

## Tópicos de análise de sobrevivência

Seja  $T$  uma variável aleatória não negativa representando o tempo de ocorrência de algum evento de interesse. Supomos também que  $T$  seja contínua.

A distribuição de probabilidade da variável aleatória  $T$  pode ser especificada de 3 modos diferentes:

- a) Pela função de sobrevivência
- b) Pela fdp.
- c) Pela função de risco (hazard function)

Definição:

a) A função de sobrevivência é definida por

$$S(t) = P[T > t], \quad t > 0.$$

b) A fdp de  $T$  é definida por

$$f(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P[t \leq T \leq t + \Delta t]}{\Delta t}.$$

c) A função de risco (hazard function) especifica a taxa instantânea de falha em  $T = t$  condicional a sobrevivência no tempo  $T = t$ , definida por

$$\lambda(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P[t \leq T \leq t + \Delta t | T > t]}{\Delta t},$$

de modo que

$$\lambda(t) = \frac{f(t)}{S(t)}.$$

Assim,

$$f(t) = \lambda(t)S(t).$$

Interpretação:

$$\begin{aligned}\lambda(t)\Delta t &\approx P[t \leq T < t + \Delta t | T > t] \\ &= P[\text{Morte em } [t, t + \Delta t) | \text{Sobreviver até o tempo } t].\end{aligned}$$

Observamos que

$$\begin{aligned}\int_0^t \lambda(u) du &= \int_0^t \frac{f(u)}{1 - F(u)} du = -\log[1 - F(u)] \Big|_0^t \\ &= -\log[1 - F(t)] = -\log(S(t)),\end{aligned}$$

onde  $F(t) = 1 - S(t)$  é a f.d.

Temos então a relação:

$$S(t) = \exp\left\{-\int_0^t \lambda(u)du\right\},$$

de modo que

$$f(t) = \lambda(t) \exp\left\{-\int_0^t \lambda(u)du\right\}.$$

## Tipos de censuras

$T_1, \dots, T_N$  uma a.a. da f.d.  $F(\cdot)$ .

Dados de sobrevivência são censurados dada a limitação do tempo de análise ou a limitação de custo. Para dados médicos é comum a perda de pacientes pelas mais diversas causas.

## Censura do tipo I

Sendo  $t_c$  fixado (tempo de censura), em lugar de  $T_1, \dots, T_n$ , observa-se somente

$$Y_i = \begin{cases} T_i, & \text{se } T_i \leq t_c, \\ t_c, & \text{se } T_i > t_c. \end{cases}$$

Note que  $Y$  tem massa positiva  $P[T > t_c] > 0$  em  $Y = t_c$ .

## Censura do tipo II

Seja  $r < n$  fixado e  $T_{(1)}, \dots, T_{(n)}$  as estatísticas de ordem correspondentes a  $T_1, \dots, T_n$ . O experimento termina depois da ocorrência da  $r$ -ésima falha, de modo que a amostra observada é  $T_{(1)}, \dots, T_{(r)}$ . Portanto, observamos

$$Y_i = \begin{cases} T_{(i)}, & i=1, \dots, r, \\ T_{(r)}, & i=r+1, \dots, n. \end{cases}$$

Censuras do tipo 1 e 2 ocorrem em geral na indústria.



## Censura aleatória

$L_1, \dots, L_n$  v.a. iid com f.d.  $H(\cdot)$ .  $L_i$  tempo de censura correspondente a  $T_i$ .  
Observa-se

$$(Y_1, \delta_1), \dots, (Y_n, \delta_n),$$

where

$$Y_i = \min(T_i, L_i)$$

com

$$\delta_i = I(T_i \leq L_i), \quad i = 1, \dots, n.$$

Suposição importante:  $T_i$  e  $L_i$  são independentes.

