Análise de Sobrevivência

Modelagem paramétrica
Revisão

Valeska Andreozzi[1]

valeska.andreozzi@fc.ul.pt

&

Marilia Sá Carvalho[2]

cavalho@fiocruz.br

[1]Centro de Estatística e Aplicações da Universidade de Lisboa, Portugal

[2]Escola Nacional de Saúde Pública e Programa de Computação Científica da Fundação Oswaldo Cruz, Brasil

Julho, 2008

Revisando

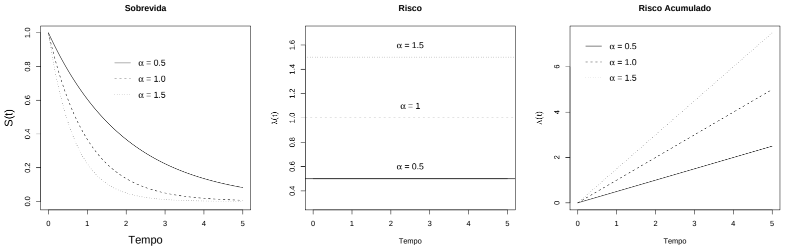
- Estimação paramétrica
- Regressão paramétrica

Distribuição Exponencial

$$
\begin{aligned}
T &\sim Exp(\alpha) \\
f(t) = \alpha \exp(-\alpha t), &\qquad \alpha > 0 \\
E(t) &= \frac{1}{\alpha} \\
S(t) &= \exp(-\alpha t) \\
\lambda(t) = \frac{f(t)}{S(t)} &= \alpha = \text{constante} \\
\Lambda(t) = -\ln S(t) &= \alpha t
\end{aligned}
$$

Algumas exponenciais

Função de sobrevida, de risco e de risco acumulado para a distribuição exponencial considerando diferentes valores de α



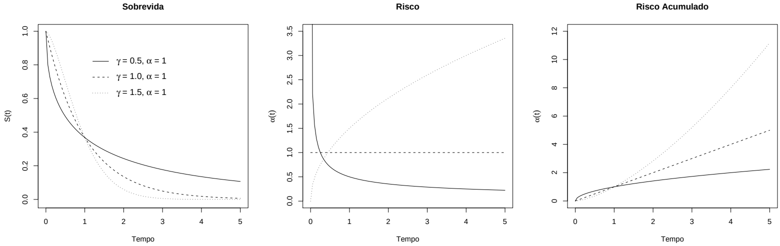
A distribuição exponencial é conhecida como distribuição exponencial padrão quando $\alpha = 1$.

Distribuição Weibull

$$
\begin{aligned}
T &\sim Weibull(\alpha, \gamma) \\
f(t) &= \gamma\alpha^{\gamma} t^{\gamma-1} \exp(-(\alpha t)^{\gamma}) \\
S(t) &= \exp(-(\alpha t)^{\gamma}) \\
\lambda(t) &= \gamma\alpha^{\gamma} t^{\gamma-1} \\
\Lambda(t) &= -\ln S(t) = (\alpha t)^{\gamma-1}
\end{aligned}
$$

Algumas Weibull

Função de sobrevida, de risco e de risco acumulado com parâmetro escala $\alpha = 1$ e diferentes valores do parâmetro de forma γ



Exemplo no R

```
> leite <- read.table("leite.txt", header=T)
> leite

   crianca tempo status
1        1     6      1
2        2    12      1
3        3    10      1
4        4     3      1
5        5     5      1
6        6     1      1
7        7     6      1
8        8     8      1
9        9     1      1
10      10     5      1
11      11     2      1
12      12     2      1
13      13     5      1
14      14     8      1
15      15     1      1
```

Exemplo no R - Exponential

```
> modeloexp <- survreg(Surv(leite$tempo,leite$status)~1,
        data=leite, dist="exponential")
> summary(modeloexp)
Call:
survreg(formula = Surv(leite$tempo, leite$status) ~ 1, data = leite,
    dist = "exponential")
            Value Std. Error    z        p
(Intercept)  1.61      0.258 6.23 4.57e-10
Scale fixed at 1
Exponential distribution
Loglik(model)= -39.1   Loglik(intercept only)= -39.1
Number of Newton-Raphson Iterations: 4
n= 15
```

