

MAE514 - Introdução à Análise de Sobrevivência e Aplicações

Modelo de Cox: Análise de Resíduos

Antonio Carlos Pedroso de Lima

IME-USP



Análise de resíduos no modelo de Cox

Verificação de suposições na modelagem

- Suposições do modelo
 - Suposição TFP
 - Forma funcional para introduzir covariáveis no modelo
- Detecção de valores aberrantes
- Estudo de observações influentes

Resíduos mais usados

- Cox-Snell
- Martingal
- Deviance
- Escore
- Shoenfeld



Análise de resíduos no modelo de Cox

Modelo de Regressão Semi-paramétrico

$$\alpha(t | \mathbf{X}) = \alpha_o(t) \exp\{\mathbf{X}(t)\beta\}$$

- $\alpha_o(t)$: Função não negativa, arbitrária e desconhecida
- Poucas suposições
- Forma funcional das covariáveis
 - Taxas de falhas proporcionais (TFP) quando $X(t) = X$, $\forall t \geq 0$



Resíduos mais utilizados para o modelo de Cox

Cox-Snell

$$r_i = \hat{A}_o(t) \exp\left\{\sum_{k=1}^p X_{ik} \hat{\beta}_k\right\}$$

- $\hat{A}_o(t)$: Estimador de Breslow
- $\hat{\beta}_k$: Estimador de máxima verossimilhança parcial

Principal uso:

Avaliação da qualidade geral do ajuste

- Sem indicações sobre causas de falta de ajuste
- Envolve estimadores obtidos em duas etapas: cautela!



Resíduos mais utilizados para o modelo de Cox

Martingal

$$\widehat{M}_i = \delta_i - r_i$$

- δ_i : indicador de falha
- r_i : resíduo Cox-Snell (covariáveis fixas)

Principais usos:

- Identificação de observações que não são previstas adequadamente pelo modelo ajustado (positiva)
- Verificação de possíveis transformações nas covariáveis

- Assimetria negativa
- Derivado a partir da decomposição Doob-Meier para processos pontuais



Resíduos mais utilizados para o modelo de Cox

Escore

$$\mathbf{U}_i(\widehat{\beta}) = \int_0^{\tau} [\mathbf{X}_i(s) - \bar{\mathbf{x}}_{\widehat{\beta}}(s)] dM_i(s)$$

- Definido para cada unidade experimental ($i = 1, \dots, n$).
- $\dim(\mathbf{U}_i) = p \times 1$: resíduos para cada uma das covariáveis.
- Contêm a contribuição de cada indivíduo para o vetor escore

Principal Uso

- Avaliação de observações influentes [Therneau & Grambsch (2000)].
- Verificação da suposição de proporcionalidade [Klein & Moeschberger (1997)].



Resíduos mais utilizados para o modelo de Cox

Deviance

$$d_j = \text{sign}(\widehat{M}_j) \times \sqrt{-2\{\widehat{M}_j + \delta_j \log(\delta_j - \widehat{M}_j)\}}$$

Equivalente ao resíduo de Pearson, derivado em MLG

Principais usos:

- Evidenciar valores aberrantes (*outliers*)

Obs.:

- Fornece maior simetria que o resíduo martingal
- Os resíduos martingal e deviance evidenciam tempos de falha maiores que o esperado. Somente o resíduo deviance detecta tempos muito menores que o esperado.



Resíduos mais utilizados para o modelo de Cox

Supondo t_1, \dots, t_L instantes de falha,

$$\mathbf{s}_k = \int_{t_{k-1}}^{t_k} \sum_{i=1}^n [\mathbf{X}_i(s) - \bar{\mathbf{x}}_{\beta}(s)] d\widehat{M}_i(s)$$

Segue que

$$\mathbf{s}_k = \int_{t_{k-1}}^{t_k} \sum_{i=1}^n [\mathbf{X}_i(s) - \bar{\mathbf{x}}_{\beta}(s)] dN_i(s)$$

Schoenfeld

Principais usos

- Avaliação de observações influentes
- **Suposição de riscos proporcionais**



Exemplo no R

Dados de Cirrose Biliar Primária (PBC)