Controle:

6.2532787, 3.1666603, 3.8574731, 1.5204410, 6.2007783, 13.7953171,  
6.0151210, 8.1982002, 4.9077456, 2.5740268, 0.1553647, 3.6454795,  
1.1961790, 0.5703925, 5.1219489

Tratamento:

0.040216554, 0.145055999, 0.100801694, 0.228209581, 0.484319718,  
0.680433042,  
0.272605316, 0.002199762, 0.149156341, 0.608479438, 0.270311250,  
0.011081122,  
1.116012546, 0.326174937, 0.062484456

Exemplo 1. Tempos de sobrevivência (em semanas) d pacientes com hepatite

controle: 1+, 2+, 3, 3, 3+, 5+, 5+, 16+,16+,16+,16+,16+,16+,16+,16+

Esteróide: 1, 1, 1, 1+, 4+, 5, 7, 8, 10, 10+, 12+, 16+, 16+, 16+

Exemplo 2. Tempo de reincidência de cancer na bexiga

$T_i$ : 3, 5, 6, 7, 8, 9, 10, 10, 10+, 12, 15, 15+, 18, 19, 20, 22, 25, 28, 30, 40, 45+

Modelo mais adequado? Exponencial, Weibull. Log-Normal

Dados vacinação sarampo no Haiti

Modelo tobit-normal, tobit-log-normal. log-potência-normal, excesso de zeros.

Modelo tobit com excesso de zeros.

## 2. Modelos paramétricos

### 1. Distribuição exponencial

Temos:

$$\lambda(t) = \lambda > 0, \quad t > 0,$$

so that

$$\int_0^t \lambda(u) du = \lambda t$$

e

$$S(t) = \exp\left\{-\int_0^t \lambda(u) du\right\} = e^{-\lambda t}.$$

de modo que

$$f(t) = -\frac{dS(t)}{dt} = \lambda e^{-\lambda t}, \quad t > 0.$$

Também

$$E[T] = \frac{1}{\lambda}, \quad Var(T) = \frac{1}{\lambda^2}.$$

Também é possível trabalhar com a parametrização

$$f(t) = \frac{1}{\lambda} e^{-\frac{t}{\lambda}},$$

$t > 0, \lambda > 0$ . Neste caso,

$$E[T] = \lambda, \quad Var[T] = \lambda^2.$$

Para amostra  $t_1, \dots, t_n$  completa, temos o MLE  $\hat{\lambda} = \bar{t}$  e pelo TLC,

$$\sqrt{n}(\hat{\lambda} - \lambda) \rightarrow N(0, \lambda^2).$$

Para construir IC e testes pode-se usar também

$$V = \frac{2 \sum_{i=1}^n T_i}{\lambda} \sim \chi_{2n}^2 \quad (\text{Verificar}).$$



Dadas  $X \sim \text{Exp}(\lambda_1)$  e  $Y \sim \text{Exp}(\lambda_2)$  independentes, temos

$$\begin{aligned} P[X < Y] &= \int_0^{\infty} P[X < Y | X = t] \lambda_1 e^{-\lambda_1 t} dt \\ &= \int_0^{\infty} P[t < Y] \lambda_1 e^{-\lambda_1 t} dt \\ &= \int_0^{\infty} e^{-\lambda_2 t} \lambda_1 e^{-\lambda_1 t} dt \\ &= \lambda_1 \int_0^{\infty} e^{-(\lambda_1 + \lambda_2)t} dt \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2}. \end{aligned}$$

Pode ser generalizado para o caso de mais que duas exponenciais.

## Distribuição de Weibull

É uma generalização da distribuição Weibull com função de sobrevivência dada por

$$S(t) = e^{-(\lambda t)^\alpha}, \quad \alpha > 0, \quad \lambda > 0.$$

Temos

$$\int_0^t \lambda(u) du = (\lambda t)^\alpha, \quad \text{e} \quad \lambda(t) = \alpha \lambda (\lambda t)^{\alpha-1}.$$

Portanto, a fdp é dada por

$$f(t) = \alpha \lambda (\lambda t)^{\alpha-1} \exp^{-(\lambda t)^\alpha}, \quad t > 0.$$

Temos:

1.  $\lambda(t)$  crescente para  $\alpha > 1$
2.  $\lambda(t)$  decrescente para  $\alpha < 1$
3.  $\lambda(t)$  constante para  $\alpha = 1$  (exponencial).

## Distribuição valor-extremo (VE)

$$S(t) = \exp\left[-\exp\left(\frac{x - \mu}{\sigma}\right)\right], \quad x \in R.$$

Note que se  $Y \text{ Weibull}(\alpha, \lambda)$ , então  $X = \log(Y)$  tem distribuição V.E

## Distribuição log-normal

$T \log - Normal(\mu, \sigma^2)$  se  $\log(T) \sim N(\mu, \sigma^2)$ .

