MAE0514 - Distribuição Weibull

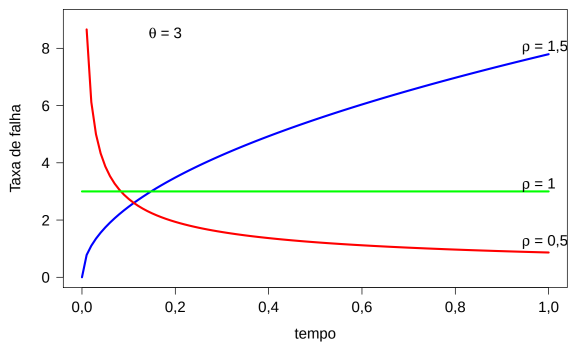
Antonio Carlos Pedroso de Lima

IME-USP

Distribuição Weibull - Caracterização

Caracterização ($\theta > 0$, $\rho > 0$)

$$T \sim \text{Weibull}(\rho, \theta) \Rightarrow \begin{cases} f_T(t) = \rho\theta(\theta t)^{\rho-1} e^{-(\theta t)^\rho}, \\ S_T(t) = e^{-(\theta t)^\rho}, \\ \alpha_T(t) = \rho\theta(\theta t)^{\rho-1}, \; t \geq 0. \end{cases}$$



Distribuição Weibull - Estimação

Supondo $\{(Z_i, \delta_i),\ i = 1, \ldots, n\}$ com

$$Z_i = \min\{T_i, C_i\} \quad \text{e} \quad \delta_i = I(T_i \le C_i),$$

com T_i ~Weibull(ρ, θ) e C_i v.a. com distribuição qualquer.

Suposições

- Censura independente
- Censura não informativa

Equações de estimação: $\psi = (\rho, \theta)$

$$\frac{\partial \ell(\psi)}{\partial \rho} = 0 \qquad \text{e} \qquad \frac{\partial \ell(\psi)}{\partial \theta} = 0$$

$\ell(\psi)$: log-verossimilhança Weibull

O sistema acima não tem solução explícita

Método iterativo de Newton-Raphson

Vetor escore:

$$\mathbf{U}(\psi) = \begin{pmatrix} \frac{\partial \ell(\psi)}{\partial \rho} \\ \frac{\partial \ell(\psi)}{\partial \theta} \end{pmatrix}$$

Note que se $\widehat{\psi}$ é o EMV de $\psi \Rightarrow \mathbf{U}(\widehat{\psi}) = \mathbf{0}$.

Expansão de Taylor de primeira ordem de $\mathbf{U}(\widehat{\psi})$ em torno de ψ,

$$\mathbf{U}(\widehat{\psi}) \approx \mathbf{U}(\psi) - \mathbf{I}(\psi)(\widehat{\psi} - \psi), \tag{1}$$

com

$$\mathbf{I}(\psi) = \begin{pmatrix} -\frac{\partial^2 \ell(\psi)}{\partial \rho^2} & -\frac{\partial^2 \ell(\psi)}{\partial \rho \partial \theta} \\ -\frac{\partial^2 \ell(\psi)}{\partial \theta \partial \rho} & -\frac{\partial^2 \ell(\psi)}{\partial \theta^2} \end{pmatrix}$$

matriz de informação *observada*.

Método iterativo de Newton-Raphson (cont.)

Como $\mathbf{U}(\widehat{\psi}) = \mathbf{0}$, isolando $\widehat{\psi}$ em (1) vem

$$\widehat{\psi} \approx \psi + \mathbf{I}^{-1}(\psi)\mathbf{U}(\psi)$$

Substituindo $\widehat{\psi}$ e ψ por ψ_{k+1} e ψ_k, respectivamente, obtemos

Núcleo do método iterativo de Newton-Raphson

$$\psi_{k+1} = \psi_k + \mathbf{I}^{-1}(\psi_k)\mathbf{U}(\psi_k), \quad k = 0, 1, 2, \ldots$$

- ψ_0 é o *chute* inicial.
- ψ_k é atualizado até que $||\psi_{k+1} - \psi_k|| < \varepsilon$, em que ε é pequeno (por exemplo, 0,0001).

Estimação Weibull no R

Para obter um algoritmo mais estável,

$$Y = \log(T)$$

Resultado - Distribuição de Valor extremo

Se $T \sim \textit{Weibull}(\rho, \theta)$, então $Y = \log(T)$ é tal que

$$S_Y(y) = \exp\{-e^{(y-\mu)/\sigma}\}, \quad \sigma = 1/\rho \text{ e } \theta^\rho = \exp(-\mu/\sigma).$$

ou ainda

$$Y = \log(T) = \mu + \sigma\varepsilon,$$

com ε uma v.a. com distribuição de Valor Extremo padrão com

$$S_\varepsilon(e) = \exp\{-\exp(e)\}$$

Estimação Weibull no R

A verossimilhança utilizada no R é

$$L(\mu, \sigma) = \prod_{i=1}^{n} [f_Y(y_i)]^{\delta_i} [S_Y(y_i)]^{1-\delta_i},$$

com $y_i = \min(\log(t_i), \log(c_i))$, valores observados na amostra.

