Var}(\mathrm{Razao}_i) &= \sigma^2 \pi_i (1-\pi_i), \end{aligned}$$

em que π_i denota a proporção esperada de assinantes na i-ésima área, $i = 1, \ldots, 40$. A parte sistemática do modelo será dada por

$$\log\left\{\frac{\pi_i}{1-\pi_i}\right\} = \alpha + \beta_1 \mathrm{Percap}_i + \beta_2 \mathrm{Taxa}_i + \beta_3 \mathrm{Custo}_i + \beta_4 \mathrm{Ncabo}_i + \beta_5 \mathrm{Ntv}_i.$$

Na Figura 6.3 é apresentado o gráfico da distância de Cook contra das observações com destaque para as áreas #5 e #14. A observação #5 corresponde a uma área de renda alta porém com uma proporção pequena de assinantes de TV a cabo, talvez devido aos altos custos de instalação e manutenção. Já a área #14 tem uma proporção alta de assinantes de TV a cabo embora as taxas também sejam altas. Também na Figura 6.3 tem-se o gráfico do resíduo

$$\hat{r}_{P_i} = \frac{(\mathrm{Razao}_i - \hat{\pi}_i)}{\hat{\sigma}\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}}$$

contra o logito dos valores ajustados e como pode-se notar há um ligeiro aumento da variabilidade com o aumento da proporção de áreas com o TV a cabo.

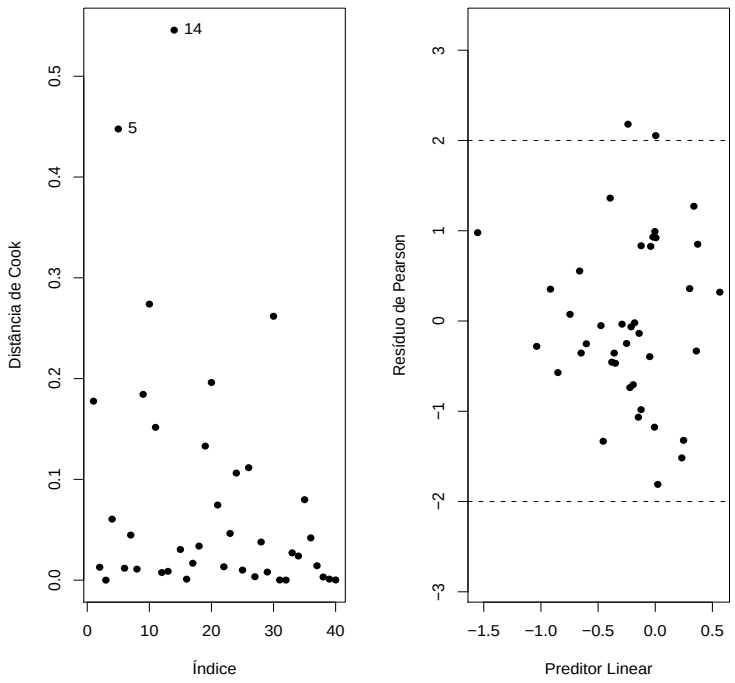

Figura 6.3: Distância de Cook e gráfico do resíduo de Pearson contra o logito de $\hat{\pi}$ para o modelo ajustado com função $V(\pi) = \pi(1-\pi)$ aos dados sobre demanda de TV a cabo.

Tabela 6.2
Estimativas dos parâmetros do modelo de quase-verossimilhança com função $V(\pi) = \pi(1-\pi)$ ajustado aos dados sobre demanda de TV a cabo.

	Com todos os pontos		Sem áreas 5 e 14	
Efeito	Estimativa	E/E.Padrão	Estimativa	E/E.Padrão
Intercepto	-2,407	-1,72	-2,440	-1,60
Percap	4×10^{-4}	2,50	4×10^{-4}	2,80
Taxa	0,023	0,93	0,016	0,64
Custo	-0,203	-1,79	-0,252	-2,27
Ncabo	0,073	1,94	0,079	2,22
Ntv	-0,216	-2,61	-0,201	-2,61
σ^2	0,114		0,098	

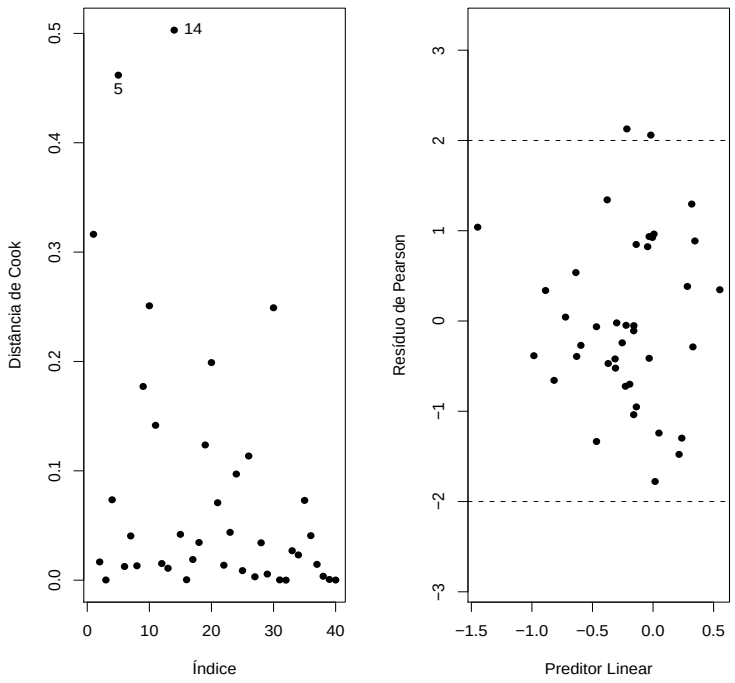

Figura 6.4: Distância de Cook e gráfico do resíduo de Pearson contra o logito de $\hat{\pi}$ para o modelo ajustado com função $V(\pi) = \pi^2(1-\pi)^2$ aos dados sobre demanda de TV a cabo.

A eliminação dessas duas áreas, como pode ser observado pela Tabela 6.2, não altera os resultados inferenciais (ao nível de 5%) com todas as observações, embora aumente a significância dos coeficientes. Nota-se que apenas o coeficiente da variável `Taxa` parece não ser significativo marginalmente.

Uma tentativa no sentido de tentar reduzir a variabilidade observada na Figura 6.3 é utilizando uma função do tipo $V(\pi) = \pi^2(1-\pi)^2$. Na Figura 6.4 temos o gráfico da distância de Cook e o gráfico do resíduo de Pearson contra o logito dos valores ajustados supondo $V(\pi) = \pi^2(1-\pi)^2$. Nota-se comportamentos muito similares àqueles encontrados na Figura 6.3. Assim, podemos assumir para esse exemplo o ajuste com a função $V(\pi) = \pi(1-\pi)$.

Nota-se, que sob esse ajuste, mais variáveis permanecem no modelo do que sob o ajuste do número esperado de domicílios com TV a cabo com resposta binomial negativa, como foi visto no Capítulo 5.

Para o ajuste do modelo de quase-verossimilhança com $V(\pi) = \pi^2(1-\pi)^2$ é preciso requerer a *library* `gnm` e usar a família `wedderburn` conforme os comandos dados abaixo

```
require(gnm)
ajuste.tvcabo = glm(razao ~ percap + taxa + custo + ncabo + ntv,
family=wedderburn).
```

Todavia, os resultados com a família `wedderburn` ficaram muito parecidos com aqueles resultados apresentados com a função $V(\pi) = \pi(1-\pi)$.

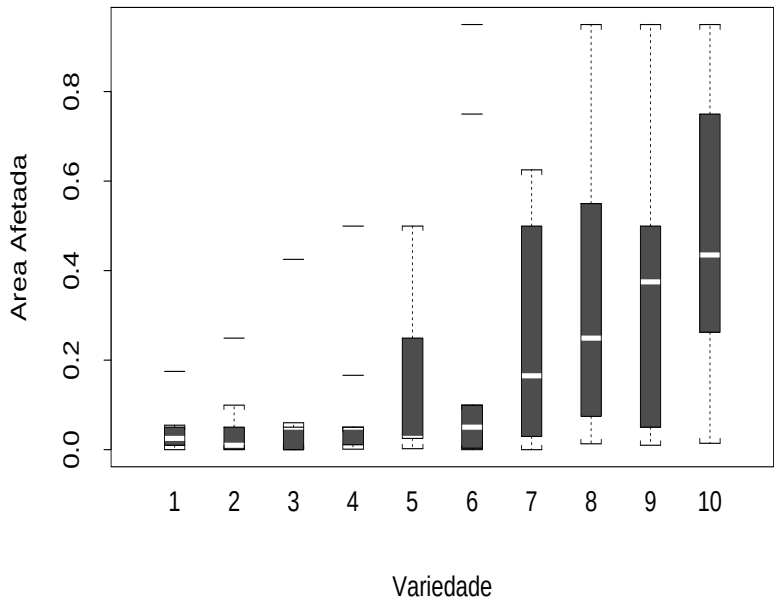

Figura 6.5: Boxplots da proporção da área afetada segundo a variedade para os dados sobre manchas na folha da cevada.

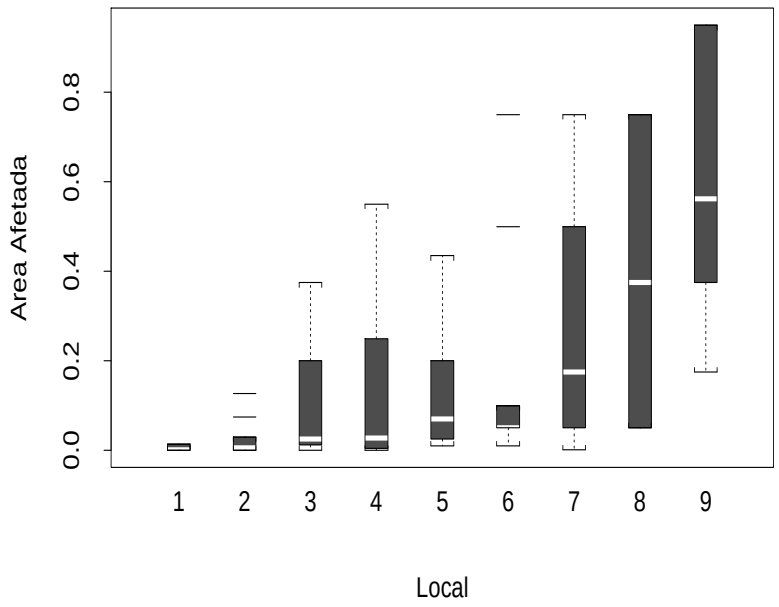

Figura 6.6: Boxplots da proporção da área afetada segundo o local para os dados sobre manchas na folha da cevada.

Manchas na folha da cevada

Esses dados estão descritos em McCullagh e Nelder (1989, Tabela 9.2) e no arquivo **cevada.txt**, em que a incidência de um tipo de mancha é observada na folha da cevada segundo 10 variedades em 9 locais diferentes. A amostra consiste de 90 observações em que a resposta é a área afetada da folha (em proporção) e os fatores são a variedade e o local.

Nas Figuras 6.5 e 6.6 são apresentados os boxplots da área afetada (em proporção) segundo a variedade e local, respectivamente. Nota-se no primeiro gráfico um aumento da mediana da proporção da área afetada e também da dispersão com a variedade. Tendência similar pode ser observada no segundo gráfico. Seja Y_{ij} a proporção da área afetada da folha da cevada correspondente ao i-ésimo local e j-ésima variedade para $i = 1, \ldots, 9$ e $j = 1, \ldots, 10$. Conforme sugerido por McCullagh e Nelder (1989, Cap. 9) vamos

supor o seguinte modelo de quase-verossimilhança:

$$\begin{aligned} \mathrm{E}(Y_{ij}) &= \pi_{ij} \quad \text{e} \\ \mathrm{Var}(Y_{ij}) &= \sigma^2 V(\pi_{ij}), \end{aligned}$$

com parte sistemática dada por

$$\log\left\{\frac{\pi_{ij}}{1-\pi_{ij}}\right\} = \alpha + \beta_i + \gamma_j,$$

em que π_{ij} denota a proporção esperada da área afetada para a j-ésima variedade do i-ésimo local, $\beta_1 = 0$ e $\gamma_1 = 0$.

Nas Figuras 6.7 e 6.8 são apresentados gráficos de diagnóstico para ajustes do modelo de quase-verossimilhança supondo $V(\pi_{ij}) = \pi_{ij}(1-\pi_{ij})$ e $V(\pi_{ij}) = \pi_{ij}^2(1-\pi_{ij})^2$, respectivamente. Nota-se ao compararmos os gráficos de resíduos que o segundo ajuste é mais adequado embora algumas observações sejam destacadas como possivelmente influentes. As observações #24, #65 e #76 apresentam proporções amostrais acima das proporções médias amostrais das variedades e locais correspondentes, enquanto a observação #52 tem uma proporção amostral abaixo da proporção média do local correspondente. A eliminação dessas observações não muda a inferência com relação às proporções médias dos locais, porém muda a inferência com relação às menores proporções médias das variedades. Em geral as estimativas de quase-verossimilhança indicam um aumento da proporção esperada da área afetada com o aumento da variedade e do local conforme descrito nos boxplots apresentados nas Figuras 6.5 e 6.6.

6.3 Classe estendida

O logaritmo da função de quase-verosssimilhança $\mathrm{Q}(\mu; y)$ assume que a função $V(\mu)$ é conhecida, logo a mudança dessa função significa que um novo modelo está sendo definido. No sentido de permitir comparações de diferentes

funções $V(\mu)$ para um mesmo modelo como também possibilitar a obtenção de uma estimativa para o erro padrão assintótico de $\hat{\sigma}^2$, Nelder e Pregibon (1987) propuseram uma (log) quase-verossimilhança estendida, definida por

$$\mathrm{Q}^{+}(\mu; y) = -\frac{1}{2\sigma^2}\mathrm{D}(y; \mu) - \frac{1}{2}\log\{2\pi\sigma^2 V(y)\},$$

em que $\mathrm{D}(y; \mu) = 2\int_{\mu}^{y}\{(y-t)/V(t)\}dt$ é o quase-desvio e $\phi = \frac{1}{\sigma^2}$ o parâmetro de dispersão.

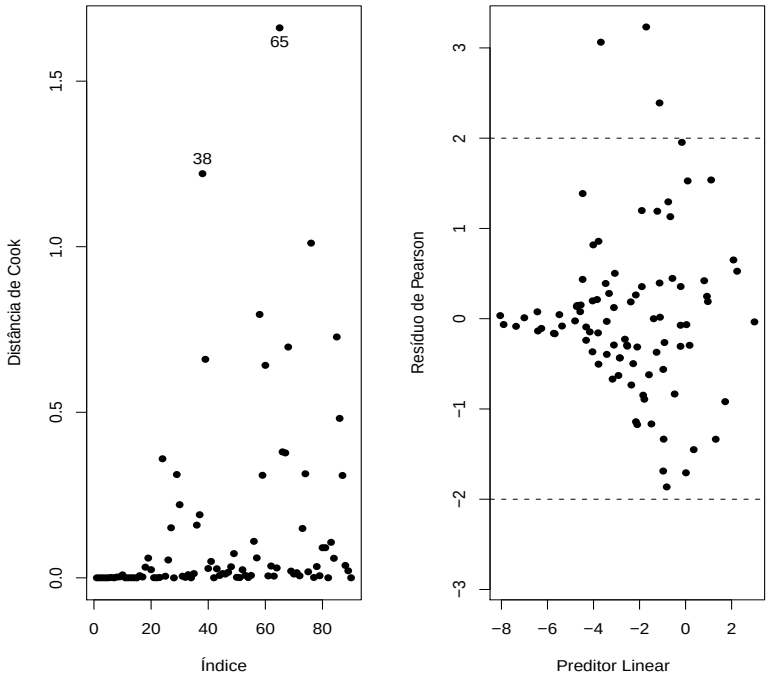

Figura 6.7: Distância de Cook e gráfico do resíduo de Pearson contra o logito de $\hat{\pi}$ para o modelo ajustado com função $V(\pi) = \pi(1-\pi)$ aos dados sobre manchas na folha da cevada.

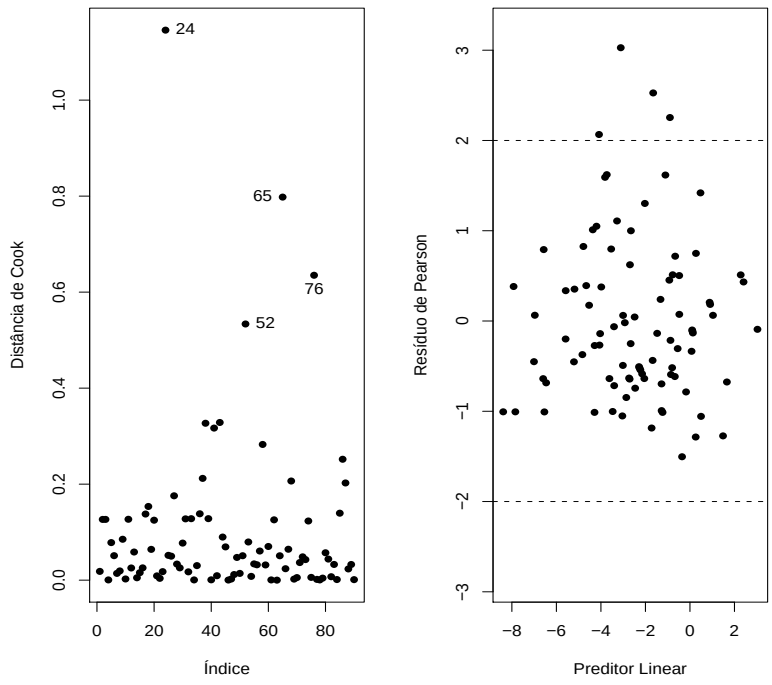

Figura 6.8: Distância de Cook e gráfico do resíduo de Pearson contra o logito de $\hat{\pi}$ para o modelo ajustado com função $V(\pi) = \pi^2(1-\pi)^2$ aos dados sobre manchas na folha da cevada.