Função de Sobrevivência:

$$\begin{aligned} S(t) &= 1 - F(t) = 1 - P[\log(T) < \log(t)] \\ &= 1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right), \end{aligned}$$

onde  $\Phi(\cdot)$  é a fdp da  $N(0, 1)$ .

A fdp é dada por

$$f(t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{(\log(t)-\mu)^2}{2\sigma^2}}.$$

A função de risco ( $f(t)/S(t)$ ) tem forma complicada.  
É necessário trabalhar computacionalmente.

## Modelo Pareto

Para este modelo temos

$$S(t) = \left(\frac{a}{t}\right)^\alpha I_{[a,\infty)}(t), \quad \alpha > 0, \quad a > 0,$$

de modo que

$$f(t) = \frac{\alpha a^\alpha}{t^{\alpha+1}} I_{[a,\infty)}(t),$$

e

$$\lambda(t) = \frac{\alpha}{t} I_{[a,\infty)}(t).$$

Note que a fdp acima pode ser escrita como

$$f(t) = \frac{\alpha a^\alpha}{(t + a)^{\alpha+1}} I_{[0, \infty)}(t),$$

por transformação de variáveis.

Este modelo é usado em Turnbull et al. (1976). Bem como o modelo exponencial.



O modelo Pareto acima pode ser obtido como mistura das distribuições exponencial e gama. Ou seja, sendo

$$X \sim Exp(\theta),$$

com fdp

$$f_{T|\theta}(t) = \theta e^{-\theta t}, \quad \theta > 0$$

e

$$g(\theta) = \frac{\lambda}{\Gamma(p)} (\lambda\theta)^{p-1} e^{-\lambda\theta}, \quad \theta > 0, \lambda > 0, p > 0,$$

temos que

$$f(t) = \int_0^{\infty} \theta e^{-\theta t} \frac{\lambda}{\Gamma(p)} (\lambda\theta)^{p-1} e^{-\lambda\theta}.$$

Pode-se então implementar o modelo Pareto como uma mistura das distribuições exponencial e gama.

## Estimação por máxima verossimilhança

Consideramos primeiramente uma representação geral para a verossimilhança. Suponhamos que as variáveis  $T_i$  e  $L_i$  são contínuas com funções de distribuição  $F_\theta(\cdot)$  e  $H_\eta(\cdot)$ , e que esta última não envolve  $\theta$ . A densidade conjunta de  $(t_i, \delta_i)$ , denotada por  $g_{\delta_i}(t_i)$  será dada por

$$g_{\delta_i}(t_i) = \begin{cases} f_\theta(t_i)[1 - H_\eta(t_i)], & \text{se } \delta_i = 1, \\ h_\eta(t_i)[1 - F_\theta(t_i)], & \text{se } \delta_i = 0. \end{cases}$$

$i = 1, \dots, n.$

Deste modo, segue que a função de verossimilhança pode ser escrita como

$$L(\theta|t, \delta) = \prod_{i=1}^n [f_{\theta}(t_i)]^{\delta_i} [1 - F_{\theta}(t_i)]^{1-\delta_i} [h_{\eta}(t_i)]^{1-\delta_i} [1 - H_{\eta}(t_i)]^{\delta_i}.$$

Igulando o escore a zero temos as equações

$$\frac{\partial \log L(\theta)}{\theta_i} = 0, \quad i = 1, \dots, k.$$

Pelas condições de regularidade do EMV,

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow N(0, I^{-1}(\theta)),$$

e pelo método delta

$$\sqrt{n}(S_{\hat{\theta}} - S_{\theta}) \rightarrow N(0, I^{-1}(\theta) \left( \frac{\partial S_{\theta}(t)}{\partial \theta} \right)^2),$$

onde

$$I(\theta) = \int_0^\infty \left( \frac{\partial \log S_\theta(t)}{\partial \theta} \right)^2 S_\theta(t) h(u) du \\ + \int_0^\infty \left( \frac{\partial \log f_\theta(u)}{\partial \theta} \right)^2 f_\theta(u) [1 - H(u)] du.$$

