

Modelo de Regressão paramétrica

Vicente G. Cancho
garibay@icmc.usp.br

Departamento de Matemática Aplicada e Estatística
Universidade de São Paulo

11 de fevereiro de 2020

Modelo de regressão paramétricos

- 1 Em muitas situações práticas temos que o tempo de vida é influenciado por uma ou mais variáveis regressoras $\mathbf{x}' = (x_1, \dots, x_p)$.
- 2 Por exemplo, em estudos com pacientes com câncer de pulmão, normalmente deseja-se medir a influência no tempo de vida dos pacientes, os fatores(ou das covariáveis) tais como idade, sexo, número de cigarros fumado por dia, outros casos na família, etc.
- 3 Em estudos na área industrial, por exemplo, o tempo de vida de um determinado isolante é naturalmente afetado pelo nível de voltagem a que ele é sujeito.

Modelo de regressão paramétricos

- 1 Uma maneira de determinar o relacionamento entre o tempo de vida e um conjunto de covariáveis é através de um modelo de regressão.
- 2 Três classes importantes de modelos de regressão:
 - Modelo de Locação-Escala (ou Modelos de Tempos de Falha Acelerada (TFA))
 - Modelos de chances proporcionais "proportional odds (PO)"
 - Modelo de Riscos Proporcionais.

Modelos de Riscos proporcionais

Uma família de riscos proporcionais (RP) é uma classe de modelos com a propriedade de que indivíduos diferentes tem função de *riscos proporcionais* entre si. Isto é, se $h_i(t)$ e $h_j(t)$ são as funções de riscos de dois indivíduos, a razão de riscos ou risco relativo (RR),

$$RR = \frac{h_i(t)}{h_j(t)}, \text{ não varia com } t.$$

Modelos de chances proporcionais "proportional odds (PO)"

- Chances de sobrevivência: chances de sobreviver além do tempo t

$$\frac{S(t)}{1 - S(t)} = \frac{P(T > t)}{P(T \leq t)}.$$

- Chances de falhas: chances de obter o evento no tempo t

$$\frac{F(t)}{1 - F(t)} = \frac{1 - S(t)}{S(t)} = \frac{P(T \leq t)}{P(T > t)}.$$

Modelo proportional odds (PO)

Em um modelo de sobrevivência de PO, a razão de chances de falha é constante ao longo do tempo. Isto é, Se $S_1(t)$ e $S_2(t)$ são as funções de sobrevivência de dois indivíduos, a razão de chance (odds ratio ,OR) para dois indivíduos (1 e 2) é

$$OR(1 \text{ vs } 2) = \frac{\frac{1-S_1(t)}{S_1(t)}}{\frac{1-S_2(t)}{S_2(t)}}, \text{ é constante ao longo de } t.$$

Modelos TFA

[Exemplo : Humanos vs Cachorros]

- Sejam $S_H(t)$ e $S_C(t)$ as funções de sobrevivência de seres humanos e cachorros respectivamente.
- Conhecido: Em geral os cachorros envelhecem sete vezes mais rápido do que humanos.
- Na terminologia TFA: A probabilidade de um cachorro sobreviver pelo menos t anos é igual a probabilidade sobreviver de humano pelo menos $7t$ anos.
- Isto significa que:

$$S_C(t) = S_H(7t)$$

- Assim, a probabilidade de que um cão sobreviver 10 anos ou mais, é a mesma probabilidade de que um ser humano sobreviver 70 anos ou mais.

Modelos TFA

- Em outras palavras, cães aceleram uns sete vezes mais rápido sua vida, do que os seres humanos.

Em geral

- Sejam $S_1(t)$ e $S_2(t)$ as funções de sobrevivência de duas populações,
- Os modelos TFA diz que existe uma constante $\gamma > 0$ tal que

$$S_1(t) = S_2(\gamma t). \quad t > 0$$

- Este modelo implica que a taxa de envelhecimento da população 1 é γ vezes mais do que a de população 2.

Modelo TFA

- Seja μ_i o tempo médio da população $i = 1, 2$ e t_{q_i} é o quantil populacional da população i , tal que, $S_i(t_{q_i}) = \theta$, com $0 < \theta < 1$. Então,

$$\begin{aligned}\mu_2 &= \int_0^{\infty} S_2(t) dt, \quad (t = \gamma u) \\ &= \gamma \int_0^{\infty} S_2(\gamma u) du \\ &= \gamma \int_0^{\infty} S_1(u) du \\ \mu_2 &= \gamma \mu_1\end{aligned}$$

e

$$S_2(t_{q_2}) = \theta = S_1(t_{q_1}) = S_2(\gamma t_{q_1})$$

Modelo TFA

- Supondo que $S_1(t)$ e $S_2(t)$ são funções estritamente decrescente. Então, temos

$$t_{q_2} = \gamma t_{q_1}$$

- Por exemplo, $q_i = 0,5$ (mediana), a relação dos tempos medianos é

$$md_2 = \gamma md_1$$

onde md_i é o tempo médiano da população $i = 1, 2$

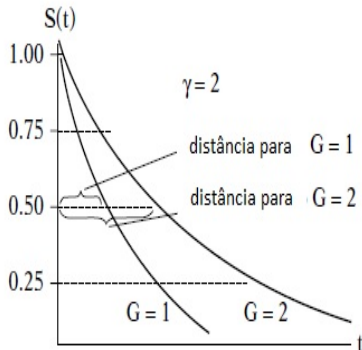
- Este argumento simples nos diz que sob o modelo de TFA, o tempo de sobrevivência esperado e o tempo de sobrevivência mediano da população 2 são γ vezes mais do que os de população 1.

Factor de aceleração

O factor de aceleração permite avaliar o efeito das covariáveis sobre o tempo de sobrevivência.