Similarmente a Q, Q^+ não pressupõe que a distribuição completa de Y seja conhecida, mas somente os dois primeiros momentos. A estimativa de $\boldsymbol{\beta}$ maximizando-se $Q^+(\mathbf{y}; \boldsymbol{\mu})$, para uma amostra aleatória de tamanho n, coincide com a estimativa de quase-verossimilhança para $\boldsymbol{\beta}$, uma vez que Q^+ é uma função linear de Q. A estimativa de ϕ maximizando Q^+ é dada por $\hat{\phi} = D(\mathbf{y}; \hat{\boldsymbol{\mu}})/n$. Portanto, para os casos especiais em que Q^+ corresponde às distribuições normal e normal inversa, $\hat{\phi}$ corresponde à estimativa de máxima verossimilhança de ϕ. Para a distribuição gama, Q^+ difere do logaritmo da função de verossimilhança por um fator dependendo somente de ϕ. Para as

distribuições de Poisson, binomial e binomial negativa, Q^+ é obtida do logaritmo da função de verossimilhança correspondente substituindo qualquer fatorial $k!$ pela aproximação de Stirling $k! \cong (2\pi k)^{\frac{1}{2}} k^k e^{-k}$. Discussões mais interessantes e aplicações da classe estendida são dadas em Nelder e Pregibon (1987).

6.4 Respostas correlacionadas

A fim de estabelecermos a notação a ser utilizada nesta seção, denotaremos por $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ir_i})^T$ o vetor resposta multivariado para a i-ésima unidade experimental, $i = 1, \ldots, n$, e assumiremos em princípio que apenas é conhecida a distribuição marginal de Y_{it}, dada por

$$f(y; \theta_{it}, \phi) = \exp[\phi\{y\theta_{it} - b(\theta_{it})\} + c(y, \phi)], \tag{6.6}$$

em que $\mathrm{E}(Y_{it}) = \mu_{it} = b'(\theta_{it})$, $\mathrm{Var}(Y_{it}) = \phi^{-1} V_{it}$, $V_{it} = d\mu_{it}/d\theta_{it}$ é a função de variância e $\phi^{-1} > 0$ é o parâmetro de dispersão, em geral desconhecido. Podemos definir um modelo linear generalizado para cada instante t acrescentando a (6.6) a parte sistemática

$$g(\mu_{it}) = \eta_{it}, \tag{6.7}$$

em que $\eta_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta}$ é o preditor linear, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ é um vetor de parâmetros desconhecidos a serem estimados, $\mathbf{x}_{it} = (x_{it1}, \ldots, x_{itp})^T$ representa os valores de variáveis explicativas observadas para a i-ésima unidade experimental no tempo t e $g(\cdot)$ é a função de ligação.

A função escore e a matrix de informação para $\boldsymbol{\beta}$, ignorando-se a estrutura de correlação intraunidade experimental, ficam, respectivamente, dadas por

$$\mathbf{U}_\beta = \phi \sum_{i=1}^{n} \mathbf{D}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \tag{6.8}$$

e

$$\mathbf{K}_{\beta\beta} = \phi \sum_{i=1}^{n} \mathbf{D}_i^T \mathbf{V}_i \mathbf{D}_i, \tag{6.9}$$

em que $\mathbf{D}_i = \mathbf{W}_i^{\frac{1}{2}} \mathbf{V}_i^{\frac{1}{2}} \mathbf{X}_i$, $\mathbf{X}_i$ é uma matriz $r_i \times p$ de linhas $\mathbf{x}_{it}^T$, $\mathbf{W}_i = \text{diag}\{\omega_{i1}, \ldots, \omega_{ir_i}\}$ é a matriz de pesos com $\omega_{it} = (d\mu_{it}/d\eta_{it})^2/V_{it}$, $\mathbf{V}_i = \text{diag}\{V_{i1}, \ldots, V_{ir_i}\}$, $\mathbf{y}_i = (y_{i1}, \ldots, y_{ir_i})^T$ e $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{ir_i})^T$. Quando há ligação canônica a função escore e a matriz de informação de Fisher ficam dadas por $\mathbf{U}_\beta = \phi \sum_{i=1}^{n} \mathbf{X}_i^T (\mathbf{y}_i - \boldsymbol{\mu}_i)$ e $\mathbf{K}_{\beta\beta} = \phi \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}_i \mathbf{X}_i$, respectivamente. O estimador de $\boldsymbol{\beta}$, ignorando-se a estrutura de correlação intraunidade experimental, sai da equação $\mathbf{U}_\beta = \mathbf{0}$. Esse estimador é consistente e assintoticamente normal. Note que podemos supor que a distribuição marginal de Y_{it} é desconhecida assumindo uma função $V(\mu_{it})$ diferente daquela que caracteriza a distribuição de Y_{it}. Nesse caso, teremos um modelo de quase-verossimilhança em cada instante t com função escore e matriz de informação, ignorando-se a estrutura de correlação, dadas por (6.8) e (6.9), respectivamente.

Um tópico de pesquisa importante, que tem interessado a vários pesquisadores, é o desenvolvimento de metodologias para a estimação dos parâmetros de interesse quando os dados são correlacionados e a distribuição marginal não é normal, como é o caso introduzido nesta seção. Uma maneira de resolver o problema é ignorar a estrutura de correlação, como vimos acima, produzindo estimadores consistentes e assintoticamente normais, porém muitas vezes com perda de eficiência. Uma outra maneira, que descreveremos a seguir, é introduzindo alguma estrutura de correlação na função escore, produzindo um novo sistema de equações para estimar $\boldsymbol{\beta}$. A fim de facilitarmos o entendimento dessa metodologia, vamos supor inicialmente que os dados são não correlacionados e que a matriz de correlação correspondente ao i-ésimo grupo é denotada por $\mathbf{R}_i$. Logo, teremos $\mathbf{R}_i = \mathbf{I}_{r_i}$. A matriz de

variância-covariância para $\mathbf{Y}_i$, por definição, é dada por

$$\text{Var}(\mathbf{Y}_i) = \phi^{-1}\mathbf{V}_i^{\frac{1}{2}}\mathbf{R}_i\mathbf{V}_i^{\frac{1}{2}}, \tag{6.10}$$

que no caso de dados não correlacionados fica simplesmente dada por $\phi^{-1}\mathbf{V}_i$. A ideia é introduzirmos em (6.10) uma matriz de correlação não diagonal, por exemplo dada por $\mathbf{R}_i(\boldsymbol{\beta})$, com reflexos na função escore que passaria a depender também de $\mathbf{R}_i(\boldsymbol{\beta})$. O incoveniente dessa proposta é o fato da correlação, que é restrita ao intervalo $[-1, 1]$, depender de $\boldsymbol{\beta}$, o que aumentaria a complexidade do processo de estimação. A solução encontrada para contornar esse problema foi dada por Liang e Zeger (1986) que propuseram uma matriz de correlação dada por $\mathbf{R}_i(\boldsymbol{\rho})$, em que $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_q)^T$ é um vetor de parâmetros de perturbação que não dependem de $\boldsymbol{\beta}$. Ou seja, os parâmetros da matriz de correlação não dependem dos parâmetros de posição.

Para entender melhor essa proposta definimos

$$\boldsymbol{\Omega}_i = \phi^{-1}\mathbf{V}_i^{\frac{1}{2}}\mathbf{R}_i(\boldsymbol{\rho})\mathbf{V}_i^{\frac{1}{2}},$$

em que $\boldsymbol{\Omega}_i$ é a matriz de variância-covariância de $\mathbf{Y}_i$ se a verdadeira correlação entre os elementos de $\mathbf{Y}_i$ for dada por $\mathbf{R}_i(\boldsymbol{\rho})$. Note que $\mathbf{R}_i(\boldsymbol{\rho})$ é uma matriz $r_i \times r_i$ que depende de um número finito de parâmetros $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_q)^T$, sendo denominada matriz trabalho. Para estimarmos $\boldsymbol{\beta}$ devemos resolver o seguinte sistema de equações:

$$\mathbf{S}_\beta(\hat{\boldsymbol{\beta}}_G) = \mathbf{0}, \tag{6.11}$$

denominado equações de estimação generalizadas (EEGs), em que

$$\mathbf{S}_\beta(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{D}_i^T\boldsymbol{\Omega}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i).$$

Note que (6.11) reduz-se a $\mathbf{U}_\beta = \mathbf{0}$ quando $\mathbf{R}_i(\boldsymbol{\rho}) = \mathbf{I}_{r_i}$, isto é, quando é ignorada a estrutura de correlação intraunidade experimental. Na verdade $\mathbf{S}_\beta(\boldsymbol{\beta})$

depende também de ϕ e $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_q)^T$ que são estimados separadamente de $\boldsymbol{\beta}$.

6.4.1 Estimação

O processo iterativo para a estimação de $\boldsymbol{\beta}$, que é uma modificação do método escore de Fisher, é dado por

$$\begin{aligned} \boldsymbol{\beta}_G^{(m+1)} &= \boldsymbol{\beta}_G^{(m)} + \{\sum_{i=1}^{n} \mathbf{D}_i^{(m)T} \boldsymbol{\Omega}_i^{-(m)} \mathbf{D}_i^{(m)}\}^{-1} \times \\ & \quad [\sum_{i=1}^{n} \mathbf{D}_i^{(m)T} \boldsymbol{\Omega}_i^{-(m)} \{\mathbf{y}_i - \boldsymbol{\mu}_i^{(m)}\}], \end{aligned} \tag{6.12}$$

$m = 0, 1, 2 \ldots$. As estimativas $\hat{\phi}$ e $\hat{\boldsymbol{\rho}}$ são dadas inicialmente e modificadas separadamente a cada passo do processo iterativo.

Supondo que $\hat{\boldsymbol{\rho}}$ e $\hat{\phi}$ são estimadores consistentes de $\boldsymbol{\rho}$ e ϕ, respectivamente, temos que

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) \rightarrow_d N_p(\mathbf{0}, \boldsymbol{\Sigma}),$$

em que

$$\boldsymbol{\Sigma} = \lim_{n\to\infty} [n(\sum_{i=1}^{n} \mathbf{D}_i^T \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i)^{-1} \{\sum_{i=1}^{n} \mathbf{D}_i^T \boldsymbol{\Omega}_i^{-1} \mathrm{Var}(\mathbf{Y}_i) \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i\} (\sum_{i=1}^{n} \mathbf{D}_i^T \boldsymbol{\Omega}_i^{-1} \mathbf{D}_i)^{-1}].$$

Se a matriz de correlação $\mathbf{R}_i(\boldsymbol{\rho})$ é definida corretamente, então um estimador consistente para $\mathrm{Var}(\hat{\boldsymbol{\beta}}_G)$ é dado por $\mathbf{H}_1^{-1}(\hat{\boldsymbol{\beta}}_G)$, em que

$$\mathbf{H}_1(\hat{\boldsymbol{\beta}}_G) = \sum_{i=1}^{n} (\hat{\mathbf{D}}_i^T \hat{\boldsymbol{\Omega}}_i^{-1} \hat{\mathbf{D}}_i),$$

com $\hat{\mathbf{D}}_i$ sendo avaliado em $\hat{\boldsymbol{\beta}}_G$ e $\hat{\boldsymbol{\Omega}}_i$ avaliado em $(\hat{\phi}, \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\beta}}_G)$. Entretanto, se a matriz trabalho $\mathbf{R}_i(\boldsymbol{\rho})$ é definida incorretamente $\mathbf{H}_1^{-1}(\hat{\boldsymbol{\beta}}_G)$ pode ser inconsistente. Um estimador robusto para $\mathrm{Var}(\hat{\boldsymbol{\beta}}_G)$, sugerido por Liang and Zeger

(1986), é dado por

$$\hat{\mathbf{V}}_G = \mathbf{H}_1^{-1}(\hat{\boldsymbol{\beta}}_G)\mathbf{H}_2(\hat{\boldsymbol{\beta}}_G)\mathbf{H}_1^{-1}(\hat{\boldsymbol{\beta}}_G),$$

em que $\mathbf{H}_2(\hat{\boldsymbol{\beta}}_G) = \sum_{i=1}^{n}\{\hat{\mathbf{D}}_i^T\hat{\boldsymbol{\Omega}}_i^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^T\hat{\boldsymbol{\Omega}}_i^{-1}\hat{\mathbf{D}}_i\}$. O estimador $\hat{\mathbf{V}}_G$ é consistente mesmo se a matriz trabalho for definida incorretamente.

6.4.2 Estruturas de correlação

Não estruturada

Quando a matriz de correlação $\mathbf{R}_i$ é não estruturada teremos $r_i(r_i-1)/2$ parâmetros para serem estimados. Denotando $\mathbf{R}_i = \{R_{ijj'}\}$, o (j,j')-ésimo elemento de $\mathbf{R}_i$ poderá ser estimado por

$$\hat{R}_{jj'} = \frac{\phi}{n}\sum_{i=1}^{n}\frac{(y_{ij}-\hat{\mu}_{ij})}{\sqrt{\hat{V}_{ij}}}\frac{(y_{ij'}-\hat{\mu}_{ij'})}{\sqrt{\hat{V}_{ij'}}}.$$

Simétrica ou permutável

Neste caso assumimos $\mathbf{R}_i = \mathbf{R}_i(\rho)$, em que o (j,j')-ésimo elemento de $\mathbf{R}_i$ fica dado por $R_{ijj'} = 1$, para $j = j'$, e $R_{ijj'} = \rho$, para $j \neq j'$. Um estimador consistente para ρ fica dado por

$$\hat{\rho} = \frac{\phi}{n}\sum_{i=1}^{n}\frac{1}{r_i(r_i-1)}\sum_{j=1}^{r_i}\sum_{j'=1,j'\neq j}^{r_i}\frac{(y_{ij}-\hat{\mu}_{ij})}{\sqrt{\hat{V}_{ij}}}\frac{(y_{ij'}-\hat{\mu}_{ij'})}{\sqrt{\hat{V}_{ij'}}}.$$

Autoregressiva AR(1)

Aqui também assumimos $\mathbf{R}_i = \mathbf{R}_i(\rho)$, em que o (j,j')-ésimo elemento de $\mathbf{R}_i$ fica dado por $R_{ijj'} = 1$, para $j = j'$, e $R_{ijj'} = \rho^{|j-j'|}$, para $j \neq j'$. Um estimador consistente para ρ fica dado por

$$\hat{\rho} = \frac{\phi}{n}\sum_{i=1}^{n}\frac{1}{(r_i-1)}\sum_{j=1}^{r_i-1}\frac{(y_{ij}-\hat{\mu}_{ij})}{\sqrt{\hat{V}_{ij}}}\frac{(y_{i(j+1)}-\hat{\mu}_{i(j+1)})}{\sqrt{\hat{V}_{i(j+1)}}}.$$

Parâmetro de dispersão

O parâmetro de dispersão ϕ^{-1} pode ser estimado consistentemente por

$$\hat{\phi}^{-1} = \frac{1}{(N-p)} \sum_{i=1}^{n} \sum_{j=1}^{r_i} \frac{(y_{ij} - \hat{\mu}_{ij})^2}{\hat{V}_{ij}},$$

em que $N = \sum_{i=1}^{n} r_i$. Assim, o processo iterativo (6.12) deve alternar com as estimativas para ρ e ϕ até a convergência.

Testes de hipóteses para $\boldsymbol{\beta}$ ou para subconjuntos de $\boldsymbol{\beta}$ podem ser desenvolvidos através de estatísticas tipo Wald com a matriz de variância-covariância estimada $\hat{\mathbf{V}}_G$.

6.4.3 Métodos de diagnóstico

Técnicas de diagnóstico para EEGs podem ser encontradas, por exemplo, em Hardin e Hilbe (2003) e Venezuela et al. (2007) e mais recentemente em Venezuela et al. (2011). Os procedimentos apresentados a seguir foram extraídos de Venezuela et al. (2007).