Assumindo esquema de censuras aleatórias (observar que este esquema inclui censura do tipo I, com  $L_i = t_c$ ), com amostra observada  $(t_1, \delta_1), \dots, (t_n, \delta_n)$ , de modo que a verossimilhança para uma observação (assumindo  $T_i$  independente de  $L_i$ ) é dada por

$$L(\theta|t_i, \delta_i) = \begin{cases} f(t_i), & \text{se } \delta_i = 1, \\ S(t_i), & \text{se } \delta_i = 0, \end{cases}$$

ou seja,

$$L(\theta|t_i, \delta_i) = f^{\delta_i}(t_i)S^{1-\delta_i}(t_i),$$

de modo que para a amostra observada

$$\mathbf{t} = (t_1, \dots, t_n)', \quad \delta = (\delta_1, \dots, \delta_n)',$$

temos a função de verossimilhança

$$\begin{aligned} L(\theta|\mathbf{t}, \delta) &= \prod_{i=1}^n L(t_i, \delta_i) \\ &= \prod_{i \in C} f(t_i) \prod_{i \in C^0} S(t_i), \end{aligned}$$

onde  $C^0$  corresponde a observações censuradas e  $C$  as não censuradas.

Para o modelo log-normal, usamos a transformação  $Y = \log(T) \sim N(\mu, \sigma^2)$  e aí calculamos os estimadores de MV para  $\mu$  e  $\sigma^2$  e usamos o princípio da invariância do EMV.



## O modelo tobit

o modelo tobit foi proposto por Tobin (1956) e corresponde a um modelo censurado a esquerda de zero (ou algum limite de detecção mínimo - LDM).

Formalmente, o modelo tobit pode ser formulado como

$$y_i = \begin{cases} 0 & \text{se } w_i \leq 0, \\ w_i & \text{caso contrário,} \end{cases}$$

onde a variável latente é  $w_i = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i$ , com  $\epsilon_i \sim N(0, \sigma^2)$ ,  $i = 1, \dots, n$ . Consequentemente, denotamos as respostas observadas por  $y_i$ , o valor das  $k$  variáveis explanatórias para a  $i$ -ésima observação por  $\mathbf{x}_i \in R^k$ , os parâmetros de regressão por  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)'$  e o  $i$ -ésimo termo residual por  $\epsilon_i$ .

Temos a função de verossimilhança

$$L_N(\beta, \sigma^2) = \prod_{i=1}^n [1 - \Phi(\frac{1}{\sigma} \mathbf{x}'_i \beta)]^{1-\delta_i} [(\frac{1}{\sigma} \phi(\frac{1}{\sigma} (y_i - \mathbf{x}'_i \beta))]^{\delta_i},$$

com  $\phi$  e  $\Phi$  sendo a fdp e a fda da  $N(0,1)$ .

Podemos ter uma representação geral para a verossimilhança para o modelo tobit

A função de log-verossimilhança para o modelo tobit (para  $T=c$ ) para a situação onde o erro  $\epsilon_i$  segue uma função de distribuição  $F$ , pode ser escrita como

$$\ell(\theta; Y) = \sum_i (1 - I_i) \ln[F(\frac{c - \mu}{\sigma})] + \sum_i I_i \{-\ln(\sigma) + \ln(f(\frac{y_i - \mu}{\sigma}))\}$$

onde  $f = F'$ , e

$$I_i = \begin{cases} 1, & \text{se } y_i > c, \\ 0, & \text{se } y_i \leq c. \end{cases}$$

Verificar no artigo (*Tobit power-normal model, CS-TM*) o uso do modelo potência-normal com o modelo tobit.

O estimador de MV,  $\hat{\theta}$  maximiza a função de verossimilhança  $L(\mathbf{t}, \delta)$  se sob condições de regularidade temos

$$\hat{\theta} \stackrel{a}{\sim} N(\theta, I_F(\theta)),$$

onde  $I_F(\theta)$  é a matriz de informação de Fisher de  $\theta$ , que é consistentemente estimada por

$$i(\hat{\theta}) = -\frac{1}{n} \left\{ \frac{\partial^2 \log L(\mathbf{t}, \delta)}{\partial \theta \partial \theta'} \right\}_{|\hat{\theta}}.$$

Para testar  $H_0 : \theta = \theta_0$ , podemos usar as estatísticas

1. Wald:

$$(\hat{\theta} - \theta_0)' \mathbf{I}_F(\theta_0) (\hat{\theta} - \theta_0) \sim \chi_p^2$$

sob  $H_0$ .



## 2. Razão de verossimilhanças.

$$-2 \log\left(\frac{L(\hat{\theta})}{L(\hat{\theta})}\right) \sim \chi_p^2,$$

under  $H_0$ .