Factor de aceleração

O factor aceleração é um razão dos tempos quantiais correspondentes a qualquer valor fixo de $S(t)$.



Interpretação do factor de aceleração

Comparando dois grupos (tratamento vs placebo) leva à seguinte interpretação geral:

$\gamma > 1 \implies$ tratamento foi benéfico à sobrevivência do paciente

$\gamma < 1 \implies$ tratamento não foi benéfico à sobrevivência do paciente

Para a razão de riscos (RR) temos:

$RR > 1 \implies$ tratamento não foi benéfico à sobrevivência do paciente

$RR < 1 \implies$ tratamento foi benéfico à sobrevivência do paciente

$\gamma = RR = 1 \implies$ tratamento não afeta a sobrevivência do paciente

Uma forma mais geral do modelo TFA

Considere um modelo TFA com um preditor, x . O modelo pode ser expressa na escala log como:

$$\log T = \beta_0 + \beta_1 X + \epsilon$$

onde ϵ é um erro aleatório com alguma distribuição.

T	$\log(T)$
Exponencial	Valor Extremo
Weibull	Valor Extremo
Log-Nornal	Normal
Log-logística	Logística

Uma forma mais geral do modelo TFA

- Algumas distribuições (por exemplo Weibull) têm um parâmetro σ adicional, que dimensiona ϵ :

$$\log T = \beta_0 + \beta_1 X + \sigma \epsilon$$

- Aqui (e não R) o modelo é parametrizado por $\sigma = 1/\alpha$.

$$\log T = \beta_0 + \beta_1 X + \frac{1}{\alpha} \epsilon$$

Uma forma mais geral do modelo TFA

- O modelo em termos dos tempos de sobrevivência T é dado:

$$T = \exp\{\beta_0 + \beta_1 X\} \exp\left\{\frac{1}{\alpha}\epsilon\right\},$$

- modelo TFA é multiplicativo em termos de T e aditivo em termos de $\log(T)$.

Suposição do modelo TFA

A relação de tempo-quantil é constante (γ) para todos os valores fixos

$$S(t_q) = q.$$

Modelo de regressão de riscos proporcionais

- Objetivo: estender o conceito de riscos proporcionais para um cenário de regressão, para que possamos incorporar covariáveis.

Modelo de riscos proporcionais

$$h(t|x_1, \dots, x_p) = h_0(t) \exp\{b_1x_1 + \dots + x_p b_p\}$$

ou

$$S(t|x_1, \dots, x_p) = [S_0(t)]^{\exp\{x_1 b_1 + \dots + x_p b_p\}}$$

onde $h_0(t)$ é a função de risco básica (ou basal), x_1, \dots, x_p são as variáveis regressoras (ou covariáveis) e b_1, \dots, b_p os respectivos parâmetros de regressão.

Interpretação dos coeficientes de regressão

$$h(t|x_1, \dots, x_p) = h_0(t) \exp\{b_1x_1 + \dots + x_p b_p\}$$

- A função de risco basal é dada por

$$h_0(t) = h(t|x_1 = x_2 = \dots, x_p = 0).$$

- O risco relativo (RR) de comparar os valores de covariáveis x_1, \dots, x_p com $x_1 = \dots, x_p = 0$ é

$$RR = \frac{h(t|x_1, \dots, x_p)}{h(t|x_1 = x_2 = \dots, x_p = 0)} = \exp\{b_1x_1 + \dots + x_p b_p\}$$

Interpretação dos coeficientes de regressão

$$h(t|x_1, \dots, x_p) = h_0(t) \exp\{b_1x_1 + \dots + x_p b_p\}$$

- b_k : indica se aumentarmos em uma unidade a covariável x_k , assumindo constante todas as outras covariáveis, RR é aumentado em e^{b_k} . Isto é,

$$\begin{aligned} RR &= \frac{h(t|x_1, \dots, x_k + 1, \dots, x_p)}{h(t|x_1, \dots, x_k, \dots, x_p)} \\ &= \exp\{b_1(x_1 - x_1) + \dots + b_k(x_k - x_k + 1) + \beta_p(x_p - x_p)\} \\ &= e^{b_k}. \end{aligned}$$

- $\beta_k = 0$ implica nenhum efeito da covariável x_k no risco de um evento. Isto é $RR = 1$
- Note que b_k independe de t , o qual implica que RR é assumido constante sobre o tempo.

Modelo de riscos proporcionais

- A razão de riscos (ou taxa de falha) de dois indivíduos é constante. Isto é, a razão das funções de taxa de falha para os indivíduos i e j é dada por

$$\frac{h_i(t|\mathbf{x}_i)}{h_j(t|\mathbf{x}_j)} = \frac{h_0(t) \exp\{\mathbf{x}_i^\top \mathbf{b}\}}{h_0(t) \exp\{\mathbf{x}_j^\top \mathbf{b}\}} = \exp\{(\mathbf{x}_i^\top - \mathbf{x}_j^\top) \mathbf{b}\}$$

e não depende do tempo.

Dados leucemia

Os dados a continuação trata dos tempos de remissão (em meses) do câncer (leucemia aguda) de 42 crianças tratadas com:6-mercaptopurine (6-MP) ou placebo.

Placebo	1	22	3	12	8	17	2	11	8	12	2
	5	4	15	8	23	5	11	4	1	8	
6-MP	10	7	32 ⁺	23	22	6	16	34 ⁺	32 ⁺	25 ⁺	11 ⁺
	20 ⁺	19 ⁺	6	17 ⁺	35 ⁺	6	13	9 ⁺	6 ⁺	10 ⁺	

+: indica tempo censurado.