- Dados obtidos na *Mayo Clinic - EUA* entre 1974 e 1984
- 312 pacientes com PBC, aleatorizados em dois grupos:
 - Placebo
 - Droga Penicilamina-D

Resposta:

- Tempo, em dias, entre a entrada no protocolo e
 - Óbito
 - Transplante de fígado
 - Fim do estudo (julho de 1986)
- Indicador de censura (Delta)
 - 0: Paciente vivo sem transplante de fígado
 - 1: Paciente vivo com transplante de fígado
 - 2: Óbito devido à PBC

Exemplo no R - Dados PBC

Variáveis Laboratoriais:

- Bilirrubina sérica (mg/dL)
- Colesterol sérico (mg/dL)
- Albumina (mg/dL)
- Urina copper (ug/dia)
- Fosfatase alcalina (U/L)
- SGOT (U/mL)
- Triglicérides (mg/dL)
- Plaquetas (contagem por mL³/1000)
- Protrombina (segundos)
- Estágio histológico da doença (classificatória)

Exemplo no R - Dados PBC

Principal variável explicativa:

- Tratamento:
 - 1: Tratamento ativo (Penicilamina-D)
 - 2: Tratamento placebo

Outras variáveis:

- Idade (dias), Sexo (0:masc, 1:fem)
- Ascite(0:não, 1:sim), Hepatomegalia(0:não, 1:sim), Spiders(0:não, 1:sim)
- Presença de Edema:
 - 0: sem edema e sem terapia com diuréticos
 - 0.5: sem edema sem uso de diuréticos ou edema resolvido com diuréticos
 - 1: presença de edema apesar do uso de diuréticos

Exemplo no R - Transformação de Covariável

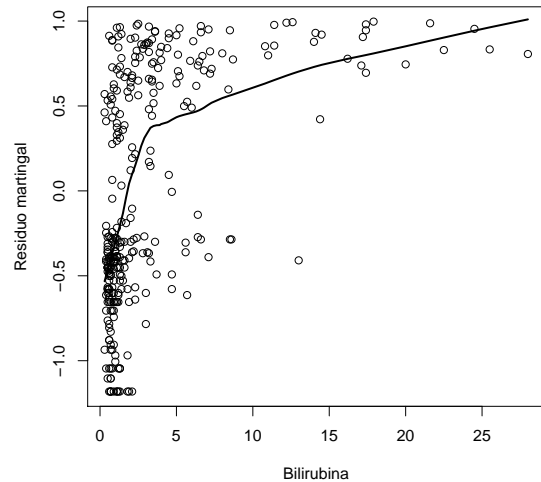
Nesta análise:

- Evento: Óbito ou Transplante (delta = 1 ou 2)
- Censura: Paciente vivo sem transplante ao final do acompanhamento

Resíduo Martingal

```
> Pbc<-read.table('Pbc.dat', heade=T)
> delta1<-ifelse(Pbc$delta==0, 0, 1)
> pbc.fit0<-coxph(Surv(tempo, delta1)~1, data=Pbc)
> plot(Pbc$bili, resid(pbc.fit0), xlab="Bilirubina",
+      ylab="Residuo martingal")
> smooth<-lowess(Pbc$bili, resid(pbc.fit0), iter=0)
> lines(smooth, lwd=2)
```