Resíduos

Aplicando para as EEGs um procedimento similar àquele apresentado na Seção 2.8.2 chega-se ao seguinte resíduo de Pearson:

$$\hat{r}_{P_{ij}} = \frac{\mathbf{e}_{ij}^T \hat{\mathbf{A}}_i^{\frac{1}{2}} (\hat{\mathbf{V}}_i \hat{\mathbf{W}}_i)^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)}{\sqrt{1 - \hat{h}_{ijj}}},$$

para $i = 1, \ldots, n$ e $j = 1, \ldots, r_i$, em que $\mathbf{A}_i^{\frac{1}{2}} = \phi \mathbf{W}_i^{\frac{1}{2}} \mathbf{R}_i^{-1} \mathbf{W}_i^{\frac{1}{2}}$ é uma matriz de dimensão $r_i \times r_i$, $\mathbf{e}_{ij}^T$ é um vetor de dimensão $1 \times r_i$ de zeros com 1 na j-ésima posição e h_{ijj} é o j-ésimo elemento da diagonal principal da matriz

$$\mathbf{H}_i = \mathbf{A}_i^{\frac{1}{2}} \mathbf{X}_i (\mathbf{X}^T \mathbf{A} \mathbf{X})^{-1} \mathbf{X}_i^T \mathbf{A}_i^{\frac{1}{2}},$$

em que $\mathbf{X} = (\mathbf{X}_1^T, \ldots, \mathbf{X}_n^T)^T$ tem dimensão $N \times p$ e $\mathbf{A} = \text{diag}\{\mathbf{A}_1, \ldots, \mathbf{A}_n\}$ tem dimensão $N \times N$ com $N = \sum_{i=1}^{n} r_i$.

Alavanca

Duas medidas de alavanca são usualmente aplicadas em EEGs. Medida de alavanca referente ao j-ésimo indivíduo do i-ésimo grupo, dada por $\hat{h}_{ijj}$ e medida de alavanca referente ao i-ésimo grupo, definida por

$$\hat{h}_i = \frac{1}{r_i} \sum_{j=1}^{r_i} h_{ijj}.$$

Gráficos de índices para $\hat{h}_{ijj}$ e $\hat{h}_i$ ou contra os valores ajustados são recomendados.

Influência

Uma versão aproximada da distância de Cook para avaliar o impacto da eliminar individual das observações na estimativa $\hat{\boldsymbol{\beta}}_G$ é dada por

$$\text{LD}_{ij} = \frac{\hat{h}_{ijj}}{(1 - \hat{h}_{ijj})} \hat{r}^2_{P_{ij}}.$$

Gráficos de índices para LD_{ij} são recomendados.

6.4.4 Seleção de modelos

Uma proposta de critério para seleção de modelos em EEGs (ver, por exemplo, Hardin e Hilbe, 2003) é dado por

$$\text{QIC} = -2Q(\hat{\boldsymbol{\beta}}_G) + 2\text{tr}(\hat{\mathbf{V}}_G \hat{\mathbf{H}}_{1I}),$$

em que $\hat{\boldsymbol{\beta}}_G$ é a estimativa de quase-verossimilhança para uma matriz específica de correlação $\mathbf{R}_i(\boldsymbol{\rho})$ e $\mathbf{H}_{1I}$ é a matriz $\mathbf{H}_1$ avaliada sob a estrutura

de independência. Esse critério pode ser aplicado para selecionar submodelos encaixados ou para selecionar a matriz de correlação para um modelo específico.

6.5 Exemplos

6.5.1 Ataques epilépticos

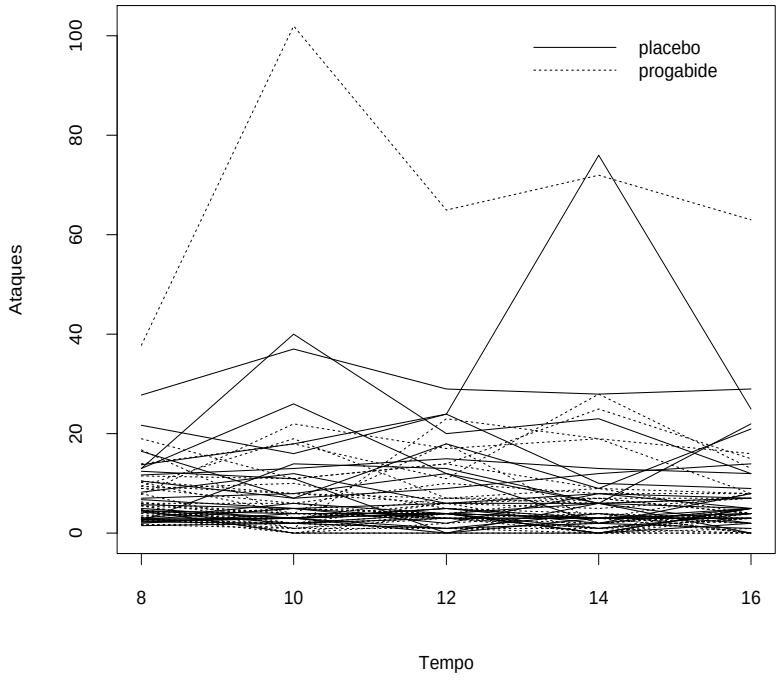

Figura 6.9: Gráfico de perfis com o número de ataques por período de 2 semanas.

No arquivo **ataques.txt** (Diggle et al., 1994, Seção 8.4) são resumidos os resultados de um ensaio clínico com 59 indivíduos epilépticos os quais foram aleatorizados de modo que cada um recebesse uma droga antiepiléptica

denominada *progabide* ou placebo. Os dados de cada indivíduo consistiram de um número inicial de ataques epilépticos num período de oito semanas antes do tratamento, seguido do número de ataques em cada período de duas semanas, num total de quatro períodos, após o tratamento. O interesse da pesquisa é saber se a droga reduz a taxa de ataques epilépticos.

Para ajustar esses modelos no `R` usaremos a *library* `gee`, que deve ser acionada através do comando

```
require(gee).
```

Os ajustes podem ser feitos de forma muito similar aos MLGs desde que os dados estejam descritos de forma apropriada. Existem outras formas de gerar dados longitudinais através de outras subrotinas que facilitam, por exemplo, a elaboração de gráficos de perfis. Nesses casos, será necessário informarmos nos comandos de ajuste como as unidades experimentais estão dispostas e o tipo de correlação intraunidade experimental a ser assumida.

No caso dos ataques epilépticos uma possível distribuição marginal para os dados é a distribuição de Poisson, uma vez que tem-se dados de contagem. Contudo, observando-se a tabela abaixo, onde estão descritos os valores amostrais para a razão variância/média para os 10 grupos experimentais, nota-se um forte indício de sobredispersão sugerindo que o parâmetro de dispersão ϕ não deve ser fixado como sendo igual a um.

	Antes	Per1	Per2	Per3	Per4
Placebo	22,13	10,98	8,04	24,50	7,24
Progradibe	24,76	38,77	16,70	23,75	18,79

Para compararmos o número de ataques epilépticos nos 10 períodos experimentais, devemos padronizar os valores referentes ao período anterior ao tratamento em que os pacientes foram observados por 8 semanas. Assim,

será possível uma comparação com os demais períodos de 2 semanas. Na Figura 6.9 temos o gráfico de perfis com os dois tratamentos. Nota-se que pelo menos um paciente (#49), que foi tratado com a droga *progabide*, apresenta um número alto de ataques antes e depois do tratamento.

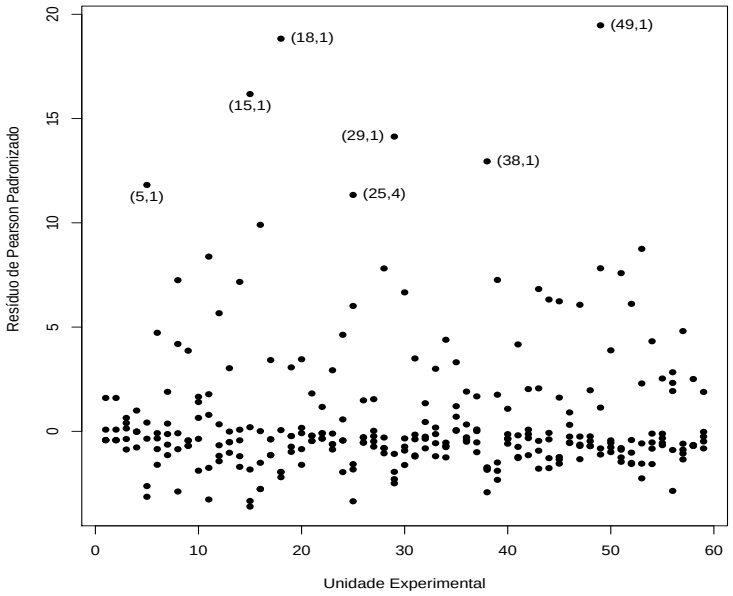

Figura 6.10: Gráfico do resíduo de Pearson referente ao modelo de Poisson com estrutura de correlação permutável ajustado aos dados sobre ataques epilépticos.

Vamos supor então que Y_{ijk} representa o número de ataques epilépticos ocorridos com o k-ésimo indivíduo do i-ésimo grupo no j-ésimo período. Assumimos que $Y_{ijk} \sim \mathrm{P}(\lambda_{ij} t_j)$, t_j denota o número de semanas do j-ésimo período, $i = 1, 2$; $j = 0, 1, 2, 3, 4$ e $k = 1, \ldots, r_{ij}$, em que $r_{1j} = 28$ (grupo placebo), $r_{2j} = 31$ (grupo tratado), $t_0 = 8$ e $t_1 = t_2 = t_3 = t_4 = 2$. Assumi-

mos também uma estrutura de correlação permutável para cada indivíduo, isto é, $\text{Corr}(Y_{ijk}, Y_{ijk'}) = \rho$, para $k \neq k'$ e (i, j) fixos. A parte sistemática do modelo será dada por

$$\begin{aligned}
\log(\lambda_{10}) &= \alpha, \\
\log(\lambda_{1j}) &= \alpha + \beta, \\
\log(\lambda_{20}) &= \alpha + \gamma \quad \text{e} \\
\log(\lambda_{2j}) &= \alpha + \gamma + \beta + \delta,
\end{aligned}$$

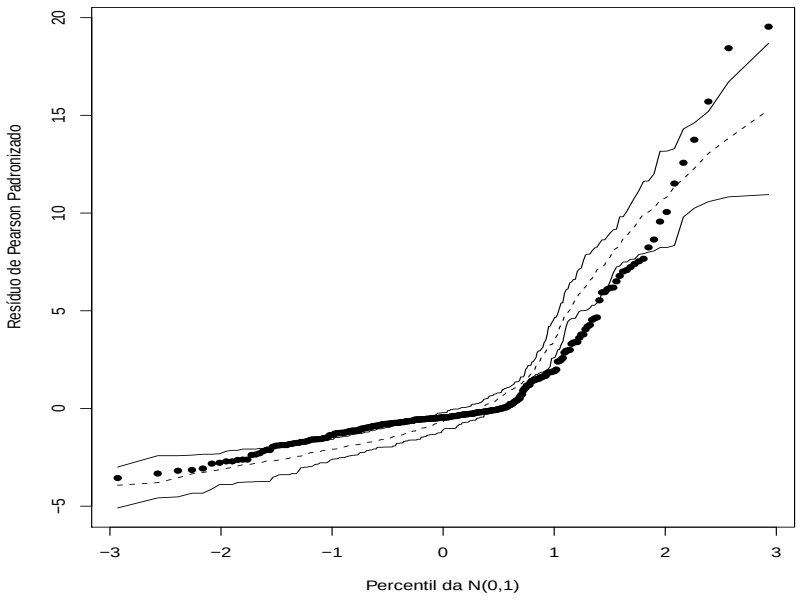

Figura 6.11: Gráfico normal de probabilidades referente ao modelo de Poisson com estrutura de correlação permutável ajustado aos dados sobre ataques epilépticos.

para $j = 1, 2, 3, 4$, em que α denota o nível base, β o efeito de tratamento, γ

o efeito de grupo e δ a interação entre tratamento e grupo. Note que, antes do tratamento, o logaritmo da razão entre as taxas dos dois grupos é dado por

$$\log\{\lambda_{20}/\lambda_{10}\} = \alpha + \gamma - \alpha = \gamma. \tag{6.13}$$

Após o tratamento, o logaritmo da razão entre as taxas fica dado por

$$\log\{\lambda_{2j}/\lambda_{1j}\} = \alpha + \gamma + \beta + \delta - \alpha - \beta = \gamma + \delta. \tag{6.14}$$

Portanto, se o tratamento não é eficaz espera-se que o logaritmo da razão não mude após o tratamento. Logo, avaliar a eficiência do tratamento equivale a testar $\mathrm{H}_0 : \delta = 0$ contra $\mathrm{H}_1 : \delta \neq 0$.

Tabela 6.3
Estimativas dos parâmetros do modelo log-linear de Poisson aplicado aos dados sobre ataques epilépticos.

	Com todos os pacientes		Sem o paciente #49	
Parâmetro	Estimativa	z-robusto	Estimativa	z-robusto
α	1,347	8,564	1,347	8,564
β	0,112	0,965	0,112	0,965
γ	0,027	0,124	-0,107	-0,551
δ	-0,105	-0,491	-0,302	-1,768
ρ	0,771		0,593	
ϕ^{-1}	19,68		10,53	

Se denotarmos por $\mu_{ij} = \mathrm{E}(Y_{ijk})$, a parte sistemática do modelo em função das médias fica dada por

$$\log(\mu_{ij}) = \log(t_j) + \log(\lambda_{ij}),$$

em que $\log t_j$ desempenha o papel de *offset*. Para ajustarmos esse modelo no `R` deve-se seguir a sequência abaixo de comandos

```
fit1.ataques = gee(ataques ~ grupo + periodo + grupo*perido +
offset(log(semanas)), id=paciente, family=poisson,
corstr="exchangeable"),
```

em que `grupo` representa o grupo (=0 placebo, =1 progabide), `periodo` representa o período (=0 antes, =1 depois), `semanas` o número de semanas, `paciente` o número do paciente (são 59 pacientes) e `corstr` o tipo de correlação a ser assumida.

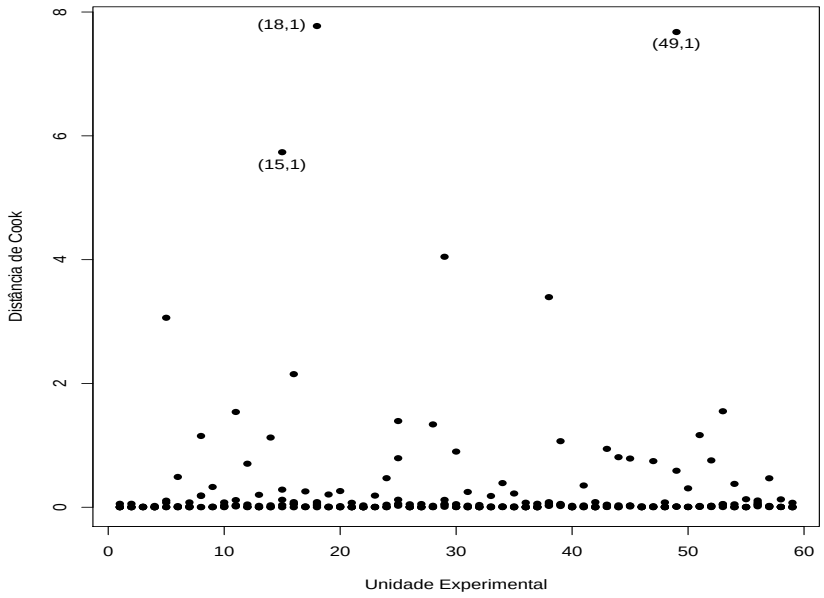

Figura 6.12: Distância de Cook referente ao modelo de Poisson com estrutura de correlação permutável ajustado aos dados sobre ataques epilépticos.

As estimativas dos parâmetros (erro padrão aproximado) são apresentadas na Tabela 6.3. Não há portanto nenhum indício de efeito de tratamento. Para a análise de resíduos vamos considerar o resíduo de Pearson $\hat{r}_{P_{ij}}$ definido na Seção 6.4.3. A geração de envelopes para esse resíduo é um pouco mais complexa do que no caso usual de respostas independentes, uma vez que requer o conhecimento da distribuição conjunta das respostas de cada

indivíduo. No entanto, mesmo quando essa distribuição não é totalmente desconhecida é possível, em alguns casos, gerar a distribuição empírica dos dados (vide, por exemplo, Venezuela et al., 2007).

Nota-se pela Tabela 6.3 que a estimativa do parâmetro de dispersão ϕ^{-1} é muito diferente da suposição de $\phi = 1$ para modelos com resposta de Poisson, sugerindo indícios fortes de sobredispersão. Assim, para uma análise de resíduos mais apropriada deve-se considerar o resíduo de Pearson padronizado $\hat{r}^*_{P_{ij}} = \sqrt{\hat{\phi}}\hat{r}_{P_{ij}}$ cujo gráfico é descrito na Figura 6.10. Nota-se 7 resíduos com valores superiores a 10,0, todos referentes a medidas de diferentes pacientes. O gráfico normal de probabilidades para o resíduo de Pearson padronizado $\hat{r}^*_{P_{ij}}$ (Figura 6.11) mostra alguns afastamentos da suposição de modelo marginal de Poisson, provavelmente devido à sobredispersão que não foi totalmente controlada.

Finalmente, na Figura 6.12, temos a distância de Cook aproximada em que três medidas se destacam. Nota-se novamente uma medida referente ao paciente (#49) cujo perfil destoa na Figura 6.9. Vamos fazer um estudo das estimativas não considerando esse paciente no ajuste. Os comandos em `R` são dados abaixo:

```
fit2.ataques = gee(ataques ~ grupo + periodo + grupo*perido +
offset(log(semanas)), id=paciente, subset=-c(241, 242, 243, 244,
245), family=poisson, corstr="exchangeable").
```

As novas estimativas (vide Tabela 6.3) indicam evidência de que o tratamento com a droga *progabide* reduz o número médio de ataques epilépticos, ou seja, há mudança inferencial em relação ao modelo com todos os pontos.