Considere o modelo TFA Weibull com uma única covariável, $x = 1$, se a criança tratada com 6-MP e $x = 0$ se se a criança tratada com placebo.

modelo de TFA Weibull

Relembre: $S(t) = \exp\{-\lambda t^\alpha\}$

- 1 Resolvendo para t resulta:

$$\begin{aligned} S(t) = \exp\{-\lambda t^\alpha\} &\iff -\log(S(t)) = \lambda t^\alpha \\ &\iff t = [-\log(S(t))]^{1/\alpha} \lambda^{-1/\alpha} \end{aligned}$$

- 2 Reparametrizando, $\lambda^{-1/\alpha} = \exp\{\beta_0 + \beta_1 x\}$.

$$t = [-\log(S(t))]^{1/\alpha} \exp\{\beta_0 + \beta_1 x\}$$

- 3 Em termos de qualquer probabilidade fixa $S(t_q) = q$ obtemos

$$t_q = [-\log(q)]^{1/\alpha} \exp\{\beta_0 + \beta_1 x\}$$

Modelo de TFA Weibull: Interpretação do fator de aceleração

O factor de aceleração γ para um valor fixo de $S(t) = q$ é calculada do seguinte modo:

Níveis de covariáveis: $x = 1$ e $x = 0$

Fator de aceleração γ é

$$\begin{aligned}\gamma &= \gamma(x = 1, \text{ vs, } x = 0) \\ &= \frac{[-\log(q)]^{1/\alpha} \exp\{\beta_0 + \beta_1\}}{[-\log(q)]^{1/\alpha} \exp\{\beta_0\}} \\ &= \exp\{\beta_1\}.\end{aligned}$$

- Por exemplo, suponha que $\exp\{\beta_1\} = 2$, isso indica que o tempo mediano de crianças tratados com 6-MP é duas vezes o tempo médiano de crianças tratados com placebo.

modelo de regressão Weibull de RP

O modelo de regressão de Weibull -RP para os dados de leucemia é dado

$$h(t|x) = h_0(t) \exp(b_1x)$$

onde o risco basal é dado por $h_0(t) = \alpha \lambda t^{\alpha-1}$. Dai tem-se

$$h(t|x) = \alpha t^{\alpha-1} \exp(b_0 + b_1x)$$

Ou seja, $T|x \sim W(\alpha, \exp(b_0 + b_1x))$

Note que,

$$RR(x = 1, x = 0) = \frac{\alpha t^{\alpha-1} \exp(b_0 + b_1)}{\alpha t^{\alpha-1} \exp(b_0)} = \exp(b_1).$$

Relação entre os coeficientes de regressão do modelo Weibull-TFA e Weibull-RP

TFA:

$$\begin{aligned}\frac{1}{\lambda^{1/\alpha}} &= \exp\{\beta_0 + \beta_1 x\} \\ \implies (1/\alpha) \log(\lambda) &= -(\beta_0 + \beta_1 x) \\ \implies \log(\lambda) &= -\alpha(\beta_0 + \beta_1 x)\end{aligned}$$

RP:

$$\begin{aligned}\lambda &= \exp\{b_0 + b_1 x\} \\ \implies \log(\lambda) &= b_0 + b_1 x\end{aligned}$$

Isto indica a seguinte relação entre os coeficientes de:

$$b_j = -\alpha\beta_j, \quad j = 0, 1.$$

Modelo exponencial TFA e RP

Obtivemos $b_j = \alpha\beta_j$ para o modelo de Weibull. No caso especial do modelo exponencial em que $\alpha = 1$, temos

$$b_j = -\beta_j, \quad j = 0, 1$$

Observação

O modelo exponencial RP e TFA são, na verdade o mesmo modelo, com exceção que a parametrização seja diferente.

Modelo de TFA Weibull

Inferência para $\theta = (\beta, \sigma)$

Suponha que temos um conjunto de n observações independentes dos pares (t_i, δ_i) , $i = 1, \dots, n$ onde t_i é o tempo de falha ou censura, $\delta_i = 1$ se i -ésima unidade experimental falha e $\delta_i = 0$ se censura, ou seja,

$$y_i = \min(T_i, C_i), \quad \delta_i = \begin{cases} 1 & T_i < C_i \\ 0 & T_i \geq C_i \end{cases},$$

onde T_i é o tempo de falha do i -ésima unidade experimental, $i = 1, \dots, n$, e C_i é tempo de censura do i -ésima unidade experimental.

Modelo de TFA Weibull

Inferência para $\theta = (\beta, \sigma)$

A função de verossimilhança θ dado os dados observados e tempo de censura não informativa

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n [f(y_i|\mathbf{x}_i, \theta)]^{\delta_i} [S(y_i|\mathbf{x}_i, \theta)]^{1-\delta_i} \\ &= \prod_{i=1}^n [h(y_i|\mathbf{x}_i, \theta)]^{\delta_i} S(y_i|\mathbf{x}_i, \theta), \end{aligned}$$

em que δ_i é o indicador de falha para a i -ésima observação, $y_i = \log t_i = \min(\log(T_i), \log(C_i))$ é logaritmo do tempo de falha ou censura.

Código em R e Saída

```
model.wei=survreg(Surv(y,delta)~ x, dist='weibull')
```

```
summary(model.wei)
```

Call:

```
survreg(formula = Surv(y, delta) ~ x, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	2.248	0.166	13.55	8.30e-42
x	1.267	0.311	4.08	4.51e-05
Log(scale) -	0.312	0.147	-2.12	3.43e-02

```
Scale= 0.732
```

Weibull distribution

```
Loglik(model)= -106.6 Loglik(intercept only)= -116.4
```

```
Chisq= 19.65 on 1 degrees of freedom, p= 9.3e-06
```

```
Number of Newton-Raphson Iterations: 5
```

```
n= 42
```

Código em R e Saída

O fator de aceleração estimada $\hat{\gamma}$ comparando o grupo de tratamento (6-MP) para o grupo de placebo ($x=1$ vs $x=0$) é

$$\hat{\gamma} = \exp\{\hat{\beta}_1\}$$