- Função `survreg`
- As estimativas fornecidas referem-se a valores do EMV para μ e σ
- $\sigma = 1$ equivale a ajustar o modelo exponencial.
- Para obter as estimativas originais é necessário fazer as contas da reparametrização à mão.
- Mudando a distribuição de ε obtem-se outros modelos paramétricos (gaussian, logistic, lognormal e loglogistic).

Dados de Leucemia - Modelo Weibull no R

Tempo livre de sintomas para pacientes com leucemia

- Origem: Administração de Placebo ou Droga 6MP
- Evento: Retorno dos sintomas da doença

Dados:

id	tempo	delta
1	6	1
2	6	1
3	6	1
4	7	1
5	10	1
6	13	1
7	16	1
8	22	1
9	23	1
10	6	0
11	9	0

id	tempo	delta
12	10	0
13	11	0
14	17	0
15	19	0
16	20	0
17	25	0
18	32	0
19	32	0
20	34	0
21	35	0

Dados de Leucemia - Modelo Weibull no R (cont.)

Programa R

```
> library(survival)
> leucemia <- read.csv("leucemia.csv", header = T)
> ajuste <- survreg(Surv(tempo,delta)~1,
                    data=leucemia,dist="w")
> summary(ajuste)
```

Saída

```
              Value Std. Error      z        p
(Intercept)  3,519      0,273 12,87 6,28e-38
Log(scale)  -0,303      0,278 -1,09 2,77e-01
Scale= 0,739
Weibull distribution
Loglik(model)= -41,7  Loglik(intercept only)= -41,7
Number of Newton-Raphson Iterations: 5  n= 21
```

Dados de Leucemia - Modelo Weibull no R (cont.)

Obtemos do objeto `ajuste` as estimativas de MV para μ e σ:

$$\widehat{\mu} = 3,519 \quad \text{e} \quad \widehat{\sigma} = 0,739$$

Logo, pelo princípio da invariância do EMV, as estimativas de MV para ρ e θ são

$$\widehat{\rho} = \frac{1}{\widehat{\sigma}} = \frac{1}{0,739} = 1,353$$

$$\widehat{\theta} = \left\{ \exp\left(-\frac{\widehat{\mu}}{\widehat{\sigma}} \right) \right\}^{\widehat{\sigma}} = \left\{ \exp\left(-\frac{3,519}{0,739} \right) \right\}^{0,739} = 0,030$$

Dados de Leucemia - Modelo Weibull no R (cont.)

$$\widehat{\rho} = 1,353 \quad \text{e} \quad \widehat{\theta} = 0,030$$

Estimação da função de Sobrevivência

$$S_\psi(t) = e^{-(\theta t)^\rho}$$

Logo, o EMV é

$$\widehat{S}_\psi(t) = S_{\widehat{\psi}}(t) = e^{-(\widehat{\theta} t)^{\widehat{\rho}}} = e^{(-0,030t)^{1,353}}$$

Portanto, uma estimativa da probabilidade de que um paciente fique livre dos sintomas por mais de $t = 10$ semanas é

$$\widehat{S}_\psi(10) = e^{-(0,030\times 10)^{1,353}} = 0,822$$

Dados de Leucemia - Modelo Weibull no R (cont.)

$$\widehat{\rho} = 1,353 \quad \text{e} \quad \widehat{\theta} = 0,030$$

Estimativa MV do tempo médio de sobrevivência:

$$\widehat{\mu} = \widehat{E}(T) = \frac{\Gamma(1+\frac{1}{\widehat{\rho}})}{\widehat{\theta}} = \frac{\Gamma(1+\frac{1}{1,353})}{0,030} = 53,1 \text{ semanas}$$

Para o tempo mediano,

$$\widehat{S}_\psi(t_{0,5}) = 0,5 \Rightarrow \widehat{t}_{0,5} = \frac{log(2)^{1/\widehat{\rho}}}{\widehat{\theta}}$$

$$\text{Logo} \qquad \widehat{t}_{0,5} = \frac{log(2)^{1/1,353}}{0,030} = 25,4 \text{ semanas}$$

- Note a diferença entre os tempos médio e mediano, típica de distribuições assimétricas