6.5.2 Condição Respiratória

Vamos considerar agora um exemplo discutido em Myers et al.(2002, Seção 6.5) que envolve a comparação de dois tratamentos aplicados em pacientes com problemas respiratórios. Um total de 56 pacientes foi considerado no estudo sendo que 27 receberam o tratamento com uma droga ativa enquanto que os 29 pacientes restantes receberam placebo. Cada paciente foi observado em quatro ocasiões em que mediu-se a condição respiratória (boa ou ruim). Foram também observados o sexo e a idade (em anos) de cada paciente além da pré-existência de um nível base (sim ou não). Apenas como ilustração descrevemos abaixo a incidência do problema respiratório em cada ocasião segundo os dois tratamentos.

	Visita 1	Visita 2	Visita 3	Visita 4
Tratamento	22/27	13/27	5/27	1/27
Placebo	20/29	18/29	21/29	15/29

Nota-se pela tabela acima que na primeira visita há uma incidência alta para ambos os tratamentos de pacientes em condição respiratória ruim, contudo a partir da segunda visita nota-se uma queda acentuada para os pacientes tratados com a droga ativa e pouca variação para os pacientes tratados com placebo. Portanto, há fortes indícios de que a droga reduz a chance de condição respiratória ruim. Os dados completos desse experimento estão descritos no arquivo **respiratorio.txt**.

Vamos denotar por Y_{ij} a condição (=1 ruim, =0 boa) do i-ésimo paciente na j-ésima ocasião, $i = 1, \ldots, 56$ e $j = 1, 2, 3, 4$. Como trata-se de resposta binária será assumido marginalmente que $Y_{ij} \sim Be(\pi_{ij})$ com parte sistemática dada por

$$\log\left\{\frac{\pi_{\mathrm{ij}}}{1-\pi_{\mathrm{ij}}}\right\} \alpha + \beta_1 \mathtt{Idade}_{\mathrm{i}} + \beta_2 \mathtt{Trat}_{\mathrm{i}} + \beta_3 \mathtt{Sexo}_{\mathrm{i}} + \beta_4 \mathtt{Base}_{\mathrm{i}},$$

em que $\texttt{Idade}_i$ denota a idade (em anos), $\texttt{Trat}_i$ (=0 droga ativa, =1 placebo), $\texttt{Sexo}_i$ (=0 feminino, =1 masculino) e $\texttt{Base}_i$ (=0 ausência do nível base, =1 presença do nível base) do i-ésimo paciente. Seguindo a sugestão de Myers et al.(2002, Seção 6.5) será assumida uma estrutura de correlação AR(1) para as respostas de cada paciente, ou seja, que $\text{Corr}(Y_{ij}, Y_{ij'}) = 1$ para $j = j'$ e $\text{Corr}(Y_{ij}, Y_{ij'}) = \rho^{|j-j'|}$ para $j \neq j'$. Para ajustar esse modelo no `R` deve-se usar os comandos

```
fit1.respir = gee(condicao ~ idade + trat + sexo + base,
id=paciente, family=binomial, corstr="AR-M", M=1).
```

Tabela 6.4
Estimativas dos parâmetros do modelo logístico aplicado aos dados sobre condição respiratória.

	Correlação AR(1)		Independência	
Parâmetro	Estimativa	z-robusto	Estimativa	z-robusto
α	-0,377	-0,529	-0,404	-0,563
β_1	0,043	3,380	0,048	3,683
β_2	1,001	3,066	1,070	3,254
β_3	-2,003	-2,988	-2,178	-3,207
β_4	0,492	0,586	0,498	0,585
ρ	0,275		0,00	

As estimativas dos parâmetros dos modelos com estrutura AR(1) e independente são apresentadas na Tabela 6.4. Nota-se que as estimativas não diferem muito e os resultados inferencias são os mesmos. Isso pode ser explicado pela baixa correlação entre as respostas do mesmo indivíduo, $\hat{\rho} = 0,275$.

Pelas estimativas da Tabela 6.4 pode-se concluir que o resultado da condição respiratória independe do nível base, no entanto depende da idade, do tratamento e do sexo. Por exemplo, há um aumento na chance de condição respiratória ruim com o aumento da idade, conforme esperado. A razão de chances entre sexo feminino e masculino é estimada por $\hat{\psi} = e^{2,003} = 7,41$,

ou seja, as mulheres têm aproximadamente 7,41 vezes a chance dos homens terem o problema. Pacientes que foram tratados com placebo têm $\hat{\psi} = e^{1,001} = 2,72$ vezes a chance dos pacientes que foram tratados com a droga de terem condição respiratória ruim. Em todos os cálculos acima supõe-se que as demais variáveis estão fixadas.

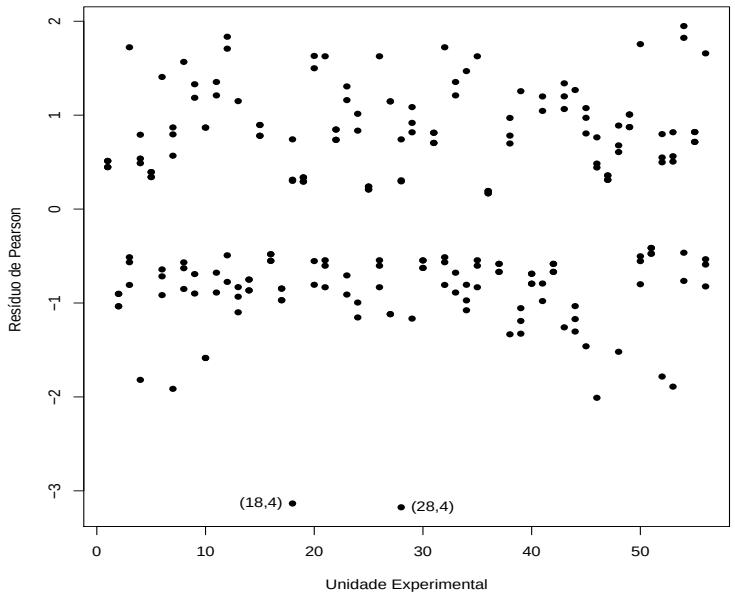

Figura 6.13: Gráfico do resíduo de Pearson referente ao modelo binomial com estrutura de correlação AR(1) ajustado aos dados sobre condição respiratória.

Na Figura 6.13 é apresentado o gráfico do resíduo de Pearson contra a ordem das observações e como podemos observar, com exceção de 2 resíduos referentes a medidas dos pacientes #18 e #28, todos os demais caem no intervalo [-2,2], indicando um bom ajuste do modelo com estrutura de correlação AR(1). O gráfico normal de probabilidades com o resíduo de Pearson

(Figura 6.14) não indica afastamentos da suposição de distribuição marginal Bernoulli com estrutura de correlação AR(1).

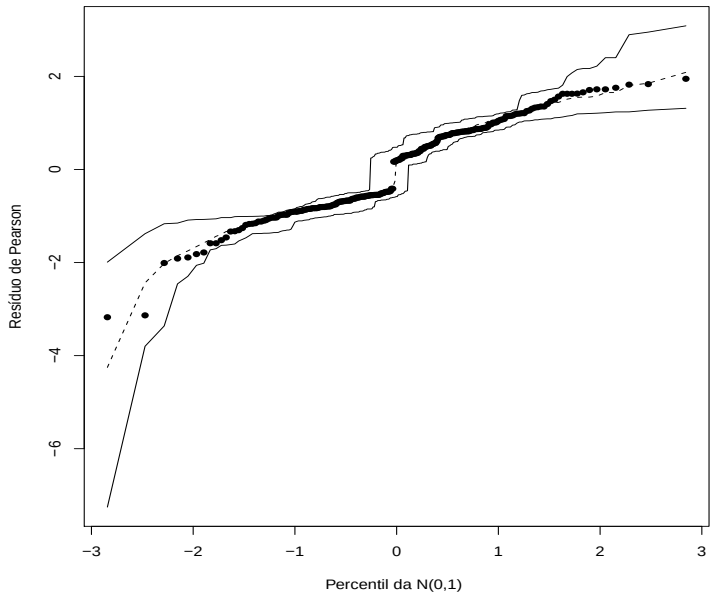

Figura 6.14: Gráfico normal de probabilidades referente ao modelo binomial com estrutura de correlação AR(1) ajustado aos dados sobre condição respiratória.

Já o gráfico da distância de Cook descrito na Figura 6.15 destaca três medidas de pacientes diferentes sendo duas dessas medidas destacadas também no gráfico com o resíduo de Pearson. Contudo, o ajuste sem considerarmos esses três pacientes não causa mudanças inferenciais.

6.5.3 Placas dentárias

Hadgu e Koch(1999) discutem os resultados de um ensaio clínico com 109 adultos voluntários com pré-existência de placa dentária. Nesse estudo os indivíduos foram distribuídos de forma aleatória para receberem um líquido tipo A (34 indivíduos), um líquido tipo B (36 indivíduos) e um líquido controle (39 indivíduos). As placas dentárias de cada indivíduo foram avaliadas e classificadas segundo um escore no início do tratamento, após 3 meses e após 6 meses. Os dados encontram-se no arquivo **rinse.txt**.

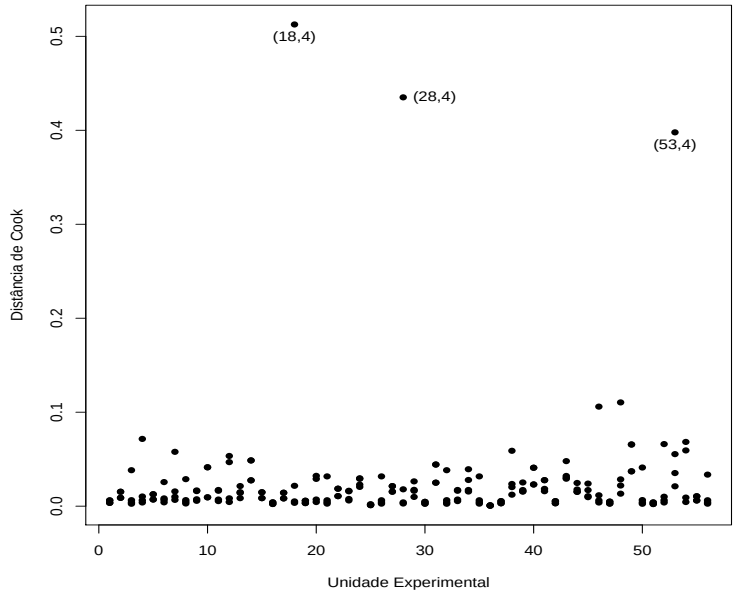

Figura 6.15: Distância de Cook referente ao modelo binomial com estrutura de correlação AR(1) ajustado aos dados sobre condição respiratória.

Tabela 6.5
Medidas resumo para os escores das placas dentárias segundo os tratamentos e períodos de escovação.

	Início	3 Meses	6 Meses
Controle	2,562 (0,343) n=39	1,786 (0,700) n=39	1,738 (0,595) n=36
Líquido A	2,568 (0,354) n=34	1,315 (0,715) n=34	1,259 (0,744) n=34
Líquido B	2,479 (0,296) n=36	1,255 (0,550) n=36	1,032 (0,451) n=36

O objetivo do estudo é verificar se pelo menos um dos novos líquidos reduz o número médio de placas dentárias. Seja Y_{ijk} o escore do k-ésimo indivíduo do i-ésimo grupo (=1 controle, =2 líquido A, =3 líquido B) e j-ésimo período (=1 início do tratamento, =2 após 3 meses, =3 após 6 meses), $k = 1, \ldots, n_{ij}$ com $n_{1j} = 39$, $n_{2j} = 34$ e $n_{3j} = 36$. Foram omitidas das nossas análises quatro observações para as quais não foi possível obter o valor do escore. Na Tabela 6.5 descrevemos os valores médios com os respectivos erros padrão para os grupos formados. Nota-se um decréscimo no valor médio após 3 meses de escovação para os três tratamentos, sendo a redução mais acentuada para os líquidos A e B. Nota-se também um aumento da variabilidade. De 3 meses para 6 meses de escovação o decréscimo continua para o escore médio dos grupos que receberam os líquidos A e B, havendo uma redução mais evidente para o grupo tratado com o líquido B. Esse grupo também apresenta as menores variabilidades. Essas tendências podem ser observadas quando são considerados os perfis individuais dos voluntários para os três tipos de líquido

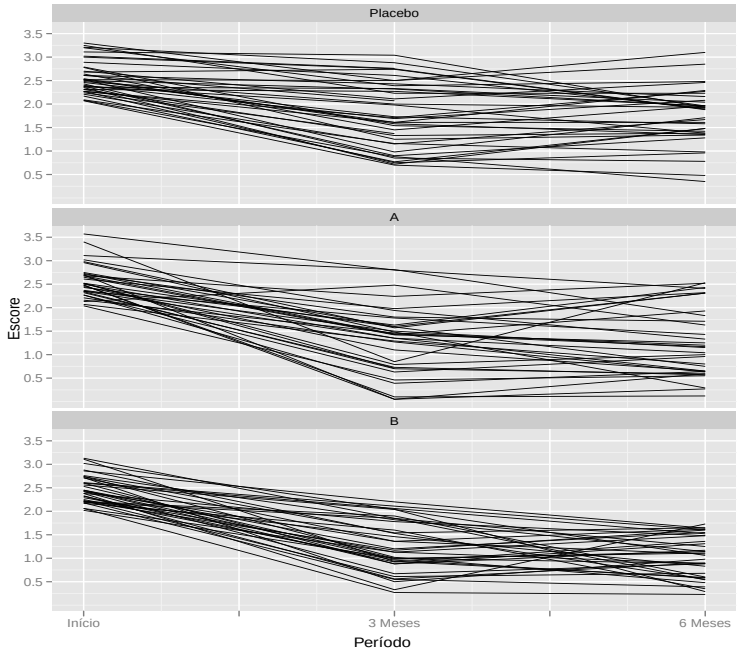

Figura 6.16: Gráfico de perfis para o escore dos voluntários que receberam placebo, líquido tipo A e líquido tipo B referente aos dados sobre placas dentárias.

ao longo do tempo conforme descrito na Figura 6.16.

Tabela 6.6
Estimativas dos parâmetros do modelo log-linear gama aplicado aos dados sobre placas dentárias.

Parâmetro	Estimativa	z-robusto	Parâmetro	Estimativa	z-robusto
α	0,941	44,407	$(\beta\gamma)_{22}$	-0,308	-3,124
β_2	0,002	0,080	$(\beta\gamma)_{32}$	-0,319	-3,835
β_3	-0,033	-1,138	$(\beta\gamma)_{23}$	-0,333	-3,266
γ_2	-0,278	-7,335	$(\beta\gamma)_{33}$	-0,492	-5,792
γ_3	-0,004	-8,321			
ρ	0,38				
ϕ^{-1}	5,68				

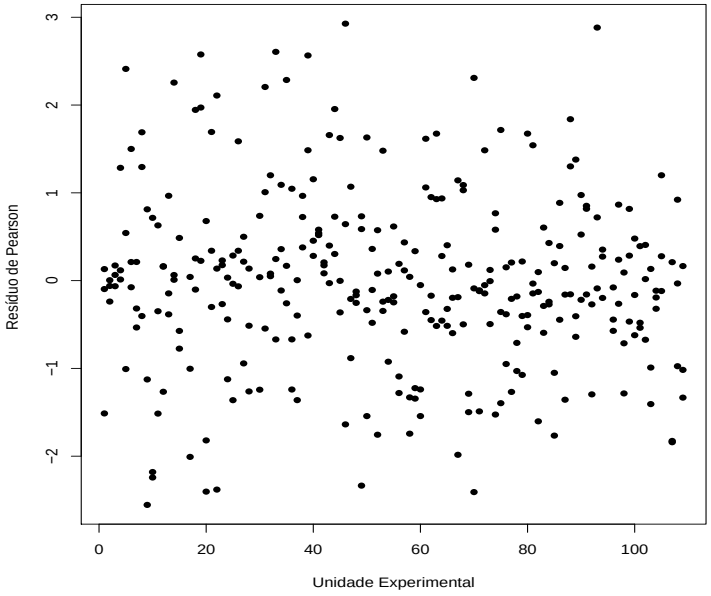

Figura 6.17: Gráfico do resíduo de Pearson referente ao modelo binomial com estrutura de correlação simétrica ajustado aos dados sobre placas dentárias.

Os pesquisadores verificaram após uma análise descritiva dos dados que a distribuição gama é mais apropriada para descrever a resposta do que a distribuição normal. Assim, vamos assumir que $Y_{ijk} \sim \mathrm{G}(\mu_{ij}, \phi)$. Seguindo ainda os pesquisadores vamos supor um modelo log-linear com interação entre tratamento e período, porém com uma parametrização um pouco diferente,

$$\log(\mu_{ij}) = \alpha + \beta_i + \gamma_j + (\beta\gamma)_{ij},$$

em que $(\beta\gamma)_{ij}$ representa a interação entre tratamento e período, sendo β_i e γ_j os efeitos principais. Teremos as restrições $\beta_1 = 0$, $\gamma_1 = 0$ e $(\beta\gamma)_{1j} = (\beta\gamma)_{i1} = 0$, para $i = 1, 2, 3$ e $j = 1, 2, 3$. As estimativas dos parâmetros são descritas na Tabela 6.6 supondo correlação simétrica entre as medidas de um mesmo indivíduo. Nota-se que a estimativa da correlação não é muito alta.