Exemplo no R - Transformação de Covariável

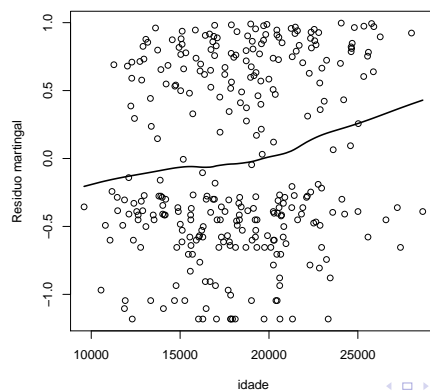


Indicação de transformação logaritmica

Exemplo no R - Transformação de Covariável

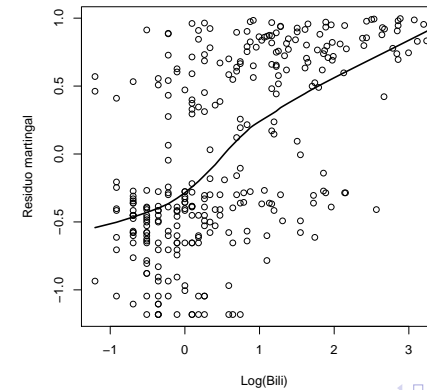
Verificação para Idade

```
> plot(Pbc$idade, resid(pbc.fit0), xlab="idade",
+      ylab="Residuo martingal")
> smooth2<-lowess(Pbc$idade, resid(pbc.fit0), iter=0)
> lines(smooth2, lwd=2)
```



Exemplo no R - Transformação de Covariável

```
> plot(log(Pbc$bili), resid(pbc.fit0), xlab="log(bili)",
+      ylab="Residuo martingal")
> smooth1<-lowess(log(Pbc$bili), resid(pbc.fit0),
+                 iter=0)
> lines(smooth1, lwd=2)
```



Exemplo no R - Transformação de Covariável

```
> pbc.fit1<-coxph(Surv(tempo, delta1)~idade+log(bili),
+                 data=Pbc)
> summary(pbc.fit1)
```

Saída

	coef	exp(coef)	se(coef)	z	p
idade	6.74e-05	1.00	2.11e-05	3.19	0.0014
log(bili)	1.05e+00	2.85	8.54e-02	12.25	0.0000

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	1.00	1.000	1.00	1.00
log(bili)	2.85	0.351	2.41	3.36

Rsquare= 0.389 (max possible= 0.991)
Likelihood ratio test= 154 on 2 df, p=0
Wald test = 162 on 2 df, p=0
Score (logrank) test = 186 on 2 df, p=0

Exemplo no R - Valores aberrantes

Resíduo Deviance

- Transformação do resíduo martingal
- Maior simetria → apropriado para detecção de *outliers*

Para a i -ésima observação, $i = 1, \dots, n$,

- d_i : resíduo deviance para obs. i , $i = 1, \dots, n$
- *preditor linear*

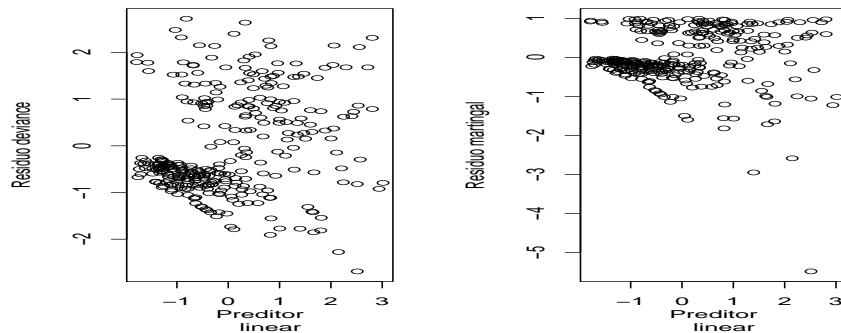
$$\sum_{j=1}^p X_{ij} \hat{\beta}_{ij}$$

Sugere-se analisar o gráfico $d_i \times$ o preditor linear, $i = 1, \dots, n$.



Exemplo no R - Valores aberrantes

```
> par(mfrow=c(1,2))
> plot(pbc.fit1$linear.predictors, resid(pbc.fit1, type='deviance'),
+ xlab='Preditor linear', ylab='Residuo deviance')
> plot(pbc.fit1$linear.predictors, resid(pbc.fit1, type='martingal'),
+ xlab='Preditor linear', ylab='Residuo martingal')
```



Notar a assimetria acentuada no resíduo martingal



Exemplo no R - Valores aberrantes

```
> attributes(pbc.fit1)

$names
 [1] "coefficients" "var"      "loglik"
 [4] "score"        "iter"    "linear.predictors"
 [7] "residuals"    "means"   "method"
[10] "n"            "terms"   "assign"
[13] "wald.test"    "y"       "formula"
[16] "call"

$class
 [1] "coxph"
```