$T \sim Exp(\alpha)$

$\alpha = \exp(-1.61) = 0.2$

$\lambda(t) = 0.2$

$S(t) = exp(-0.2t)$

$E(t) = 1/0.2 = 5.002 \rightarrow$ `mean(leite$time)=5`

Exemplo no R - Weibull

```
> modeloweib <- survreg(Surv(leite$tempo,leite$status)~1,
      data=leite, dist="weib")
> summary(modeloweib)
Call:
survreg(formula = Surv(leite$tempo, leite$status) ~ 1, data = leite,
    dist = "weib")
             Value Std. Error     z        p
(Intercept)  1.713      0.180  9.54 1.38e-21
Log(scale)  -0.415      0.209 -1.99 4.70e-02
Scale= 0.66
Weibull distribution
Loglik(model)= -37.5   Loglik(intercept only)= -37.5
Number of Newton-Raphson Iterations: 6
n= 15
```

$T \sim Weibull(\alpha, \gamma)$

$\alpha = \exp(-1.713) = 0.18$ e $\gamma = 1/0.66 = 1.51$

$S(t) = \exp(-(\alpha t)^{\gamma}) = \exp(-(0.18t)^{1.51})$

$\lambda(t) = \gamma \alpha^{\gamma} t^{\gamma - 1} = 1.51 \times 0.18^{1.51} t^{1.51-1} = 0.11 t^{0.51}$

Regressão Paramétrica

- Assumimos que o parâmetro da distribuição depende de covariáveis segundo uma função
- Exemplo: $\alpha(\boldsymbol{x}) = \exp(\boldsymbol{x\beta})$
- Regressão Exponencial:

$$S(t|\boldsymbol{x}) = \exp(-\alpha(\boldsymbol{x})t) = \exp(-\exp(\boldsymbol{x\beta})t)$$

$$\lambda(t|\boldsymbol{x}) = \alpha(\boldsymbol{x}) = \exp(\boldsymbol{x\beta})$$

- Regressão Weibull:

$$S(t) = \exp(-(\alpha(\boldsymbol{x})t)^\gamma) = \exp(-(\exp(\boldsymbol{x\beta})t)^\gamma)$$

$$\lambda(t) = \gamma\alpha(\boldsymbol{x})^\gamma t^{\gamma-1} = \gamma(\exp(\boldsymbol{x\beta}))^\gamma t^{\gamma-1}$$

Exemplo regressão paramétrica no R

```
> dialise <- read.csv("dialise.csv",header=T)
> regexp<-survreg(formula=Surv(tempo,status) ~ idade,
        data=dialise, dist="exponential")
> summary(regexp)

Call:
survreg(formula = Surv(tempo, status) ~ idade, data = dialise,
    dist = "exponential")
             Value Std. Error     z         p
(Intercept)  6.136    0.10585  58.0  0.00e+00
idade       -0.037    0.00174 -21.3 1.36e-100
Scale fixed at 1
```

$$\lambda(t|idade) = \exp(-6,14 + 0,037 idade)$$
$$\frac{\lambda(t|x_1{=}70)}{\lambda(t|x_1{=}30)} = \frac{\exp(\beta_0{+}30\beta_1)}{\exp(\beta_0{+}70\beta_1)} = \frac{0{,}02886}{0{,}00657} = 4,39$$

```
> regweib<-survreg(formula=Surv(tempo,status) ~ idade,
      data=dialise, dist="weibull")
> summary(regweib)

Call:
survreg(formula = Surv(tempo, status) ~ idade, data = dialise,
    dist = "weibull")
              Value Std. Error      z        p
(Intercept)  6.7512    0.14693  45.95 0.00e+00
idade       -0.0436    0.00224 -19.50 1.12e-84
Log(scale)   0.1987    0.02083   9.54 1.49e-21
Scale= 1.22
```

$\gamma = 1/1.22 = 0.82$
$\lambda(t) = 0.82 \times t^{-0.18} \times (\exp(-6.75 + 0.04 idade))^{0.82}$
Para cada ano a mais de idade, o risco relativo de morrer aumente 4% $(\exp(0,04) = 1,0408)$.

Inferência

- Teste de Wald $z = \frac{\hat{\beta}}{EP(\hat{\beta})}$
  - H_0: parâmetro β da regressão é igual a zero
- Análise da função desvio
  - estatística global do ajuste do modelo $D = 2(l_{modelo} - l_{nulo})$
  - comparar modelos aninhados $D = 2(l_{maior} - l_{menor})$