Claramente confirma-se a existência de interação entre período e tratamento. Os líquidos A e B reduzem em média a quantidade de placas dentárias, havendo indícios de uma redução mais acentuada com o líquido B de 3 meses para 6 meses de escovação.

Para ajustar esse modelo no `R` deve-se usar os comandos

```
tratm = factor(tratm)
mes = factor(mes)
fit1.placas = gee(score ~ + tratm + mes + tratm*mes,
id=voluntar, family=Gamma(link=log), corstr="exchangeable").
```

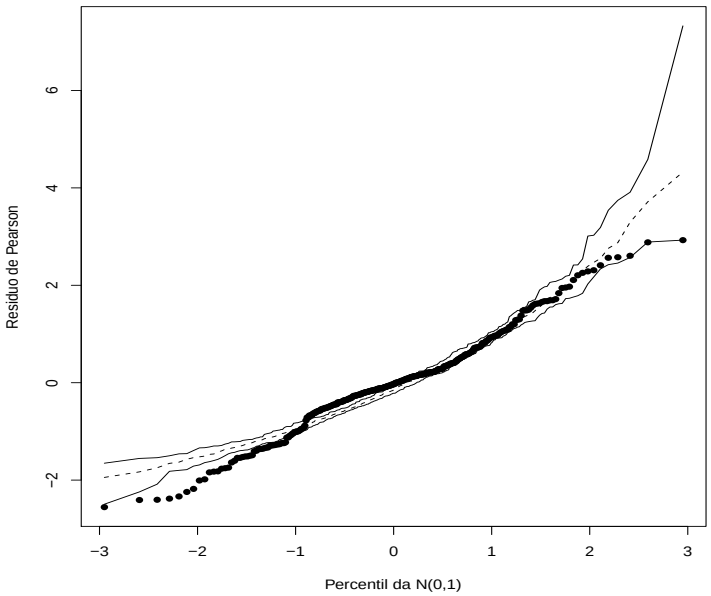

Figura 6.18: Gráfico normal de probabilidades referente ao modelo gama com estrutura de correlação simétrica ajustado aos dados sobre placas dentárias.

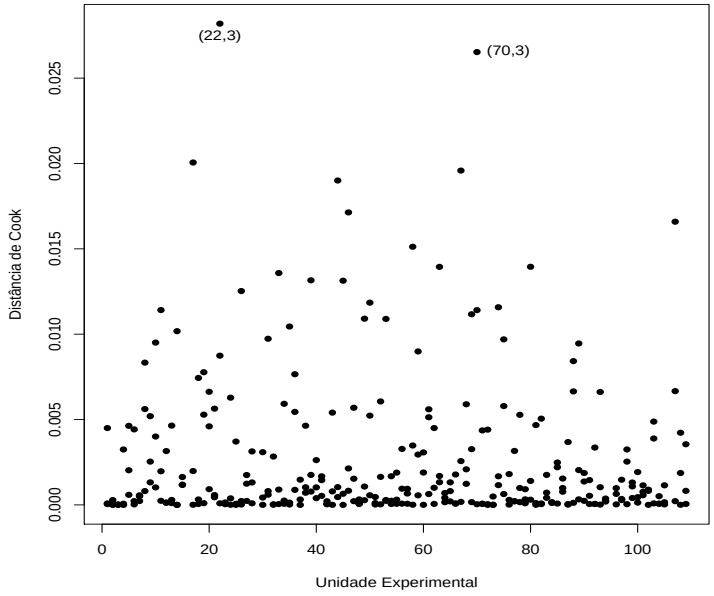

Figura 6.19: Distância de Cook referente ao modelo gama com estrutura de correlação simétrica ajustado aos dados sobre placas dentárias.

A Figura 6.17 descreve o gráfico de índices do resíduo de Pearson. Nota-se uma distriuição simétrica dos resíduos que ficam concentrados no intervalo [-3,3]. Pelo gráfico normal de probabilidades com o resíduo de Pearson (Figura 6.18) nota-se alguns afastamentos, em particular para os resíduos com valores negativos mais extremos, indicando uma falta de ajuste nesses casos. Finalmente, tem-se na Figura 6.19 o gráficos de índices da distância de Cook. Destaque para a $3^{\underline{a}}$ medida dos voluntários #22 (líquido B) e #70 (líquido B). Espera-se para ambos os voluntários um decréscimo no escore ao longo do tempo. Todavia, para o voluntário #22 tem-se a sequência 2,56; 2,04 e 0,29, ou seja, uma queda muito acentuada da $2^{\underline{a}}$ medida para a $3^{\underline{a}}$ medida. Já para o voluntário #70 tem-se a sequência 2,38; 0,33 e 1,75, ou seja, um decréscimo muito acentuado da $1^{\underline{a}}$ para a $2^{\underline{a}}$ medida, porém um aumento

após a $2^{\underline{a}}$ medida. Essas tendências que destoam do esperado para o líquido tipo B podem ter elevado o valor da distância de Cook para a $3^{\underline{a}}$ medida desses voluntários. A retirada desses dois vonluntários, contudo, altera muito pouco as estimativas e não altera os resultados inferenciais. Cardoso-Neto e Paula (2001) analisaram este exemplo supondo restrições em alguns dos parâmetros e encontraram evidências mais fortes com relação aos resultados obtidos por Hadgu e Koch(1999).

6.6 Exercícios

1. Supor as funções de variância $V(t) = t^3$ e $V(t) = t + t^2/k$ para $t > 0, k > 0$. Encontre para cada caso a função $Q(\mu; y)$ e verifique sob quais restrições as funções encontradas são proporcionais a funções de verossimilhança da família exponencial.

2. Considere a seguinte função de quase-verossimilhança:

$$Q(\mu; y) = \frac{1}{\sigma^2} \int_y^\mu \frac{y - t}{V(t)} dt,$$

   em que $V(t) = t(1 + t)$ para $t > 0$. (i) Desenvolva essa função de quase-verossimilhança. (ii) Verifique se é possível recuperar alguma distribuição da família exponencial. Em caso afirmativo qual é a distribuição? (iii) Supor agora uma amostra aleatória de n variáveis aleatórias independentes com função de quase-verossimilhança $Q(\mu_i; y_i)$ dada acima. Como fica a função quase-desvio? (iv) Como estimar σ^2?

3. Considere novamente o arquivo **claims.txt**, em que 9 variáveis são observadas para uma amostra aleatória de 996 apólices de seguros de veículos extraída do livro de de Jong e Heller (2008). A variável `expos` (exposição do veículo), que varia no intervalo (0,1), será considerada

agora como variável resposta. Inicialmente, faça uma análise descritiva dos dados e procure agrupar as variáveis categóricas em um número menor de categorias. Aplique modelos de quase-verossimilhança com funções $V(\mu) = \mu(1-\mu)$ e $V(\mu) = \mu^2(1-\mu)^2$, em que μ denota o valor esperado para a exposição do veículo, para explicar a variável resposta dadas as demais variáveis explicativas. Para o modelo selecionado faça uma análise de diagnóstico e procure interpretar os coeficientes estimados através de razões de chances.

4. Supor $Y_1, \ldots, Y_n$ variáveis aleatórias independentes com logaritmo da função de quase-verossimilhança $Q(\mu_i; y_i)$, $i = 1, \ldots, n$. Mostre que as funções escore e de informação para $\boldsymbol{\beta}$ ficam, respectivamente, dadas por:

$$\mathbf{U}_\beta = \frac{1}{\sigma^2}\mathbf{D}^T\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu})$$

e

$$\mathbf{K}_{\beta\beta} = -\mathrm{E}\left\{\frac{\partial \mathbf{U}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right\} = \frac{1}{\sigma^2}\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D}.$$

5. Sejam Y_{ij} variáveis aleatórias tais que $Y_{ij} \sim \mathrm{FE}(\mu_i, \phi)$, $i = 1, 2$ e $j = 1, \ldots, m$. A estatística de Wald para testar $H_0 : \mu_1 - \mu_2 = 0$ contra $H_1 : \mu_1 - \mu_2 \neq 0$ é dada por $\xi_W = (\bar{Y}_1 - \bar{Y}_2)^2/\mathrm{Var}(\bar{Y}_1 - \bar{Y}_2)$. Sob H_0 e para $m \to \infty$ segue que $\xi_W \sim \chi^2_1$. Calcular $\mathrm{Var}(\bar{Y}_1 - \bar{Y}_2)$ para as seguintes situações:

   (a) supondo que $\mathrm{Corr}(Y_{ij}, Y_{ij'}) = \rho$ para ($j \neq j'$; i fixo) e =0 em caso contrário;

   (b) supondo que $\mathrm{Corr}(Y_{ij}, Y_{i'j}) = \rho$ para ($i \neq i'$; j fixo) e =0 em caso contrário;

Para $\mu_1 - \mu_2$ e ϕ fixos e $\rho \geq 0$ discutir o comportamento do poder de ξ_W conforme ρ cresce para as situações (a) e (b). São esperados esses comportamentos? Comente.

6. Supor $Y_{ij} \sim \text{FE}(\mu, \phi)$, para $i = 1, \ldots, n$ e $j = 1, \ldots, r_i$, $\text{Corr}(Y_{ij}, Y_{ij'}) = \rho$ para $j \neq j'$ (=1 caso contrário), $r_i \geq 2$. Obter $\text{E}(Y_i)$ e $\text{Var}(Y_i)$, em que $Y_i = Y_{i1} + \cdots + Y_{ir_i}$. Mostre que $-1/(r_{\min} - 1) \leq \rho \leq 1$, comente. Use os resultados $\text{Var}(X + Z) = \text{Var}(X) + \text{Var}(Z) + 2\text{Cov}(X, Z)$ e $\text{Cov}(X, Z) = \rho\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Z)}$.

7. Supor $Y_i \overset{\text{ind}}{\sim} \text{Q}(\pi_i; y_i)$, em que $\text{E}(Y_i) = \pi_i$ e $\text{Var}(Y_i) = \sigma^2\pi_i(1 - \pi_i)$, para $i = 1, \ldots, n$, com parte sistemática dada por $\text{arcosen}(\sqrt{\pi_i}) = \beta_0 + \beta_1(x_i - \bar{x})$. Obtenha a matriz de variância-covariância assintótica $\text{Var}(\hat{\boldsymbol{\beta}})$, em que $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$. Desenvolva uma estatística tipo-escore para testar $\text{H}_0 : \beta_1 = 0$ contra $\text{H}_1 : \beta_1 \neq 0$? Qual a distribuição nula assintótica da estatística do teste? Resultados úteis: $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ e $\frac{d}{dx}\text{arcsen}\{u(x)\} = \frac{1}{\sqrt{1-u^2}}\frac{du}{dx}$.

8. Supor o modelo de quase-verossimilhança em que $Y_1, \ldots, Y_n$ são variáveis aleatórias independentes tais que $\text{E}(Y_i) = \mu_i$ e $\text{Var}(Y_i) = \sigma^2\mu_i^2$ com parte sistemática dada por $\log(\mu_i) = \beta_0 + \beta_1(x_i - \bar{x})$. Responda aos itens abaixo:

   (a) como ficam as variâncias assintóticas de $\hat{\beta}_0$ e de $\hat{\beta}_1$?

   (b) Como fica o teste de Wald para testar $\text{H}_0 : \beta_0 = 0$ contra $\text{H}_1 : \beta_0 \neq 0$?

   (c) Proponha um teste tipo escore para testar $\text{H}_0 : \beta_1 = 0$ contra $\text{H}_1 : \beta_1 \neq 0$.

9. Como fica a diferença entre desvios para testar $\mathrm{H}_0 : \beta_1 = 0$ contra $\mathrm{H}_1 : \beta_1 \neq 0$ num modelo de quase-verossimilhança com $V(\mu_i) = \mu_i^2(1-\mu_i)^2$, $g(\mu_i) = \eta_i = \mathbf{x}_i^T\boldsymbol{\beta}$ e $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$?

10. Supor que o vetor de respostas seja dado por $\mathbf{Y}_{ij} = (Y_{ij1}, \ldots, Y_{ijT})^T$, em que Y_{ijt} denota a resposta para o j-ésimo elemento do i-ésimo grupo no instante t, $i = 1, \ldots, g$ e $j = 1, \ldots, r_i$ (Park et al., 1998). Supor ainda que $\mathrm{E}(Y_{ijt}) = \mu_i$, $\mathrm{Var}(Y_{ijt}) = V_i\phi^{-1}$ e que Y_{ijt} pertence à família exponencial. Mostre que dado $\hat{\boldsymbol{\rho}}$ a equação de estimação generalizada para μ_i pode ser expressa na forma $\mathbf{S}(\hat{\mu}_i) = 0$, em que

$$\mathbf{S}(\mu_i) = \sum_{j=1}^{r_i} \mathbf{1}_T^T \mathbf{R}_{ij}(\boldsymbol{\rho})(\mathbf{y}_{ij} - \mu_i \mathbf{1}_T),$$

$\mathbf{R}_{ij}$ é a matriz trabalho para o j-ésimo indivíduo do i-ésimo grupo e $\mathbf{1}_T$ é um vetor $T \times 1$ de uns. Expresse a estimativa de μ_i em forma fechada.

11. Supor que $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ir_i})^T$, $i = 1, \ldots, n$, são vetores aleatórios independentes tais que $Y_{ij} \sim \mathrm{Be}(\pi_i)$. Assumir ainda que a matriz trabalho para $\mathbf{Y}_i$ é permutável e que

$$\log\left\{\frac{\pi_i}{1-\pi_i}\right\} = \mathbf{x}_i^T\boldsymbol{\beta}.$$

Mostre que, dado $\hat{\rho}$, as EEGs para $\boldsymbol{\beta}$ ficam dadas por

$$\mathbf{S}_\beta(\hat{\boldsymbol{\beta}}_G) = \sum_{i=1}^{n} \{1 + (r_i - 1)\hat{\rho}\}^{-1} \mathbf{x}_i(y_i - n_i\hat{\pi}_i) = \mathbf{0},$$

em que $y_i = y_{i1} + \cdots + y_{ir_i}$. Sugestão: use a relação abaixo

$$\mathbf{R}_i^{-1}(\rho) = (1-\rho)^{-1}[\mathbf{I}_{r_i} - \rho\{1 + (r_i - 1)\rho\}^{-1}\mathbf{J}],$$

em que $\mathbf{J}$ é uma matriz $r_i \times r_i$ de uns. Como fica o processo iterativo para estimar $\boldsymbol{\beta}$?

12. Supor que $Y_{ij} \sim \text{Be}(\mu)$ para $i = 1, \ldots, n$ e $j = 1, \ldots, r_i$, em que $\text{Corr}(Y_{ij}, Y_{ij'}) = \rho$ (fixado) para $j \neq j'$ com parte sistemática dada por $\log\left\{\frac{\mu}{1-\mu}\right\} = \beta$. Responda às seguintes questões: (i) como fica a equação de estimação generalizada para estimar β? (ii) expresse em forma fechada a estimativa $\hat{\beta}_G$ (obtenha inicialmente $\hat{\mu}_G$) e (iii) como fica a variância assintótica (não robusta) de $\hat{\beta}_G$?

13. Supor que $Y_{ij} \sim \text{Q}(\mu, \sigma^2)$ para $i = 1, \ldots, n$ e $j = 1, 2$, em que $\text{Var}(Y_{ij}) = \sigma^2\mu^2$, $\text{Corr}(Y_{ij}, Y_{ij'}) = \rho$ para $j \neq j'$ com parte sistemática dada por $\log\mu = \beta$. Responda às seguintes questões: (i) como fica a equação de estimação generalizada para estimar β? (ii) expresse em forma fechada a estimativa $\hat{\beta}_G$ (obtenha inicialmente $\hat{\mu}_G$) e (iii) como fica a variância assintótica (não robusta) de $\hat{\beta}_G$? Supor que ρ e σ^2 são estimados consistentemente.

14. Considere uma amostra aleatória de n indivíduos que são observados em 2 ocasiões cada um, sendo Y_{ij} a resposta do i-ésimo indivíduo na j-ésima ocasião para $i = 1, \ldots, n$ e $j = 1, 2$, com a suposição $Y_{i1} \overset{\text{ind}}{\sim} \text{FE}(\mu_1, \phi)$ e $Y_{i2} \overset{\text{ind}}{\sim} \text{FE}(\mu_2, \phi)$ e $\rho = \text{Corr}(Y_{i1}, Y_{i2})$ ou seja $\text{Cov}(Y_{i1}, Y_{i2}) = \rho\sqrt{\text{Var}(Y_{i1})}\sqrt{\text{Var}(Y_{i1})}$. A diferença entre as médias amostrais nas duas ocasiões $\bar{Y}_2 - \bar{Y}_1$, em que $\bar{Y}_j = n^{-1}\sum_{i=1}^{n} Y_{ij}$ para $j = 1, 2$, é utilizada para detectar eventuais diferenças entre as médias μ_2 e μ_1. Responda às seguintes questões:

    (i) calcule $\text{Var}(\bar{Y}_2 - \bar{Y}_1)$,

    (ii) chame $\Delta = \mu_2 - \mu_1$ e calcule $\text{P}(\Delta - \epsilon < \bar{Y}_2 - \bar{Y}_1 < \Delta + \epsilon) = 1 - \alpha$, $0 < \alpha < 1$ e $\epsilon > 0$, em que $1 - \alpha = \text{P}(-z < Z < z)$, $Z \sim \text{N}(0, 1)$,

    (iii) expresse n em função das quantidades z, ϵ, Δ e ρ e

(iv) discuta o comportamento de n em função de ρ mantendo-se as demais quantidades fixas.