```
> exp(model.wei$coeff[2])
```

```
x
```

```
3.551374
```

Interpretação

O tempo de sobrevivência para o grupo de tratamento ($x = 1$) é aumentada por um factor de 3,55 em comparação com o grupo placebo ($x = 0$).

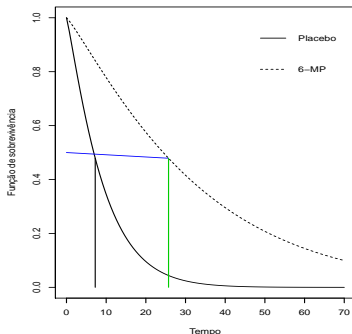
⇒ **o tratamento é positivo.**

Funções de sobrevivência

Cálculo de quantieis, por exemplo, a mediana:

$$S(t) = \exp \left\{ - \left(\frac{t}{\exp(\beta_0 + \beta_1 x)} \right)^\alpha \right\} = 0,5$$
$$\implies \widehat{t}_{md} = [\log(2)]^{1/\hat{\alpha}} \exp\{\hat{\beta}_0 + \hat{\beta}_1\}.$$

```
> mediana=predict(model.wei,  
+ newdata=list(x=c(0,1)),  
+ type='quantile',p=0.5)  
> mediana  
1          2  
7.242697 25.721526  
> mediana[2]/mediana[1]  
3.551374
```



Ajuste do Modelo exponencial TFA e RP

Exemplo: Ajuste dos dados de leucemia ao modelo Exponencial- TFA

```
model.exp=survreg(Surv(y,delta)~ x, dist='exponential')
```

```
summary(model.exp)
```

Call:

```
survreg(formula = Surv(y, delta) ~ x, dist = "exponential")
```

	Value	Std. Error	z	p
(Intercept)	2.16	0.218	9.90	4.33e-23
x	1.53	0.398	3.83	1.27e-04

Scale fixed at 1

Exponential distribution

Loglik(model)= -108.5 Loglik(intercept only)= -116.8

Chisq= 16.49 on 1 degrees of freedom, p= 4.9e-05

Number of Newton-Raphson Iterations: 4

n= 42

Modelo exponencial TFA e RP

Os valores estimados dos coeficientes a partir do exemplo acima exponencial.

$$\text{Modelo TFA : } \hat{\beta}_1 = 1,53$$

$$\text{Modelo RP : } \hat{b}_1 = -1,53$$

Também temos

$$RP(x = 1 \text{ vs } x = 0) = \exp\{\hat{b}_1\} = \exp\{-\hat{\beta}_1\} = \frac{1}{\hat{\gamma}}.$$

Propriedade do modelo Weibull

Resultado

A suposição TFA é válido se e se somente se a suposição RP é válido.

Demonstração.

Para provar considere o exemplo $RR(x=1, \text{ vs } x=0)$

$$[\implies] \quad \gamma = \exp\{\beta_1\}$$

assume que γ é constante $\implies \beta_1$ é constante

$$RR = \exp\{b_1\} = \exp\{-\alpha\beta_1\} \implies RR \text{ é constante}$$



$$[\impliedby] \quad RR = \exp\{b_1\}$$

assume que RP é constante $\implies b_1$ é constante

$$\gamma = \exp\{\beta_1\} = \exp\{-b_1/\alpha\} \implies \gamma \text{ é constante}$$

Possíveis plots

Possíveis resultados para as plotagens de $\log(-\log(\widehat{S}(t)))$ contra $\log(t)$:



- 1 Weibull (ou exponencial, se $\alpha = 1$), a suposição RP e TFA é válido.
- 2 Não Weibull, RP e não TFA.
- 3 Não Weibull, não RP e não TFA.
- 4 Weibull, não RP e não TFA (α não fixo).

Teste de hipóteses

Suponha que se deseja testar $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ contra a alternativa $H_0 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Usualmente são utilizadas três estatísticas de teste para testar H_0 .

Teste de Wald

$$W = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top I_0(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

Razão de Verossimilhança

$$TRV = -2 \log \left[\frac{L(\hat{\boldsymbol{\theta}}_0)}{L(\hat{\boldsymbol{\theta}})} \right]$$

Score

$$S = U^\top(\boldsymbol{\theta}_0)[I_0(\boldsymbol{\theta}_0)]^{-1}U(\boldsymbol{\theta}_0)$$

As três estatísticas sob H_0 , possuem distribuição assintótica qui-quadrado com p graus de liberdade

Teste de hipóteses

Suponha $\theta = (\theta_1, \theta_2)$, e estamos interessados em testar

$$H_0 : \theta_1 = \theta_{10} \text{ vs } H_1 : \theta_1 \neq \theta_{10}$$

. Sob certas condições de regularidade, os seguintes estatísticas podem ser usados para testar H_0

Teste de Wald:

Suponha a partição de $\hat{\theta}$ e C

$$\hat{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \text{ e } C = \begin{bmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{bmatrix}.$$

Então sob H_0

$$W = \left(\hat{\theta}_1 - \theta_{10} \right)^T C_{11}^{-1} \left(\hat{\theta}_1 - \theta_{10} \right) \stackrel{ass}{\sim} \chi^2_{(k)}$$

onde k é a dimensão de θ_1 .

Teste de hipóteses

Teste escore

O teste escore é baseado na normalidade assintótica do vetor escore, isto é,

$$U(\boldsymbol{\theta}) \stackrel{ass}{\sim} N(\mathbf{0}, I(\boldsymbol{\theta}))$$

Considere a partição do vetor escore $U(\boldsymbol{\theta}) = (U_1(\boldsymbol{\theta}), U_2(\boldsymbol{\theta}))$ e seja $\tilde{\boldsymbol{\theta}}_2$ o EMV de $\boldsymbol{\theta}_2$ sob $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$. Então sob H_0

$$S = U_1^\top C_{11} U_1 \stackrel{ass}{\sim} \chi_{(k)}^2,$$

onde $U_1(\boldsymbol{\theta})$ e C_{11} é avaliado em $C = (\boldsymbol{\theta}_{10}, \tilde{\boldsymbol{\theta}}_2)$. Rejeita-se H_0 se $S > \chi_{1-\alpha}^2$

Teste da razão de verossimilhança

Sob H_0 a estatística

$$RV = -2 \left(\ell(\tilde{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}) \right) \stackrel{ass}{\sim} \chi_{(k)}^2,$$