## 3. Escore

$$\frac{\partial}{\partial \theta} \log L(\theta_0)' \mathbf{I}_F(\theta_0) \frac{\partial}{\partial \theta} \log L(\theta_0) \sim \chi_p^2,$$

que não envolve o EMV. Usamos  $L(\mathbf{y}, \delta) = L(\theta)$ .

## *Inferência no modelo exponencial*

$$\begin{aligned} L(\theta) &= \prod_{i \in C} f(t_i) \prod_{i \in C^0} S(t_i) \\ &= \theta^{n_u} \exp\left\{-\theta \sum_{i=1}^n t_i\right\}, \end{aligned}$$

de modo que

$$\log L = n_u \log(\theta) - \theta \sum_{i=1}^n t_i.$$

The EMV é então dado por

$$\hat{\theta} = \frac{n_u}{\sum_{i=1}^n t_i}.$$

Também

$$i(\theta) = \frac{n_u}{\theta^2}.$$

Também, para  $S = \sum_{i=1}^n T_i$  no caso não cesurado,

$$f_S(t) = \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t},$$

de modo que

$$2\theta \sum_{i=1}^n T_i \sim \chi_{2n}^2,$$

de modo que

$$\frac{2n\theta}{\hat{\theta}} \sim \chi_{2n}^2,$$

que pode ser usada na construção de ICs e testes.



Para censura do tipo II,

$$\sum_{i=1}^n Y_i = T_{(1)} + T_{(2)} + \cdots + T_r + (N - r)T_{(r)}$$

$$= nT_{(1)} + (n - 1)[T_{(2)} - T_{(1)}] + \cdots + (N - r + 1)[T_{(r)} - T_{(r-1)}].$$

Temos que  $nT_{(1)}$ ,  $(n - 1)[T_{(2)} - T_{(1)}]$ , ...,  $(N - r + 1)[T_{(r)} - T_{(r-1)}]$  são iids com *Exponencia*( $\theta$ ).

## EMV para Weibull

$$S(t) = e^{-\{(\lambda t)^\alpha\}} = e^{-\gamma t^\alpha},$$

$$f(t) = \gamma \alpha t^{\alpha-1} e^{-\gamma t^\alpha},$$

com  $\gamma = \lambda^\alpha$ .

Sendo  $\theta = (\alpha, \gamma)'$ ,

Verossimilhança:

$$L(\theta|t, \delta) = (\gamma\alpha)^{n_u} \left( \prod_{C^c} t_i^{\alpha-1} \right) e^{-\{\gamma \sum_{i=1}^n t_i^\alpha\}}.$$

Métodos iterativos necessários para maximização.

Condicionais determinam conjunta ( marginais)

Let

$$\mathbf{Z} \sim N(\mu, \Sigma),$$

onde

$$\mathbf{Z} = \begin{pmatrix} X \\ Y \end{pmatrix} \quad \mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

Então

$$X|Y \sim N(0.5 * y, 0.75)$$

$$Y|X \sim N(0.5 * x, 0.75).$$

## Função de distribuição empírica

Ex: Amostra X:1,3,7,10

$$F_n(t) = \begin{cases} 0, & \text{se } t < 1, \\ 1/4, & \text{se } 1 \leq t < 3, \\ 2/4, & \text{se } 3 \leq t < 7, \\ 3/4, & \text{se } 7 \leq t < 10, \\ 1, & \text{se } t \geq 10 \end{cases}$$

$S_n(t) = 1 - F_n(t)$ : função de sobrevivência empírica.

Represente  $S_n$  para a amostra acima!!!!

$$S_n(5) = 1 - F_n(5) = 1 - 2/4 = 2/4.$$

$$S_n(t) = \begin{cases} 1, & \text{se } 0 \leq t < 1, \\ 3/4, & \text{se } 1 \leq t < 3, \\ 2/4, & \text{se } 3 \leq t < 7, \\ 1/4, & \text{se } 7 \leq t < 10, \\ 0, & \text{se } t \geq 10 \end{cases}$$

**Estimador de Kaplan-Meier:**

$$S_n(1) = (1 - 1/4) = 0.75$$

$$S_n(3) = (1 - 1/4)(1 - 1/3) = (0.75)(2/3) = 0.5$$

$$S_n(7) = (1 - 1/4)(1 - 1/3)(1 - 1/2) = (0.75)(2/3)(0.5) = 0.25$$

$$S_n(10) = (1 - 1/4)(1 - 1/3)(1 - 1/2)(1 - 1) = (0.75)(2/3)(0.5)(0) = 0.$$

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	4	1	0.75	0.217	0.4259	1
3	3	1	0.50	0.250	0.1877	1
7	2	1	0.25	0.217	0.0458	1
10	1	1	0.00	NaN	NA	NA