Supor para n grande $\bar{Y}_2 - \bar{Y}_1 \sim \mathrm{N}(\Delta, \mathrm{Var}(\bar{Y}_1 - \bar{Y}_2))$.

15. Um experimento é conduzido para avaliar a dispersão de um pigmento particular numa pintura. Quatro diferentes misturas do pigmento são estudadas. O procedimento consiste em preparar cada mistura e aplicá-la num painel usando três métodos diferentes: pincel, rolo e spray. O experimento é repetido três dias diferentes e a resposta é a porcentagem de reflectância do pigmento. Os dados são descritos na tabela abaixo e no arquivo **mistura.txt** (Myers et al., 2002).

		Mistura			
Dia	Método	1	2	3	4
	1	64,5	66,3	74,1	66,5
1	2	68,3	69,5	73,8	70,0
	3	70,3	73,1	78,0	72,3
	1	65,2	65,0	73,8	64,8
2	2	69,2	70,3	74,5	68,3
	3	71,2	72,8	79,1	71,5
	1	66,2	66,5	72,3	67,7
3	2	69,0	69,0	75,4	68,6
	3	70,8	74,2	80,1	72,4

Analise os dados através de equações de estimação generalizadas com estrutura de correlação simétrica. Faça análise de diagnóstico.

16. No arquivo **ratosgee.txt** (Myers et al., 2002, Seção 6.5) estão os dados de um experimento em que 30 ratos tiveram uma condição de leucemia induzida. Três drogas quimio-terápicas foram utilizadas no tratamento

dos animais. Foram coletadas de cada animal a quantidade de células brancas (WBC), a quantidade de células vermelhas (RBV) e o número de colônias de células cancerosas (RESP) em quatro períodos diferentes. Assuma distribuição de Poisson para RESP em cada período e verifique através de um modelo log-linear se existe diferenças significativas entre os três tratamentos considerando WBC e RBC como variáveis explicativas. Compare os resultados supondo estruturas de correlação independente e AR(1). Faça uma análise de diagnóstico.

17. Sejam $Y_{i1} \overset{\text{iid}}{\sim} \text{FE}(\mu_1, \phi)$ e $Y_{i2} \overset{\text{iid}}{\sim} \text{FE}(\mu_2, \phi)$, em que $\text{Corr}(Y_{i1}, Y_{i2}) = \rho$, para $i = 1, \ldots, n$. Para testar $\text{H}_0 : \mu_1 - \mu_2 = 0$ contra $\text{H}_1 : \mu_1 - \mu_2 \neq 0$ considere a estatística

$$\xi_W = \frac{(\bar{Y}_1 - \bar{Y}_2)^2}{\text{Var}(\bar{Y}_1 - \bar{Y}_2)},$$

que sob H_0 segue assintoticamente distribuição χ^2_1 central. Sob a hipótese alternativa ξ_W segue assintoticamente distribuição $\chi^2_1(\lambda)$, em que $\lambda = (\mu_1 - \mu_2)^2/2\text{Var}(\bar{Y}_1 - \bar{Y}_2)$ é o parâmetro de não centralidade. Seja $P_n(\lambda, \rho) = P(\xi_W > c|\text{H}_1)$ o poder do teste (probabilidade de rejeitar H_0 quando há diferença entre as médias populacionais). Calcule inicialmente $\text{Var}(\bar{Y}_1 - \bar{Y}_2)$ e discuta o comportamento do poder do teste à medida que varia o coeficiente de correlação linear de Pearson $-1 \leq \rho \leq 1$, supondo fixados μ_1, μ_2, ϕ e n. Procure responder para quais situações será necessário um tamanho amostral maior para detectar a mesma diferença (entre as médias populacionais) com a mesma probabilidade.

18. No arquivo **artrite.txt** (Myers et al., 2002, Seção 6.5) estão os dados de um ensaio clínico em que 20 pacientes com artrite foram aleatorizados de modo que 10 receberam o medicamento `auronofin` e os outros 10 receberam `placebo`. Foram observadas as variáveis explicativas `gênero`

(1: masculino, 0: feminino) e a `idade` do paciente em anos além do `tratamento` (0: placebo, 1: auronofin). Os pacientes foram consultados em 4 ocasiões (1: início, 2: 1 mês, 3: 2 meses e 4: 3 meses) a respeito do seu estado avaliado pelo próprio paciente (1: ruim, 2: regular, 3: bom). Faça inicialmente uma análise descritiva com os dados.

Seja Y_{ij} o estado do i-ésimo paciente na j-ésima ocasião (=1 bom, =0 regular ou ruim) para $i = 1, \ldots, 20$ e $j = 1, 2, 3, 4$. Assuma que $Y_{ij} \sim \text{Be}(\pi_{ij})$, em que π_{ij} é a probabilidade do estado ser considerado bom pelo i-ésimo paciente na j-ésima ocasião. Proponha uma EEG para explicar π_{ij} através de uma regressão logística e considerando as estruturas de correlação simétrica e AR(1) entre as ocasiões de um mesmo paciente. Considere no modelo apenas os efeitos principais `tratamento`, `idade`, `gênero` e `ocasião`. Compare os modelos através de métodos de diagnóstico e para o modelo escolhido faça uma interpretação através de razões de chances.

19. No arquivo **Milk** do GAMLSS são apresentados dados referentes a um experimento longitudinal desenvolvido na Austrália com 79 vacas que foram aleatorizadas segundo três dietas e foi observado semanalmente a quantidade de proteína no leite de cada animal. O objetivo principal do estudo é verificar se há diferenças significativas entre as quantidades médias semanais de proteína sob as três dietas. Os dados estão descritos na seguinte ordem: (i) `protein` (quantidade de proteína), (ii) `Time` (semana), (iii) `Cow` (identificação do animal) e (iv) `Diet` (cevada, cevada+tremoços e tremoços). É preciso informar que a variável `Diet` é categórica através do comando

    ```
    Diet=factor(Diet).
    ```

Fazer inicialmente uma análise descritva dos dados, por exemplo, apresentando os perfis dos animais segundo a quantidade de proteína observada ao longo das semanas e para cada dieta gráficos de densidade e boxplots. Ajustar inicialmente uma equação de estimação generalizada gama com estrutura de correlação do tipo AR(1) e considere o tempo como variável explicativa contínua. Verifique se é possível incluir interação entre `Diet` e `Time`. Faça uma análise de diagnóstico e interprete os resultados do modelo selecionado.

20. No arquivo **gross.txt** (Munnell, 1990) estão resumidos os dados de produtividade dos 48 estados norte-americanos contíguos no período de 1970 a 1986. As variáves estão descritas na seguinte ordem (os recursos estão expressos em milhões de USD): (i) `state`, nome do estado, (ii) `region`, região do estado, (iii) `yr`, ano, (iv) `pcap`, total do capital de empresas públicas, (v) `hwy`, capital das estradas e rodovias, (vi) `water`, capital das empresas de saneamento básico, (vii) `util`, capital das demais empresas públicas, (viii) `pc`, total do capital privado, (ix) `gsp`, produto interno bruto, (x) `emp`, total de empregos e (xi) `unemp`, taxa de desemprego. O objetivo do estudo é tentar relacionar o produto interno bruto de cada estado com as demais variáveis. Faça uma análise descritiva considerando apenas as variáveis, `gsp`, `water` e `yr`.

    Supor inicialmente o seguinte modelo de quase-verossimilhança:

    (i) $Y_{ij} \sim \text{Q}(\mu_{ij}; y_{ij})$, $\text{E}(Y_{ij}) = \mu_{ij}$ e $\text{Var}(Y_{ij}) = \sigma^2\mu_{ij}^2$

    (ii) $\log(\mu_{ij}) = \beta_0 + \beta_1 \log(\text{water})_{ij} + \beta_2 \text{yr}_{ij}$

    (iii) $\text{Corr}(\mathbf{Y}_{ij}) = \mathbf{R}_{ij}(\alpha)$,

    em que $\sigma^2 > 0$, $\mathbf{Y}_{ij} = (Y_{ij1}, \ldots, Y_{ij17})^\top$. Compare o modelo acima supondo as estruturas de correlação simétrica e AR(1). Para a estrutura

selecionada tente melhorar o modelo, por exemplo, incluindo interação. Faça uma análise de diagnóstico do modelo final e interpetre os resultados.

Apêndice A

Neste apêndice são descritos os conjuntos de dados usados nos exemplos e nos exercícios propostos. As variáveis são descritas na ordem em que aparecem em cada arquivo.

Capítulo 1

bateria.txt: tempo (em minutos) e queda da tensão (voltagem).

capm.txt: taxa de retorno Tbill, retorno Microsoft, retorno SP500, retorno GE e retorno Ford.

cimento.txt: calor, x_1, x_2, x_3 e x_4.

delivery.txt: tempo (em minutos), número de caixas e distância (em pés).

fluxo.txt : idade (em anos), gênero, interj (interjeiçôes por minuto), fpm (frequência de sílabas por minuto) e fsm (frequência de sílabas por minuto).

ginidh.txt: UF, idh de 2017 e gini de 2013.

octana.txt: x1, x2, x3, x4, e octanas. A resposta é o número de octanas.

reg3.txt: nome do estado, população estimada em julho de 75, renda per capita em 74 (em USD), proporção de analfabetos em 70, expectativa

de vida 69-70, taxa de criminalidade em 76 (por 100000 habitantes), proporção de estudantes que concluíram o segundo grau em 70, número de dias do ano com temperatura abaixo de zero graus Celsus e área do estado (em milhas quadradas).

salarios.txt: salário anual (em mil USD), sexo, posição na empresa (escore de 1 a 9) e experiência (em anos).

sheep.txt: peso (em kg) e energi (mcal).

takeoff.txt: id, tempo (em minutos) e cia aérea.

vendas.txt: telhados (em mil m^2), gastos (em mil USD), clientes (em milhares), marcas e potencial (escore).

wine.txt: claridade, aroma, corpo, sabor, aromac (aroma do tonel de carvalho), qualidade e região (região1, região 2, região 3, região 4).

Capítulo 2

canc3.txt: tipo de tumor (0:benigno, 1:maligno), idade (em anos), sexo (1:masculino, 2:feminino), HL e FF (1:ausente, 2:discreta, 3:moderada, 4:intensa).

censo.txt: unidade da federação, escolaridade média (anos de estudo) e renda média (em reais).

reg2.txt: sigla do estado, taxa do combustível (em USD), porcentagem de motoristas licenciados, renda per capita (em USD), ajuda federal às estradas do estado (em mil USD) e consumo per capita de combustível (em galões por ano).

perch.txt: Weight (peso do peixe, em gramas), Length1 (comprimento do peixe do nariz até o início da cauda, em cm), Length2 (comprimento do peixe do nariz até o entalhe da cauda, em cm), Length3 (comprimento do peixe do nariz até o final da cauda, em cm), Heightpc (altura máxima, como porcentagem de Length3) e Widthpc (largura máxima, como porcentagem de `Length3`).

trees.txt: diâmetro (em polegadas), altura (em pés) e volume da árvore (em pés cúbicos).

Capítulo 3

claims.txt: valor do veículo (em 10000 dolares australianos), exposição do veículo, número de sinistros no período, custo total dos sinistros (em dolares australianos), tipo do veículo (em 11 categorias), idade do veículo (em 4 categorias), sexo do condutor principal, área de residência do condutor principal (em 6 categorias) e idade do condutor principal (em 6 categorias).

dfilme.txt: tempo de duração do filme (em horas) e densidade máxima do filme.

energy.txt: total de energia consumida num mês (em kilowatts-hora) e demanda de energia na hora de pico.

insurance.txt: valor pago do seguro (dolares australianos), representação legal (0:não, 1:sim), mês em que ocorreu o acidente e tempo operacional.

milho.txt: quantidade de nitrogênio, quantidade de fosfato e produtividade de milho (libras/acre).

pesca.txt: frota (Santos e Ubatuba), ano (95 a 99), trimestre (1 a 4), latitude (de 23,25º a 28,25º), longitude (de 41,25º a 50,75º), dias de pesca, captura (quantidade em kg de peixes capturados) e cpue (captura por unidade de esforço).

raia.txt : período (seco e chuvoso), local (local da pesca), mare (maré, quadratura e sizígia), vvento (velocidade do vento m/s), tmax (temperatura máxima em graus Celsus), tmin (temperatura mínima em graus Celsus), ins (insolação em horas) e cpue (captura por unidade de esforço).

restaurante.txt: faturamento anual (em mil USD) e gastos com publicidade (em mil USD).

snacks.txt: força necessária para o cisalhamento, tipo de snack (1:A, 2:B, 3:C, 4:D, 5:E), número de semanas.

sobrev.txt: número de células brancas, tempo de sobrevivência (em semanas) e característica morfológica (AG=1 positivo, AG=0 negativo).

turbina.txt: tipo de turbina (1 a 5) e tempo de duração do motor (em milhões de ciclos).

vidros.txt: tempo de resistência (em horas), voltagem (1:200, 2:250, 3:300, 4:350) e temperatura (1:170 graus Celsus, 2:180 graus Celsus).

Capítulo 4

besouros.txt: besouros mortos, besouros expostos e dose.

caduquice.txt: escore no exame psicológico, ocorrência de caduquice (1:sim, 0:não).

camundongos: sexo (1:macho, 0:fêmea), tratamento (1:sim, 0:controle), casos e expostos.

canc3a.txt: Tipo (0:benigno, 1:maligno), Idade, Sexo (0:masculino, 1:feminino), HL (0:baixa, 1:alta) e FF (0:baixa, 1:alta).

creme.txt: centro (1,2,3,4), tratamento (droga, controle), sucessos, fracassos e total.

dengue.txt: idade (em anos) do entrevistado, nível sócio-econômico (1:alto, 2:médio, 3:baixo), setor da cidade onde mora o entrevistado (1:setor 1, 2:setor 2) e diagnóstico da doença (1:sim, 0:não).

diabetes.txt: massa corporal, histórico familiar (1:presença, 0:ausência) e atividades físicas (1:presença, 0:ausência) para os casos e para os controles, respectivamente.

dose1.txt: dose, caramujos expostos e caramujos mortos.

dose2.txt: dose, caramujos expostos e caramujos mortos.

dose3.txt: dose, caramujos expostos e caramujos mortos.

equipamentos.txt: tempo, número de equipamentos expostos, número de equipamentos que falaharam.

grahani.txt: número de lagartos da espécie grahani, total de lagartos, período do dia (1:manhã, 2:meio-dia, 3:tarde), comprimento da madeira (1:curta, 2:cumprida), largura da madeira (1:estreita, 2:larga) e local de ocupação (1:claro, 2:escuro).

hearthd: Age (idade em anos), FE 9faixa etária) e HD (doença arterial coronariana, 1:presença e 0:ausência).

insetic.txt: número de insetos mortos, número de insetos expostos, dose do inseticida, inseticida DDT, inseticida γ-DDT e inseticida DDT + γ-DDT (1:presença, 0:ausência).

leuce.txt: idade do paciente (em anos), mancha diferencial da doença, infiltração na medula, células com leucemia, malignidade da doença, temperatura máxima antes do tratamento, tratamento (1:satisfatório, 0:não), tempo de sobrevivência (em meses) e situação (1:sobrevivente, 0:não sobrevivente).

matched.txt: estrato, observação (1:caso, 2:controle), idade da paciente no momento da entrevista (em anos), diagnóstico (1:caso, 0:controle), tempo de escolaridade (em anos), grau de escolaridade (0:nenhum, 1:segundo grau, 2:técnico, 3:universitário, 4:mestrado, 5:doutorado), checkup regular (1:sim, 2:não), idade da primeira gravidez, idade do início da menstruação, número de abortos, número de filhos, peso (em libras), idade do último período menstrual e estado civil (1:casada, 2: divorciada, 3:separada, 4:viúva, 5:solteira). Observações perdidas são denotadas por NA.

meninas.txt: garotas menstruando, garotas entrevistadas e idade média.

morgan.txt: concentração (R, D, M), dose, insetos expostos, insetos mortos.

olhos.txt: cor dos olhos dos pais, cor dos olhos dos avós, número total de filhos e número de filhos com olhos claros.

prefauto.txt: preferência comprador tipo de automóvel (1:americano, 0:japonês), idade do comprador (em anos), sexo do comprador (0:masculino, 1:feminino) e estado civil do comprador (0:casado, 1:solteiro).

pregibon.txt: resposta (1:ocorrência, 0:ausência), volume e razão.

pulso.txt: pulsação em repouso (1:normal, 0:alta), hábito de fumar (1:sim, 2:não) e peso (em kg).

rotifers.txt: densidade, rotifers suspensos, rotifers expostos e espécie (1: Polyarthra, 0:Keratella).

sementes.txt: temperatura da germinação, nível da umidade, nível da temperatura, número de sementes que germinaram.