Teste de hipóteses

- (i) **Teste de Wald:** é geralmente o mais usado para testar hipóteses relativas a um único parâmetro, isto é,

$$H_0 : \theta_j = \theta_{0j}$$

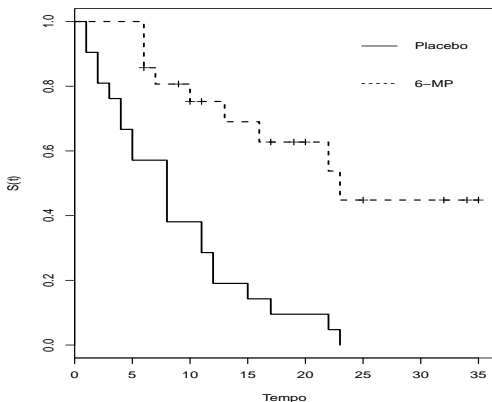
A estatística de teste:

$$W = \left(\frac{\hat{\theta}_j - \theta_{0j}}{\sqrt{\widehat{Var}(\hat{\theta}_j)}} \right)^2 \underset{H_0}{\sim} \chi^2_{(1)},$$

Valores de $W > \chi^2_{(1,1-\alpha)}$ indicam a rejeição de H_0 .

Exemplo: Leucemia

- Tempos de remissão (em meses) do câncer (leucemia aguda) entre crianças tratadas com: 6-MP ou placebo



Exemplo: Leucemia

Regressão exponencial	Regressão Weibull
$\hat{\beta}_0 = 2,159$	$\hat{\beta}_0 = 2,248$
$\hat{\beta}_1 = 1,527$	$\hat{\beta}_1 = 1,267$
$\alpha = 1$	$\hat{\alpha} = 1.366$
log-vero=-108,5240	log-vero=-106,5795

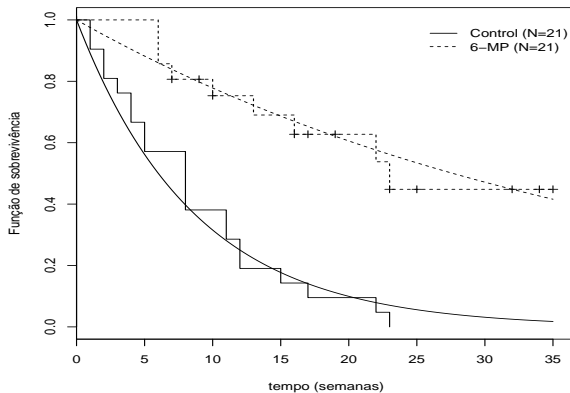
- Testando-se as hipóteses : $H_0 : \alpha = 1$ vs $H_1 : \alpha \neq 1$, obten-se para o teste de razão de verossimilhança o valor de 3.889 (p-valor=0.048, gl=1).
- Este resultado fornece indicações favoráveis para o modelo de regressão exponencial.

Exemplo: Leucemia

- Teste de hipóteses: $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$.
- O TRV resultou em 16,48521 e p – valor $< 0,0001$, conclui-se, portanto, pela rejeição de H_0 .
- A função de sobrevivência obtida pelo modelo de regressão exponencial ajustado do exemplo é dado por

$$\widehat{S}(t) = \exp \left\{ -\frac{t}{\exp(2,159 + 1,527x)} \right\}, t > 0$$

Exemplo: Leucemia



Estudo de adequação do modelo

Essencialmente trabalha-se com os resíduos.

Resíduos Cox-Snell, 1963

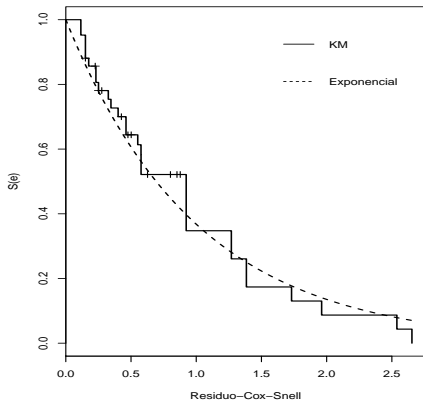
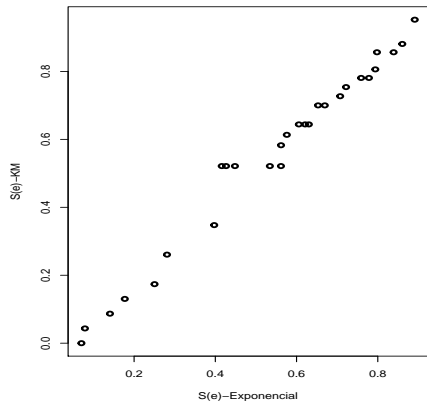
Forma do resíduo tem a seguinte forma:

$$\hat{e}_i = \hat{H}(t_i | \mathbf{x}_i),$$

em que $H(t_i | \mathbf{x}_i) = -\log(S(t_i | \mathbf{x}_i))$ é a função de risco acumulado de T .

- se o modelo for adequado os \hat{e}_i têm distribuição exponencial padrão ($\lambda = 1$).
- Cuidado: existem evidências de situações em que o modelo não é adequado e os \hat{e}_i têm a forma de uma exponencial padrão.
- Esses resíduos, juntamente como residuo de martingal, são melhores indicados para indicar a forma funcional de uma determinada covariável no modelo.

Estudo de adequabilidade: Dados de Leucemia



Estudo de adequação do modelo

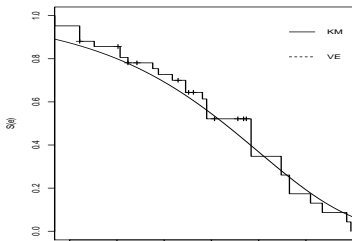
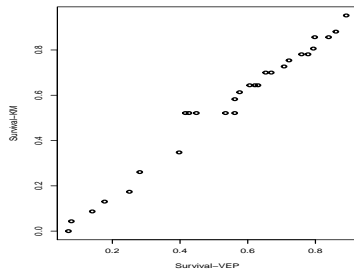
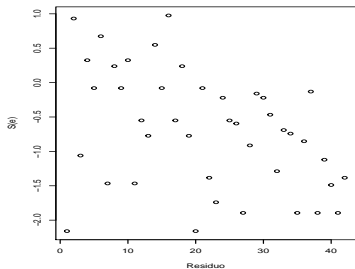
Resíduos padronizados

O resíduo padronizado é definido por

$$\hat{\epsilon}_i = \frac{Y - \mathbf{x}'\hat{\beta}}{\hat{\sigma}}, i = 1, \dots, n,$$

- Os resíduos $\hat{\epsilon}_i$ são estimativas dos erros que vêm de uma população homogênea com distribuição valor extremo padrão (VEP).
- Desta forma as probabilidades de sobrevivência $S(\hat{\epsilon}_i)$ obtidas para esse resíduo pelo estimador de K-M versus $\hat{\epsilon}_i$ devem se aproximadamente uma linha reta.

Estudo de adequabilidade: Dados de Leucemia



Estudo de adequação do modelo

Resíduos Martingal

O resíduo Martingal é definido por

$$\hat{m}_i = \delta_i - \hat{e}_i = \delta_i - \hat{H}(t_i | \mathbf{x}_i),$$

sendo δ_i o indicador de censura e \hat{e}_i é o resíduo de Cox-Snell.

- São usados para indicar a forma funcional de uma determinada covariável no modelo.
- Por exemplo, se uma curva suavizada do diagrama de dispersão de (x_{1i}, \hat{m}_i) , em que x_1 é uma covariável contínua, for linear, nenhuma transformação em x_1 é necessária.
- Outros comportamentos dessa curva pode indicar, por exemplo, a inclusão de um termo quadrático da covariável no modelo ou sugerir alguma transformação da mesma.

Estudo de adequação do modelo

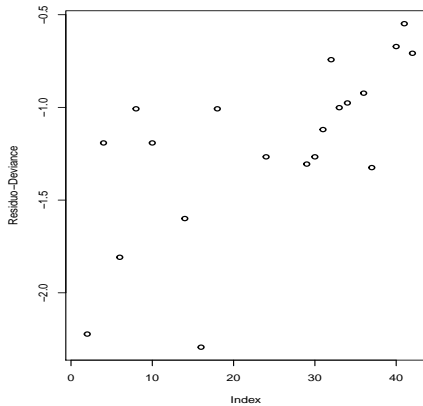
