ANÁLISE DE SOBREVIVÊNCIA

Airlane P. Alencar – IME-USP

Alessandra C. Gourlart – FM-USP

Modelo de regressão linear usual

- $Y_i = \beta_0 + \beta_1 x_i + e_i$
- A variável resposta Yi poderia ser o tempo até a ocorrência do evento. Mas como o tempo tem que ser positivo é usual propor transformações nos dados.

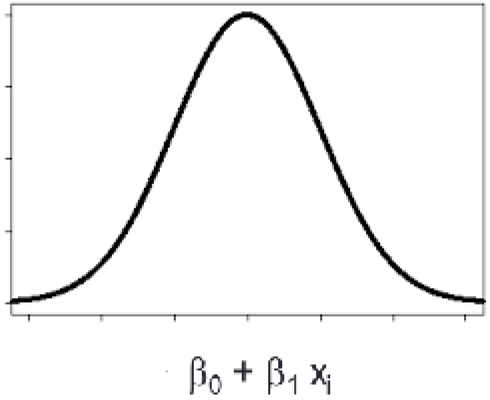
- $Y_i = \log(T_i) = \beta_0 + \beta_1 x_i + v_i$
- O modelo de T seria multiplicativo mas é linearizável, com v tendo dist.valor extremo padrão com f(v)=exp(v-exp(v)).
- T >=0 pois $T_i = \exp(\beta_0 + \beta_1 x_i + v_i)$
- $S(t|x) = P(T>t) = \exp\left[-\left(\frac{t}{\exp(\beta_0 + \beta_1 x)}\right)\right]$

Estimador de máxima verossimilhança

- Ideia: encontra os valores de parâmetros que maximizam a probabilidade ou densidade dos dados que você observou.
- Apresenta boas propriedades de consistência e variância assintótica.
- Em regressão linear usual, usamos também o método de mínimos quadrados mas como estimar a variância dos erros?

Estimador de máxima verossimilhança

- $Y_i = \beta_0 + \beta_1 x_i + e_i$
- Supondo que os erros ei são independentes e ei~N(0, σ^2) então Yi~N($\beta_0 + \beta_1 x_i$, σ^2) independentes e usamos a densidade da normal.



$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right]$$

- Vamos encontrar β_0, β_1 e σ^2 que maximizam a verossimilhança da amostra $y_1, \ldots, y_n$ que temos.

Estimador MV - Normalidade

- Função de verossimilhança
- $L(\beta_0, \beta_1, \sigma^2) = f(y_1)*f(y_2)*...*f(y_n)$

- Em regressão linear usual os estimadores MV de β_0 e β_1 são iguais aos de mínimos quadrados e obtemos também a estimação de σ^2 como a soma de quadrados dos resíduos dividido por n.

- Análise de resíduos

  Checar se os resíduos tem média 0, variância constante, são não correlacionados (séries temporais) e têm distribuição normal.

Máxima verossimilhança com censura

- Como incluir as observações censuradas?
- Usamos a densidade quando ocorreu a falha.
- Para os dados com censura à direita, sabemos que se a observação foi censurada no tempo t, $P(T>t)=S(t)$.

- 6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+
- Verossimilhança
- $L(\text{parâmetros}) = f(6)\, f(6)\, f(6)\, f(7) \ldots f(23) S(6) S(9) \ldots S(35)$
- Suponhamos que o tempo tenha distribuição $Exp(\lambda)$.

Distribuição Exponencial

- Densidade de probabilidade $f(t) = \lambda \exp(-\lambda t)$
- Sobrevivência: P(S>t) $S(t) = \exp(-\lambda t)$
- Risco constante: $h(t) = \lambda$

Risco= Taxa de falha em t condicional a sobreviver a t (T>=t)

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t)}{\Delta t\, S(t)} = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)}$$

- Considerando $\lambda_i = \beta_0 + \beta_1 x_i$

  Temos que o risco de cada pessoa i depende suas explicativas.

Weibull

- Densidade de probabilidade $f(t) = h(t)S(t)$
- Sobrevivência: P(S>t) $S(t) = \exp\left(-\lambda t^{p}\right)$
- Risco constante: $h(t) = \lambda p t^{p-1}$
- Tem um parâmetro a mais (p).
- A taxa de risco não é mais constante

Exponencial e Weibull

- Exemplo Ovarian
- fit <- survfit(Surv(futime, fustat)~1 , data=ovarian)
- summary(fit)
- plot(fit, xlab="t",ylab=expression(hat(S)*"(t)"))
- s2=survreg(Surv(futime, fustat)~1 , ovarian, dist="exponential")
- summary(s2)
- Tempo=seq(0,1200,1)
- lines(Tempo,1-pexp(Tempo,exp(-7.169)))

- s3=survreg(Surv(futime, fustat)~1 , ovarian, dist="weibul")
- summary(s3)   # p=1/scale
- lines(Tempo,1-pweibull(Tempo,1/0.902,exp(7.111)), col=2)

Diagnóstico

- Depois do modelo de Cox