Capítulo 5

breslow.txt: número de casos de câncer, total de pessoas-anos, número de cigarros por dia (1:não fumante, 2:1-9 cigarros, 3:10-30 cigarros, 4:+ 30 cigarros) e faixa-etária (1:40-49 anos, 2:50-59 anos, 3:60-69 anos, 4:70-80 anos).

nasal.txt: idade no primeiro emprego com 4 níveis (1:¡20, 2:20-27, 3:27.5-34.9, 4:35+ anos), ano do primeiro emprego com 4 níveis (1:¡1910, 2:1910-1914, 3:1915-1919, 4:1920-1924), tempo decorrido desde o primeiro emprego com 5 níveis (1:0-19, 2:20-29, 3:30-39, 4:40-49, 5:50+ anos), número de casos de câncer e o total de pessoas-anos de observação.

detergente.txt: temperatura da água, uso de M, preferência (X,M), maciez da água, número de pessoas.

emprego.txt: nível de renda (1: ¡ USD 6000, 2: USD 6000-15000, 3: USD 15000-25000, 4: ¿ USD 25000), grau de satisfação (1:alto, 2: bom, 3: médio, 4: baixo) e número de indivíduos.

geriatra.txt: número de quedas no período, intervenção (0:educação somente, 1:educação e exercícios físicos), sexo (0:feminino, 1:masculino), balanço e força.

gestantesc.txt: idadec (0:<30, 1:30+), cigarrosc (0:<5, 1:5+), gestaçaoc (0:<260 dias, 1:260+ dias), sobrevc (0:não, 1:sim) e freq.

heart.txt: doença das coronárias (1:sim, 2:não), nível de colesterol (1:menor do que 200 mg/100 cc, 2:200-219, 3:220-259, 4:260 ou +),pressão arterial (1:menor do que 127 mm Hg, 2:127-146, 3:147-166, 4:167 ou +) e número de indivíduos.

navios.txt: tipo do navio (1:A, 2:B, 3:C, 4:D, 5:E), ano da fabricação (1:60-64, 2:65-69, 3:70-74, 4:75-79), período de operação (1:60-74, 2:75-79), tempo de operação (em meses) e número de avarias.

nitrofen: dosagem de nitrofen, total de ovos eclodidos.

quine.txt: etnia (A:aborígine, N:não aborígine), sexo (M:masculino, F: feminino), ano (F0:8^{a} série, F1:1^{o} ano ensino médio, F2:2^{o} ano ensino médio, F3:3^{o} ano ensino médio), desempenho (SL:baixo, AL:normal) e dias ausentes no ano letivo.

recrutas.txt: hábito de nadar (ocasional, frequente), local onde costuma nadar (piscina, praia), faixa-etária (15-19, 20-25, 25-29), sexo (masculino, feminino) e número de infecções de ouvido.

rolos.txt: comprimento do tecido (em metros) e número de falhas.

store.txt: número de clientes, número de domicílios, renda média anual (em USD), idade média dos domicílios (em anos), distância entre a área e o

competidor mais próximo (em milhas) e distância entre a área e a loja (em milhas).

tvcabo.txt: número de domicílios na área (em milhares), porcentagem de domicílios com TV a cabo, renda per capita (em USD) por domicílio com TV a cabo, taxa de instalação de TV a cabo (em USD), custo médio mensal de manutenção de TV a cabo (em USD), número de canais a cabo disponíveis na área e número de canais não pagos com sinal de boa qualidade disponíveis na área.

visitas.txt: nvis, hosp, altacond, baixacond, nucron, genero, escol e seguro.

Capítulo 6

artrite.txt: paciente, ocasião (1:início, 2:1 mês, 3:2 meses, 4:3 meses), gênero (1:masculino, 0:feminino), idade (em anos), tratamento (0:placebo, 1:auronofin), resultado (1:ruim, 2:regular, 3:bom).

ataques.txt: indivíduo, período (1:antes do tratamento, 2:1º período após o tratamento, 3:2º período após o tratamento, 4:3º período após o tratamento), número de semanas em cada período, número de ataques em cada período e tratamento (0:placebo, 1:progabide).

cevada.txt: incidência da mancha (proporção), local (1 a 9) e variedade (1 a 10).

gross.txt: state, region, yr, pcap, hwy, water, util, pc, gsp, emp e unemp.

mosca.txt: número de ácaros coletados espécie2, espécie3, espécie6, espécie14, número de partes da placa, posição (1:lateral, 0:central), região (1:São

Roque, 2:Pindamonhangaba, 3:Nova Odessa, 4:Ribeirão Preto) e temperatura (em graus Celsus).

mistura.txt: painel, dia, método, mistura, porcentagem de reflectância do pigmento.

ratosgee.txt: animal, período, quantidade de células brancas, quantidade de células vermelhas e número de colônias de células cancerosas.

respiratorio.txt: paciente, tratamento (0:droga ativa, 1:placebo), sexo (0:feminino, 1:masculino), idade (em anos), nível base (0:ausência, 1:presença) e condição do paciente nas visitas (0:boa, 1:ruim).

rinse.txt: voluntário, período (1:início, 2:após 3 meses, 3:após 6 meses), tratamento (1:placebo, 2:rinse A, 3:rinse B) e escore.

Apêndice B

Neste apêndice são apresentados os códigos em R dos programas de envelope usados para alguns MLGs.

Modelos com resposta normal

```
X = model.matrix(fit.model)
   n = nrow(X)
   p = ncol(X)
   H = X%*%solve(t(X)%*%X)%*%t(X)
   h = diag(H)
   si = lm.influence(fit.model)$sigma
   r = resid(fit.model)
   tsi = r/(si*sqrt(1-h))
   #
   ident = diag(n)
   epsilon = matrix(0,n,100)
   e = matrix(0,n,100)
   e1 = numeric(n)
   e2 = numeric(n)
   #
   for ( i in 1:100) {
```

```
   epsilon[,i] = rnorm(n,0,1)
   e[,i] = (ident - H)%*%epsilon[,i]
   u = diag(ident - H)
   e[,i] = e[,i]/sqrt(u)
   e[,i] = sort(e[,i]) }
   #
   for ( i in 1:n) {
   eo = sort(e[i,])
   e1[i] = eo[5]
   e2[i] = eo[95] }
   #
   med = apply(e,1,mean)
   faixa = range(tsi,e1,e2)
   par(pty="s")
   qqnorm(tsi, xlab="Quantil da N(0,1)", ylab = "Resíduo Studentizado",
ylim=faixa)
   par(new=TRUE)
   qqnorm(e1,axes=FALSE, xlab=, ylab= , type="l", ylim=faixa, lty=1)
   par(new=TRUE)
   qqnorm(e2,axes=FALSE,xlab=, ylab=, type="l", ylim=faixa, lty=1)
   par(new=TRUE)
   qqnorm(med,axes=FALSE,xlab=, ylab=, type="l", ylim=faixa, lty=2)
```

Modelos com resposta gama

```
X = model.matrix(fit.model)
   n = nrow(X)
   p = ncol(X)
```

```
w = fit.model$weights
W = diag(w)
H = solve(t(X)%*%W%*%X)
H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
h = diag(H)
ro = resid(fit.model,type="response")
fi = (n-p)/sum((ro/(fitted(fit.model)))^ 2)
td = resid(fit.model,type="deviance")*sqrt(fi/(1-h))
#
e = matrix(0,n,100)
for (i in 1:100) {
resp = rgamma(n,fi)
resp = (fitted(fit.model)/fi)*resp
fit = glm(resp ~ X, family=Gamma)
w = fit$weights
W = diag(w)
H = solve(t(X)%*%W%*%X)
H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
h = diag(H)
ro = resid(fit, type="response")
phi = (n-p)/sum((ro/(fitted(fit)))^ 2)
e[,i] = sort(resid(fit, type="deviance")*sqrt(phi/(1-h))) }
#
e1 = numeric(n)
e2 = numeric(n)
#
for (i in 1:n) {
```

```
   eo = sort(e[i,])
   e1[i] = eo[5]
   e2[i] = eo[95]
   #
   med = apply(e,1,mean)
   faixa = range(td,e1,e2)
   #
   par(pty="s")
   qqnorm(td, xlab="Quantil da N(0,1)", ylab="Componente do Desvio
Padronizado", ylim=faixa)
   par(new=TRUE)
   qqnorm(e1,axes=FALSE,xlab=, ylab=, type="1", ylim=faixa,lty=1)
   par(new=TRUE)
   qqnorm(e2,axes=FALSE, xlab=, ylab=, type="1", ylim=faixa, lty=1)
   par(new=TRUE)
   qqnorm(med,axes=FALSE, xlab=, ylab=, type="1", ylim=faixa, lty=2)
```

Modelos com resposta binomial

```
X = model.matrix(fit.model)
   n = nrow(X)
   p = ncol(X)
   w = fit.model$weights
   W = diag(w)
   H = solve(t(X)%*%W%*%X)
   H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
   h = diag(H)
   td = resid(fit.model,type="deviance")/sqrt(1-h)
```

```
#
e = matrix(0,n,100)
for(i in 1:100){
dif = runif(n) - fitted(fit.model)
dif[ dif >=0 ] = 0
dif[dif < 0] = 1
nresp = dif
fit = glm(nresp ~ X, family=binomial)
w = fit$weights
W = diag(w)
H = solve(t(X)%*%W%*%X)
H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
h = diag(H)
e[,i] = sort(resid(fit, type="deviance")/sqrt(1-h)) }
#
e1 = numeric(n)
e2 = numeric(n)
#
for (i in 1:n) {
eo = sort(e[i,])
e1[i] = eo[5]
e2[i] = eo[95] }
#
med = apply(e,1,mean)
faixa = range(td,e1,e2)
#
par(pty="s")
```

```
   qqnorm(td, xlab="Quantil da N(0,1)", ylab="Componente do Desvio
Padronizado", ylim=faixa)
   par(new=TRUE)
   qqnorm(e1,axes=FALSE, xlab=, ylab=, type="l", ylim=faixa, lty=1)
   par(new=TRUE)
   qqnorm(e2,axes=FALSE, xlab=, ylab=, type="l", ylim=faixa, lty=1)
   par(new=TRUE)
   qqnorm(med,axes=FALSE, xlab=, ylab=, type="l", ylim=faixa, lty=2)
```

Modelos com resposta binomial com réplicas

```
X = model.matrix(fit.model)
   k = nrow(X)
   e = matrix(0,k,100)
   tot = numeric(k)
   w = fit.model$weights
   W = diag(w)
   H = solve(t(X)%*%W%*%X)
   H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
   h = diag(H)
   td = sort(resid(fit.model, type="deviance")/sqrt(1-h))
   #
   for(i in 1:100){
   for(j in 1:k) {
   dif = runif(n[j]) - fitted(fit.model)[j]
   dif[dif >= 0] = 0
   dif[dif<0] = 1
   tot[j] = sum(dif)}
```

```
   xmat = cbind(tot,n-tot)
   fit = glm(xmat ~ X, family=binomial)
   w = fit$weights
   W = diag(w)
   H = solve(t(X)%*%W%*%X)
   H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
   h = diag(H)
   e[,i] = sort(resid(fit, type="deviance")/sqrt(1-h)) }
   #
   e1 = numeric(k)
   e2 = numeric(k)
   #
   for(i in 1:k){
   eo = sort(e[i,])
   e1[i] = eo[5]
   e2[i] = eo[95]}
   #
   med = apply(e,1,mean)
   faixa = range(td,e1,e2)
   par(pty="s")
   qqnorm(td,xlab="Quantil da N(0,1)", ylab="Componente do Desvio",
ylim=faixa)
   #
   par(new=TRUE)
   qqnorm(e1,axes=FALSE,xlab=,ylab=,type="l",ylim=faixa,lty=1)
   par(new=TRUE)
   qqnorm(e2,axes=FALSE,xlab=,ylab=, type="l",ylim=faixa,lty=1)
```

```
par(new=TRUE)
qqnorm(med,axes=FALSE,xlab=, ylab=, type="l", ylim=faixa, lty=2)
```

Modelos com resposta de Poisson

```
X = model.matrix(fit.model)
   n = nrow(X)
   p = ncol(X)
   w = fit.model$weights
   W = diag(w)
   H = solve(t(X)%*%W%*%X)
   H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
   h = diag(H)
   td = resid(fit.model,type="deviance")/sqrt(1-h)
   #
   e = matrix(0,n,100)
   for(i in 1:100){
   nresp = rpois(n, fitted(fit.model))
   fit = glm(nresp ~ X, family=poisson)
   w = fit$weights
   W = diag(w)
   H = solve(t(X)%*%W%*%X)
   H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
   h = diag(H)
   e[,i] = sort(resid(fit,type="deviance")/sqrt(1-h)) }
   #
   e1 = numeric(n)
   e2 = numeric(n)
```

```
   #
   for(i in 1:n){
   eo = sort(e[i,])
   e1[i] = eo[5]
   e2[i] = eo[95] }
   #
   med = apply(e,1,mean)
   faixa = range(td,e1,e2)
   par(pty="s")
   qqnorm(td, xlab="Quantil da N(0,1)", ylab="Componente do Desvio
Padronizado", ylim=faixa)
   par(new=TRUE)
   qqnorm(e1,axes=FALSE,xlab=, ylab=, type="l", ylim=faixa, lty=1)
   par(new=TRUE)
   qqnorm(e2,axes=FALSE, xlab=, ylab=, type="l", ylim=faixa, lty=1)
   par(new=TRUE)
   qqnorm(med,axes=FALSE, xlab=, ylab=, type="l", ylim=faixa, lty=2)
```

Modelos com resposta binomial negativa

```
X = model.matrix(fit.model)
   n = nrow(X)
   p = ncol(X)
   fi = fit.model$theta
   w = fi*fitted(fit.model)/(fi + fitted(fit.model))
   W = diag(w)
   H = solve(t(X)%*%W%*%X)
   H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
```

```
h = diag(H)
td = resid(fit.model,type="deviance")/sqrt(1-h)
#
e = matrix(0,n,100)
for (i in 1:100) {
resp = rnegbin(n,fitted(fit.model),fi)
fit = glm.nb( resp ~ X)
fi = fit$theta
w = fit$weights
W = diag(w)
H = solve(t(X)%*%W%*%X)
H = sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
h = diag(H)
e[,i] = sort(resid(fit, type="deviance")/sqrt((1-h))) }
#
e1 = numeric(n)
e2 = numeric(n)
#
for (i in 1:n) {
eo = sort(e[i,])
e1[i] = eo[5]
e2[i] = eo[95]
#
med = apply(e,1,mean)
faixa = range(td,e1,e2)
par(pty= "s")
```

```
   qqnorm(td, xlab="Quantil da N(0,1)", ylab="Componente do Desvio
Padronizado", ylim=faixa)
   par(new=TRUE)
   qqnorm(e1,axes=FALSE, xlab=, ylab=, type="1", ylim=faixa,lty=1)
   par(new=TRUE)
   qqnorm(e2,axes=FALSE, xlab=, ylab=, type="1", ylim=faixa, lty=1)
   par(new=TRUE)
   qqnorm(med,axes=FALSE, xlab=, ylab=, type="1", ylim=faixa, lty=2)
```

Bibliografia

Agresti A (1990) *Categorical Data Analysis, First Edition.* Wiley.

Agresti A (2013) *Categorical Data Analysis, Third Edition.* Wiley.

Akaike H (1974) A new look at statistical model identification. *IEEE Transactions on Automatic Control* AU-19:716-722.

Aranda-Ordaz FJ (1981) On two families of transformations to additivity for binary response data. *Biometrika* 68:357-364.

Armitage P (1955) Test for linear trend in proportions and frequencies. *Biometrics* 11:375-386.

Armitage P (1971) *Statistical Methods in Medical Research.* Blackwell Scientific Publications.

Atkinson AC (1981) Two graphical display for outlying and influential observations in regression. *Biometrika* 68:13-20.

Atkinson AC (1985) *Plots, Transformations and Regressions.* Oxford Statistical Science Series.

Belsley DA, Kuh E, Welsch RE (1980) *Regression Diagnostics.* Wiley.

Bliss CI (1935) The calculation of the dosage-mortality curve. *Annals of Applied Biology* 22:134-167.

Bishop YMM, Fienberg SE, Holland PW (1975) *Discrete Multivariate Analysis: Theory and Practice*. MIT Press.

Boice JD, Monson RR (1977) Breast cancer in women after repeated fluoroscopic examinations of the chest. *Journal of the National Cancer Institute* 59:823-832.

Box GEP, Cox DR (1964) An analysis of transformations (with discussion). *Journal of the Royal Statistical Society B* 26:211-252.

Box GEP, Draper NR (1987) Empirical Model-Building and Response Surfaces. John Wiley & Sons.

Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88:9-25.

Breslow NE, Day NE (1980) *Statistical Methods in Cancer Research, Vol. I, The Analysis of Case-Control Studies*. IARC Scientific Publications, International Agency for Research on Cancer, Lyon.

Breslow NE, Day NE (1987) *Statistical Methods in Cancer Research, Vol. II, The Design and Analysis of Cohort Studies*. IARC Scientific Publications, International Agency for Research on Cancer, Lyon.

Buse A (1982) The likelihood ratio, Wald and Lagrange multiplier tests: an expository note. *The American Statistician* 36:153-157.