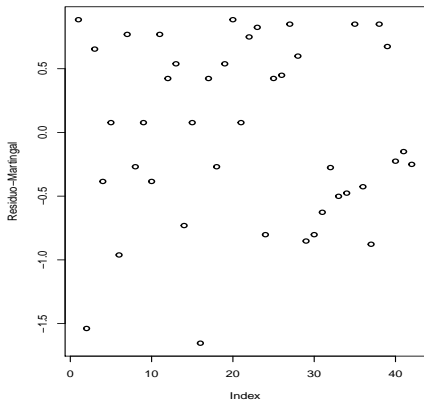
Resíduos Deviance

O resíduo deviance é definido por

$$\hat{d}_i = \text{sign}(\hat{m}_i) [-2(\hat{m}_i + \delta_i \log(\delta_i - \hat{m}_i))]^{1/2} .,$$

- Esses resíduos, são uma tentativa de tornar os resíduos Martingal mais simétricos em torno a zero,
- Se o modelo for apropriado, esses resíduos devem apresentar um comportamento aleatório em torno a zero.
- Facilitam a detecção de pontos atípicos.

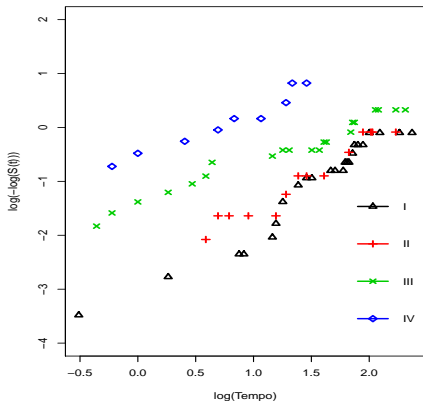
Estudo de adequabilidade: Dados de Leucemia



Exemplo: Câncer de laringe

- Neste exemplo os dados considerados referem-se a um estudo, descrito em Klein e Moeschberger (1997), realizado com 90 pacientes do sexo masculino diagnosticados no período de 1970-1978 com câncer de laringe e que foram acompanhados até 01/01/1983.
- Para cada paciente foram registrados, no diagnóstico, a idade (em anos) e o estágio da doença (I = tumor primário, II = envolvimento de nódulos, III = metástases e IV = combinações dos 3 estágios anteriores) bem como seus respectivos tempos de sobrevivência ou censura (em anos).
- Os estágios encontram-se ordenados pelo grau de seriedade da doença (menos sério para mais sério).

Exemplo: Câncer de laringe



- Retas aproximadamente paralelas \implies RP.
- Retas com intercepto \implies TFA Weibull.

Exemplo: Câncer de laringe

- Consideremos o modelo TFA Weibull

$$\log t = \beta_1 + \beta_2 I_1 + \beta_3 I_2 + \beta_4 I_3 + \beta_5 x_2 + \beta_6 (I_1 x_2) + \beta_7 (I_2 x_2) + \beta_8 (I_3 x_2) + \sigma \varepsilon,$$

onde

- t tempo de sobrevivência ou censura do paciente,
- Estágio-(I,II,III,IV)

$$I_1 = \begin{cases} 1, & \text{se II} \\ 0, & \text{cc} \end{cases}, \quad I_2 = \begin{cases} 1, & \text{se III} \\ 0, & \text{cc} \end{cases}, \quad \text{e } I_3 = \begin{cases} 1, & \text{se IV} \\ 0, & \text{cc} \end{cases},$$

- x_2 : idade dos pacientes,
- Os β_i 's $i = 1, \dots, 8$ os coeficientes de regressão.
- σ é o parâmetro de escala e
- ε é o termo aleatório com distribuição valor extremo padrão.

Exemplo: Câncer de laringe

Modelos	Covariáveis	Estimativas	Log-verossimilhança
1	nenhuma	—	$\ell_1 = -151,1101$
2	x_1 : estágio	$\hat{\beta}_1 = 2.36914628$	$\ell_2 = -142.4096$
		II $\hat{\beta}_2 = -0.086756$	
		III $\hat{\beta}_3 = -0.556629$	
		IV $\hat{\beta}_4 = -1.5785924$	
3	x_1 : estágio	$\hat{\beta}_1 = 3.52876$	$\ell_3 = -141.4234$
		II $\hat{\beta}_2 = -0.14770$	
		III $\hat{\beta}_3 = -0.58656$	
		IV $\hat{\beta}_4 = -1.54408$	
	x_2 :idade $\hat{\beta}_5 = -0.01746$		
4	x_1 : estágio	$\hat{\beta}_1 = 2.34096$	$\ell_3 = -138.5119$
		II $\hat{\beta}_2 = 6.67742$	
		III $\hat{\beta}_3 = 0.31200$	
		IV $\hat{\beta}_4 = -1.05556$	
	x_2 :idade $\hat{\beta}_5 = 0.00023$		
	$x_1 \times x_2$:	II \times Idade $\hat{\beta}_6 = -0.10112$	
		III \times Idade $\hat{\beta}_7 = -0.01330$	
		IV \times Idade $\hat{\beta}_8 = -0.00750$	

Exemplo: Câncer de laringe

Comp	Teste	GL	RV	p-valor
(1 vs 2)	$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$	3	17.401	0.0006
(1 vs 3)	$H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$	4	19,373	0.0006
(1 vs 4)	$H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$	7	25.196	0.0007
(2 vs 3)	$H_0 : \beta_5 = 0$	1	1.972	0.1602
(2 vs 4)	$H_0 : \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$	4	7.795	0.0994
(3 vs 4)	$H_0 : \beta_6 = \beta_7 = \beta_8 = 0$	3	5.823	0.1206

Exemplo: Câncer de laringe

- A partir dos resultados do teste da razão de verossimilhanças associado à interação entre estágio e idade, o resultado $RV = 7.79537$ (p -valor = 0,0994, g.l. = 4)
- Por outro lado, os resultados dos testes locais dessa interação, apresentados na tabela abaixo mostram evidências de que pelo menos um dos β s associados à referida interação difere significativamente de zero, no caso β_6 com p – valor = 0,029.