- O objeto "pbc.fit1" produz **linear.predictors**



Exemplo no R - Observações Influentes

Seja

- $\hat{\beta}$: EMVP baseado nas n observações
- $\hat{\beta}_{(i)}$: EMVP baseado excluindo-se a i -ésima observação

O 'impacto' da observação i pode ser medido por

$$J_i = \hat{\beta} - \hat{\beta}_{(i)}$$

Pode ser mostrado que J_i é proporcional a

$$(X_i - \bar{X}) \times u_i$$

com u_i o *resíduo score*. Pode ainda ser mostrado que

$$\Delta \hat{\beta} = \mathbf{1}(U I_o)$$

- $\Delta \hat{\beta}$: variação no EMVP excluindo-se cada uma das obs.
- U matriz com os resíduos scores, para cada covariável
- I_o é a matriz de informação *observada*



Exemplo no R - Observações Influentes

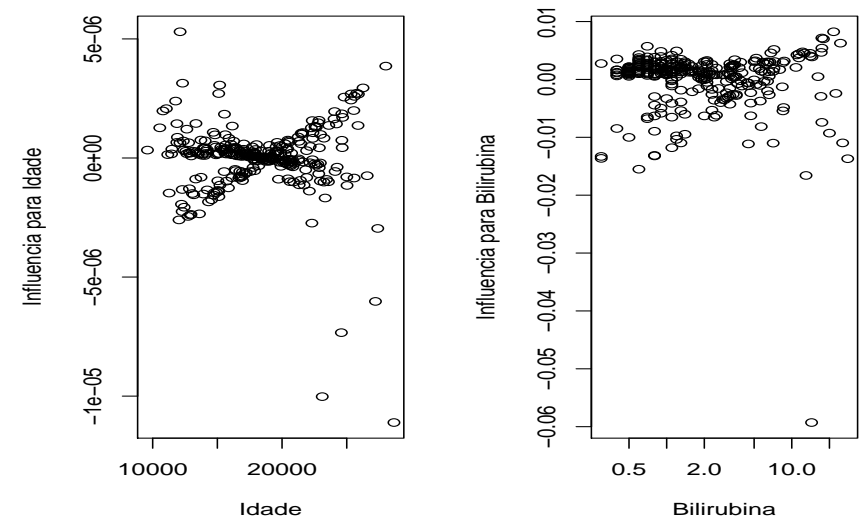
No R, definem-se os resíduos 'dfbeta' como elementos da matriz D , dada por

$$D = UI_0$$

Obtenção dos resíduos:

```
> par(mfrow=c(1,2))
> rr<-resid(pbc.fit1,type='dfbeta')
> plot(Pbc$idade, rr[,1],xlab='Idade',
+      ylab='Influencia para Idade')
> plot(Pbc$bili, rr[,2],xlab='Bilirubina',
+      ylab='Influencia para Bilirubina',log='x')
```

Exemplo no R - Observações Influentes



Note a presença de pelo menos uma obs. de grande influência

Exemplo no R - Ajuste geral do modelo

Resíduos Cox-Snell

```
# Small helper function to plot Cox-Snell residuals (modificada)
# From:
# www.stat.uni-muenchen.de/~semwiso/lebensdaueranalyse_ws0910/
# download/diagnostics.R
plot.coxsnell <- function(m, which=rep(TRUE,m$n),...) {
  delta <- m$y[,"status"]
  r.cs <- delta - residuals(m, type="martingale")
  r.surv <- survfit(Surv(r.cs[which],delta[which])~1,
  type="fleming-harrington")
  if (!is.null(list(...)$add)) {
    if (list(...)$add == TRUE) {
      lines(r.surv$time, -log(r.surv$surv),type="s",...)
    }
  } else {
    plot(r.surv$time, -log(r.surv$surv),
    xlab="Residuos Cox-Snell", ylab="Taxa de Falha Acumulada",
    type="s",ylim=c(0,max(r.cs)),...)
  }
  t <- seq(0, max(r.cs),length=100)
  lines(t,t,lwd=3)
  invisible()
}
```

Exemplo no R - Ajuste geral do modelo

Resíduos Cox-Snell

```
plot.coxsnell(pbc.fit1,las=1)
```