Canty A, Ripley B, Brazzale AR (2024) Package "Boot". https://cran.r-project.org/web/packages/boot/boot.pdf

Cardoso-Neto J, Paula GA (2001). Wald one-sided test using generalized estimating equations approach. *Computational Statistics and Data Analysis* 36:475-495.

Collett D (1991) *Modelling Binary Data*. Chapman and Hall.

Collett D (2003) *Modelling Survival Data in Medical Research, Second Edition*. CRC Press.

Colosimo ER, Giolo SR (2024). *Análise de SWobrevivência Aplicada, 2ª Edição*. Blucher.

Cook RD (1977) Detection of influential observations in linear regressions. *Technometrics* 19:15-18.

Cook RD (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B* 48:133-169.

Cook RD, Weisberg S (1982) *Residuals and Influence in Regression*. Chapman and Hall.

Cordeiro GM, Demétrio C, Moral R (2024) *Modelos Lineares Generalizados e Aplicações*. Blucher.

Cordeiro GM, Ferrari SLP, Paula GA (1993) Improved score tests for generalized linear models. *Journal of the Royal Statistical Society B* 55:661-674.

Cordeiro GM, Paula GA, Botter DA (1994) Improved likelihood ratio tests for dispersion models. *International Statistical Review* 62:257-274.

Cordeiro GM, Paula GA (1989) Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika* 76:93-100.

Cornfield J (1951) A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast and crevix. *Journal of the National Cancer Institute* 11:1269-1275.

Cornfield J (1956) A statistical problem arising from retrospective studies. *In: Proceedings of the Third Berkeley Symposium*, Berkeley, University of California Press, pgs. 133-148.

Cox DR (1970) *The Analysis of Binary Data*. Methuen.

Cox DR (1972) Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B* 74:187-220.

Cox DR, Hinkley DV (1974) *Theorical Statistics*. Chapman and Hall.

Cox DR, Oakes D (1984) *Analysis of Survival Data*. Chapman and Hall.

Cox DR, Snell EJ (1968) A general definition of residuals (with discussion). *Journal of the Royal Statistical Society B* 30:248-275.

Cox DR, Snell EJ (1989) *The Analysis of Binary Data, 2nd Edition*. Chapman and Hall.

Davison AC, Gigli A (1989) Deviance residuals and normal scores plots. *Biometrika* 76:211-221.

Day NE, Byar DP (1979) Testing hypothesis in case-control studies-equivalence of Mantel-Haenszel statistics and logit score tests. *Biometrics* 35:623-630.

de Jong P, Heller GZ (2008) *Generalized Linear Models for Insurance Data*. Cambridge University Press.

Diggle PJ, Liang KY, Zeger SL (1994) *Analysis of Longitudinal Data.* Oxford University Press.

Dixon WJ (1987) *BMDP Statistical Software*. University of California Press.

Dunkler D, Plischke M, Leffondré K, Heinze G (2014) Augmented backward elimination: a pragmatic and purposeful way to develop statistical models. *Plos One* 9(11):e113677.

Dunn PK, Smyth GK (1996) Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics* 5:236-244.

Efron B (1979) Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7:1-26.

Efron B (1988) Logistic regression, survival analysis and the Kaplan-Meier curve. *Journal of the American Statistical Association* 83:414-425.

Everitt BS (1977) *The Analysis of Contingency Tables*. Chapman and Hall.

Everitt BS (1994) *A Handbook of Statistical Analysis using S-Plus*. Chapman and Hall.

Fahrmeir L, Kaufmann H (1985) Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics* 13:342-368.

Fahrmeir L, Klinger J (1994) Estimating and testing generalized linear models under inequality constraints. *Statistical Papers* 35:211-229.

Fahrmeir L, Tutz G (2001). *Multivariate Statistical Modelling based on Generalized Linear Models*. Springer.

Faraway JJ (2016) *Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models, 2nd Edition*. Chapman and Hall/CRC.

Feigl P, Zelen M (1965) Estimation of exponential survival probabilities with concomitant information. *Biometrics* 21:826-838.

Finney DJ (1971) *Probit Analysis, 3rd. Edition.* Cambridge University Press.

Finney DJ (1978) *Statistical Methods in Biological Assay, 3rd. Edition.* Cambridge University Press.

Fieller EC (1954) Some problems in interval estimation. *Journal of the Royal Statistical Society B* 16:175-185.

Foster DP, Stine RA, Waterman RP (1998) *Business Analysis using Regression.* Springer.

Fox J, Weisberg S (2019) *An R Companion to Applied Regression, 3nd Edition.* Sage.

Goñi R, Alvarez F, Adlerstein S (1999) Application of generalized linear modeling to cath rate analysis of western mediterranean fisheries: the Castellón trawl fleet as a case study. *Fisheries Research* **42**, 291-302.

Gray JB (1989) On the use of regression diagnostics. *The Statistician* 38:97-105.

Green PJ, Silverman BW (1994) *Nonparametric Regression and Generalized Linear Models.* Chapman and Hall.

Griffiths WE, Hill RC, Judge GG (1993) *Learning and Practicing Econometrics.* John Wiley and Sons.

Gujarati D (2006) *Econometria Básica, 4ª Edição.* Campus.

Hadgu A, Koch G (1999) Application of generalized estimating equations to a dental randomized clinical trial. *Journal of Biopharmaceutical Statistics* 9:161-178.

Hand DJ, Daly F, Lunn AD, McConway KJ, Ostrowski E (1994) *A Handbook of Small Data Sets*. Chapman and Hall.

Hannan J, Harkness W (1963) Normal approximation to the distribution of two independent binomials, conditional to the sum. *Annals of Mathematical Statistics* 34:1593-1595.

Hardin JW, Hilbe JM (2012) *Generalized Estimating Equations*. Chapman and Hall.

Hastie T, Tibshirani R (1990) *Generalized Additive Models*. Chapman and Hall.

Hinde J (1982) Compoud poisson regression models. In R. Gilchrist Ed., GLIM82, pgs. 109-121. Springer.

Hinde J, Demétrio CGB (1998). Overdispersion: model and estimation. *Computational Statistics and Data Analysis* **27**, 151-170.

Hoaglin DC, Welsch RE (1978) The hat matrix in regression and ANOVA. *The American Statistician* 32:17-22.

Hosmer DW, Lemeshow S (1989) *Applied Logistic Regression, 1st Edition*. Wiley.

Hosmer DW, Lemeshow S, Sturdivant R (2013) *Applied Logistic Regression, 3nd Edition*. Wiley.

Hubert M, Vandervierin E (2008) An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis* 32:5186-5201.

Innes JRM, Ulland BM, Valerio MG, Petrucelli L, Fishbein L, Hart ER, Pallota AJ, Bates RR, Falk HL, Gart JJ, Klein M, Mitchell I, Peters J (1969) Biossay of pesticides and industrial chemicals for tumorigenicity in mice: A preliminary note. *Journal of the National Cancer Institute* 42:1101-1114.

Jørgensen B (1987) Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society B* 49:127-162.

Jørgensen B (1997) *The Theory of Dispersion Models*. Chapman and Hall.

Keen KJ (2010) Graphics for Statistics and Data Analysis with R. CRC Press.

Kwan CW, Fung WK (1998) Assessing local influence for specific restricted likelihood: Applications to factor analysis. *Psychometrika* 63:35-46.

Lawless JF (1982) *Statistical Models and Methods for Lifetime Data*. Wiley.

Lawless JF (2003) *Statistical Models and Methods for Lifetime Data, Second Edition*. Wiley.

Lawless JF (1987) Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* 15:209-225.

Lawrence AJ (1988) Regression transformation diagnostics using local influence. *Journal of the American Statistical Association* 84:125-141.

Lee ET (1991) *Statistical Methods for Survival Data Analysis, Second Edition*. Wiley.

Lee Y, Nelder JA (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society B* 58:619-678.

Lee Y, Nelder JA (2001) Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika* 88:987-1006.

Lee Y, Nelder JA, Pawitan Y (2006) *Generalized Linear Models with Random Effects: Unified Analysis via h-likelihood.* Chapman and Hall/CRC.

Leemis LM, Trivedi KS (1996) A comparison of aproximate interval estimators for the Bernoulli parameter. *The American Statistician* 50:63-68.

Lehnman EL Casella G (2011) *Theory of Point Estimation, Second Edition.* Springer.

Leiva V, Barros M, Paula GA (2009) *Generalized Birnbaum-Sanders Models using R.* Livro Texto de Minicurso da 11ª Escola de Modelos de Regressão, Recife - PE. ABE-Associação Brasileira de Estatística.

Lesaffre E, Verbeke G (1998) Local influence in linear mixed models. *Biometrics* 54:570-582.

Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73:13-22.

Lindsey JK (1997) Applying Generalized Linear Models. Springer.

Mantel N (1963) Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association* 58:690-700.

Mantel N, Haenszel BF (1959) Statistical aspects of the analysis of the data from retrospective studies of disease. *Journal of the National Cancer Institute* 22:719-748.

McCullagh P (1983) Quasi-likelihood functions. *Annals of Statistics* 11: 59-67.

McCullagh P (1987) *Tensor Methods in Statistics*. Chapman and Hall.

McCullagh P, Nelder JA (1989) *Generalized Linear Models, 2nd. Edition*. Chapman and Hall.

McCulloch CE, Searle SR (2001) *Linear and Generalized Linear Mixed Models*. Wiley.

Milicer H, Szczotka F (1966) Age at menarche in Warsaw girls in 1965. *Human Biology* 38:199-203.

Montgomery DC, Peck EA, Vining GG (2001) *Introduction to Linear Regression Analysis, Fourth Edition*. Wiley.

Montgomery DC, Peck EA, Vining GG (2021). *Introduction to Linear Regression Analysis, Sixth Edition*. Wiley.

Moolgavkar SH, Lustbader ED, Venzon DJ (1984) A geometric approach to non-linear regression diagnostics with application to matched case-control studies. *Annals of Statistics* 12:816-826.

Morgan BJT (1992) *Analysis of Quantal Response Data*. Chapman and Hall.

Munnell AH (1990) Why has productivity declined? Productivity and public investment. *New England Econonic Review* Jan-Fev 3-22.

Myers RH, Montgomery DC, Vining GG (2002) *Generalized Linear Models: With Applications in Engineering and the Sciences*. Wiley.

Mullahy J 1986 Specification and testing of some modified count data models. *Journal of Econometrics* 33:341-365.

Nelder JA, Pregibon D (1987) An extended quasi-likelihood function. *Biometrika* 74:221-232.

Nelder JA, Wedderburn RWM (1972) Generalized linear models. *Journal of the Royal Statistical Society A* 135:370-384.

Neter J, Kutner MH, Nachtsheim CJ, Wasserman W (1996) *Applied Linear Regression Models, 3rd Edition*. Irwin.

Nyquist H (1991). Restricted estimation of restricted generalized linear models. *Applied Statistics* 40:133-141.

Palmgren J (1981) The Fisher information matrix for log linear models against conditionally on observed explanatory variables. *Biometrika* 68:563-566.

Pan W (2001) Akaike's information criterion in generalized estimating equations. *Biometrics* 57:120-125.

Park TP, Shin DW, Park CG (1998) A generalized estimating equations approach for testing ordered group effects with repeated measurements. *Biometrics* 54:1645-1653.

Paula GA (2013) On diagnostics in double generalized linear models. *Computational Statistics and Data Analysis* 68:44-51.

Paula GA, Artes R (2000) One-sided test to assess correlation in logistic linear models using estimating equations. *Biometrical Journal* 42:701-714.

Paula GA, Leiva V, Barros M, Liu S (2012) Robust statistical modeling using Birnbaum-Saunders-t distribution applied to insurance. *Applied Stochastic Models in Business and Industry* 28:16-34.

Paula GA, Kumagaia GH (2014) Relatório de Análise Estatística sobre o Projeto: *Variabilidade Espaço-Temporal da Captura da Raia Branca, Dasyatis Cuttata, na Pesca Artesanal da Bahia de Todos os Santos*. RAE CEA14P04, IME-USP.

Paula GA, Oshiro CH (2001) Relatório de Análise Estatística sobre o Projeto: *Análise de Captura por Unidade de Esforço do Peixe-Batata na Frota Paulista*. RAE-CEA0102, IME-USP.

Paula GA, Tuder RM (1986) Utilização da regressão logística para aperfeiçoar o diagnóstico de processo infeccioso pulmonar. *Revista Ciência e Cultura* 40:1046-1050.

Paula GA, Sevanes M, Ogando MA (1988) Relatório de Análise Estatística sobre o Projeto: *Estudo de Plantas Brasileiras com Efeito Moluscicida em Biomphalaria Glabrata*. RAE-CEA8824, IME-USP.

Paula GA, Tavares HR (1992) Relatório de Análise Estatística sobre o Projeto: *Ácaros Associados ao Esterco Bovino. Subsídios para Controle Biológico da Mosca do Chifre*. RAECEA 9206, IME-USP.

Paula GA, Moura AS, Yamaguchi AM. Relatório de Análise Estatística sobre o Projeto *Estabilidade Sensorial de Snacks Aromatizados com*

Óleo de Canola e Gordura Vegetal Hidrogenada. RAECEA04P05, IME-USP.

Peduzzi PN, Hardy RJ, Holford TT (1980) A stepwise variable selection procedure for nonlinear regression models. *Biometrics* 36:511-516.

Piegorsch WW, Casella G (1988) Confidence bands for logistic regression with restricted predictor variables. *Biometrics* 44:739-750.

Poon W, Poon Y (1999) Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society B* 61:51-61.

Pregibon D (1981) Logistic regression diagnostics. *Annals of Statistics* 9:705-724.

Pregibon D (1982) Score tests in GLIM with applications. *Lecture Notes in Statistics* 14:87-97.

Pregibon D (1984) Data analytic methods for matched case-control studies. *Biometrics* 40:639-651.

Quenouille MH (1956) Notes on bias in estimation. *Biometrika* 43:353-360.

Ramanathan R (1993) *Statistical Methods in Econometrics*. Wiley.

Rao CR (1973) *Linear Statistical Inference and Its Applications, Second Edition*. Wiley.

Ratkowsky DA (1983) *Nonlinear Regression Modelling*. Marcel Dekker.

Rigby RA, Stasinopoulos DM (2005) Generalized Additive Models for Location, Scale and Shape. *Applied Statistics* 54:507-554.

Ruppert D (2004) *Statistical and Finance*. Springer.

Ryan BF, Joiner BL (1994) *Minitab Handbook, Third Edition*. Duxbury Press.

Schwarz G (1978) Estimating the Dimension of a Model. *Annals of Statistics* 6:461-464.

Seber GAF, Wild CJ (1989) *Nonlinear Regression*. Wiley.

Sen PK, Singer JM (1993) *Large Sample Methods in Statistics: An Introduciton with Applications*. Chapman and Hall.

Silva GL (1992) *Modelos Logísticos para Dados Binários*. Dissertação de Mestrado, IME-USP.

Smyth GK (1989) Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society B* 51:47-60.

Smyth GK, Verbyla A (1999) Adjusted likelihood methods for modelling dispersion in generalized linear models. *Environmetrics* 10:696-709.

Stasinopoulos MD, Righy RA, Gillian ZA, Voudouris V, de Bastiani F (2017) *Flexible Regression and Smoothing Using GAMLSS in R*. Chapman and Hall/CRC.

Stukel TA (1988) Generalized logistic models. *Journal of the American Statistical Association* 83:426-431.

Svetliza CF (2002) *Modelos Não-Lineares com Resposta Binomial Negativa*. Tese de Doutorado em Estatística. Instituto de Matemática e Estatística - Universidade de São Paulo..

Venables WN, Ripley BD (1999) *Modern Applied Statistics with S-Plus, Third Edition*. Springer.

Venezuela MK, Botter DA, Sandova lMC (2007) Diagnostic techniques in generalized estimating equations. *Journal of Statistical Computation and Simulation* 77:879-888.

Venezuela MK, SandovalMC, Botter DA (2011) Local influence in estimating equations. *Computational Statistics and Data Analysis* 55:1867-1883.

Venzani J (2004) *Using R for Introductory Statistics*. Chapman and Hall/CRC.

Verbyla AP (1993) Modelling variance heterogeneity: residual maximum likelihood and diagnostics. *Journal of the Royal Statistical Society B* 55:493-508.

Wang PC (1985). Adding a variable in generalized linear models. *Technometrics* 27:273-276.

Wedderburn RWM (1974) Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* 61:439-447.

Wedderburn RWM (1976) On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika* 68:27-32.

Wei BC (1998) *Exponential Family Nonlinear Models*. Lecture Notes in Statistics Vol. 130. Springer, New York.

Wei BC, Hu YQ, Fung WK (1998) Generalized leverage and its applications. *Scandinavian Journal of Statistics* 25:25-37.

Williams DA (1984) Residuals in generalized linear models. *In: Proceedings of the 12th. International Biometrics Conference*, Tokyo, pp. 59-68.

Williams DA (1987) Generalized linear model diagnostic using the deviance and single case deletion. *Applied Statistics* 36:181-191.

Wolf (1955) On estimating the relationship between blood group and disease. *Annals of Human Genetic* 19:251-253.

Wood FS (1973) The use of individual effects and residuals in fitting equations to data. *Technometrics* 15:677-687.

Wood SN (2017) *Generalized Additive Models. An Introduction with R, 2nd Edition*. Chapman and Hall/CRC.

Zeileis A, Kleiber C, Jackman S (2008) Regression models for count data in R. *Journal of Statistical Software* 27:1-25.