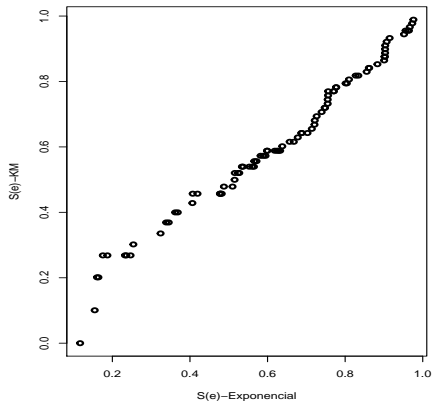
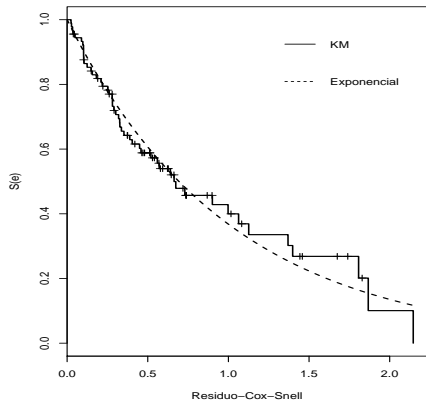
parâmetro	estimativa	erro padrão	z	p-valor
$\beta_6 = idade \times II$	-0.101117	0.0462	-2.1886	0.0286
$\beta_7 = idade \times III$	-0.013300	0.0321	-0.4139	0.6789
$\beta_8 = idade \times IV$	-0.007498	0.0307	-0.2445	0.8068

Exemplo: Câncer de laringe

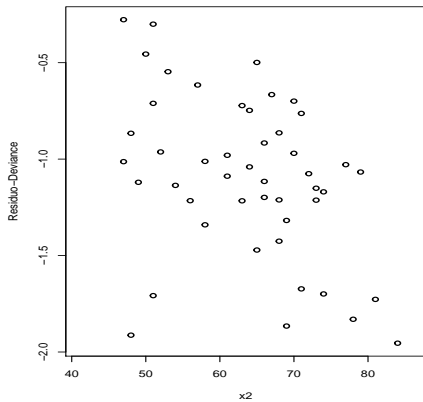
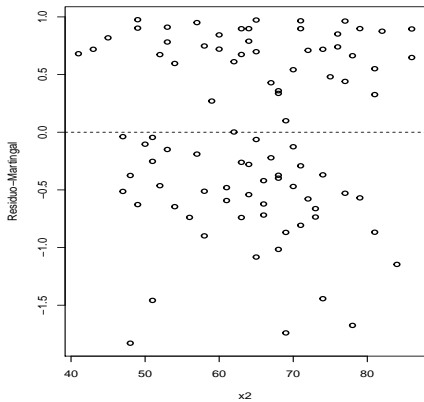
- Da análise considerou-se o modelo com a presença da interação. As estimativas obtidas para este modelo foram:

Covariável	Estimativa	Erro-Padrão	z	Valor-p
Intercepto	2.34096	1.4732	1.5891	0.1120
Estágio II	6.67742	3.2579	2.0496	0.0404
Estágio III	0.31200	2.1186	0.1473	0.8829
Estágio IV	-1.05556	2.0774	-0.5081	0.6114
Idade	0.00023	0.0222	0.0103	0.9918
Estágio II*idade	-0.101117	0.0462	-2.1886	0.0286
Estágio III*idade	-0.013300	0.0321	-0.4139	0.6789
Estágio IV*idade	-0.007498	0.0307	-0.2445	0.8068
$\log(\sigma)$	-0.145388	0.1218	-1.1934	0.2327

Exemplo: Câncer de laringe



Exemplo: Câncer de laringe



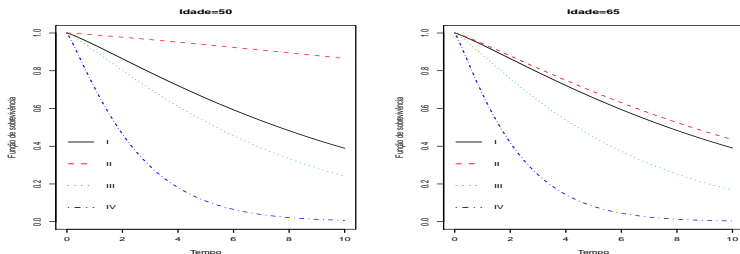
Exemplo: Câncer de laringe

As funções de sobrevivência para o modelo ajustado:

$$\widehat{S}(t|\mathbf{x}) = \begin{cases} \exp \left\{ - \left(\frac{t}{\exp\{\hat{\beta}_1 + \hat{\beta}_5 \text{Idade}\}} \right)^{\hat{\alpha}} \right\}, & \text{se estágio I,} \\ \exp \left\{ - \left(\frac{t}{\exp\{\hat{\beta}_1 + \hat{\beta}_2 + (\hat{\beta}_5 + \hat{\beta}_6) \text{Idade}\}} \right)^{\hat{\alpha}} \right\}, & \text{se estágio II,} \\ \exp \left\{ - \left(\frac{t}{\exp\{\hat{\beta}_1 + \hat{\beta}_3 + (\hat{\beta}_5 + \hat{\beta}_7) \text{Idade}\}} \right)^{\hat{\alpha}} \right\}, & \text{se estágio III} \\ \exp \left\{ - \left(\frac{t}{\exp\{\hat{\beta}_1 + \hat{\beta}_4 + (\hat{\beta}_5 + \hat{\beta}_8) \text{Idade}\}} \right)^{\hat{\alpha}} \right\}, & \text{se estágio IV} \end{cases}$$

Exemplo: Câncer de laringe

- Curvas de sobrevivência estimadas para pacientes com idades de 50 e 65 anos, em cada um dos 4 estágios da doença.



- pode-se observar que as curvas de sobrevivência estimadas para os estágios I, III e IV não apresentam diferenças muito acentuadas quando comparadas para as idades de 50 e 65 anos

Exemplo: Câncer de laringe

- Assim, por exemplo, para os pacientes i e l em que ambos encontram-se no estágio II da doença, mas um deles apresenta idade de 50 anos e outro de 65 anos, tem-se que o fator de aceleração, γ , entre eles é:

$$\gamma_{II}(x_2 = 50, \text{ vs, } x_2 = 65) = \frac{\exp\{\hat{\beta}_1 + \hat{\beta}_2 + (\hat{\beta}_5 + \hat{\beta}_6) \times 50\}}{\exp\{\hat{\beta}_1 + \hat{\beta}_2 + (\hat{\beta}_5 + \hat{\beta}_6) \times 65\}} = 1,7,$$

- o que significa que o tempo mediano de sobrevivência de pacientes com 50 anos de idade e no estágio II da doença, é de aproximadamente 2 vezes maior do que o tempo mediano de sobrevivência de pacientes com 65 anos e no mesmo estágio da doença.

Exemplo: Câncer de laringe

- Por outro lado, tem-se, por exemplo, para os pacientes j e k em que ambos têm 50 anos de idade mas um deles encontra-se no estágio IV da doença e outro no estágio III, o fator de aceleração entre eles é de:

$$\gamma(\text{III vs IV}) = \frac{\exp\{\hat{\beta}_1 + \hat{\beta}_3 + (\hat{\beta}_5 + \hat{\beta}_7) \times 50\}}{\exp\{\hat{\beta}_1 + \hat{\beta}_2 + (\hat{\beta}_5 + \hat{\beta}_8) \times 50\}} = 5.3,$$

- Desse modo, tem-se que o tempo de sobrevivência mediano de pacientes com 50 anos de idade e no estágio III da doença é de 5,3 vezes maior do que o tempo de sobrevivência mediano de pacientes também com 50 anos de idade mas que encontram-se no estágio IV da doença.
- É possível determinar o erro padrão do fator de aceleração?