Como para o caso sem censura

$$\sqrt{n}(S_n(t) - S(t))\sqrt{N}(0, S(t)F(t)),$$

de modo que um estimador consistente para  $Var[S_n(t)]$  é dado por

$$\widehat{Var}[S_n(t)] = \frac{S_n(t)F_n(t)}{n}.$$

Assim, da Tabela acima,

$$\widehat{Var}[S_n(1)] = \frac{.75 * .25}{4} = 0.0469,$$

de modo que

$$MSE(S_n(1)) = \sqrt{(0.0469)} = 0.217.$$

No caso geral, com censura

$$Var[\widehat{S_{KM}}(t)] = (S_{KM}(t))^2 \sum_{i; t_i < t} \frac{d_i}{n_i(n_i - d_i)}$$

onde  $d_i$  é o número defalhas em  $t_i$  e  $n_i$  o número de individuos em risco.  
Assim,

$$Var[\widehat{S_{KM}}(1)] = (S_{KM}(1))^2 \frac{1}{4 * 3} = ((0.75)^2)/12 = 0.0469.$$

## Comparação de dois grupos com K-M

```
require(survival)
```

```
ts=c(1,2,3,3,3,5,5,16,16,16,16,16,16,16,16,1,1,1,1,4,5,7,8,10,10,12,16,16,16)
```

```
cs=c(0,0,1,1,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,1,1,1,1,0,0,0,0,0)
```

```
grupos=c(rep(1,15),rep(2,14))
```

```
ekm=survfit(Surv(ts,cs)~grupos)
```

```
summary(ekm)
```

```
plot(ekm,lty=c(2,1),xlab="Ts",ylab="StE")
```

```
survdif(Surv(ts,cs)~grupos,rho=0)
```

## Comparação entre Empírico e paramétrico

### 1. Modelo Exponencial:

$$S_\lambda(t) = e^{-\lambda t}, \quad \frac{\partial S_\lambda(t)}{\partial \lambda} = t e^{-\lambda t},$$

de modo que

$$\sqrt{n}(S_{\hat{\lambda}}(t) - S(t)) \rightarrow N(0, \lambda^2 t^2 e^{-2\lambda t}).$$

Como  $X_1, \dots, X_n$  são iids,  $I(X_i > t)$  são Bernoullis independentes com probabilidade de sucesso  $p = P(X_i > t)$ , de modo que pela LFGN e TLC,

$$\sqrt{n}(S_n(t) - S(t)) \rightarrow N(0, S(t)(1 - S(t))),$$

de modo que

$$e_{E,MV} = \frac{\lambda^2 t^2 e^{-2t\lambda}}{e^{-\lambda t}(1 - e^{-\lambda t})} = \frac{\lambda^2 t^2 e^{-t\lambda}}{(1 - e^{-\lambda t})}$$

Para  $t = 1$ , temos as eficiências

$\lambda$	1	2	3	4	5
$S(t)$	0.58	0.63	0.47	0.30	0.17

Considere  $t_{(1)}, \dots, t_{(k)}$ , os tempos distintos de falhas

$d_i$ : número de falhas em  $t_{(i)}$ ,

$n_i$ : número de indivíduos em risco no instante do  $t_{(i)}$  (-ésima falha);

O estimador de Kaplan-Meier para o caso de não haver empates pode ser escrito como

$$S_{KM}(t_i) = \prod_{i; t_{(i)} < t} \left(1 - \frac{d_i}{n_i}\right) = \prod_{i; t_{(i)} < t} (1 - \hat{q}_i),$$

Comparação de dois grupos pelo KM.

Tratamento: 1+,2+,3,3,3+,5+,5+,16+,16+,16+,16+,16+,16+,16+,16+

Controle: 1,1,1,1+,1+,4,5,7,8,10,10+,12+,16+,16+,16+

Usando o KM no R, temos para o Grupo Controle

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
3.000	13.000	2.000	0.846	0.100	0.671	1.000

## Grupo Tratamento

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	14	3	0.786	0.110	0.598	1.000
5	9	1	0.698	0.128	0.488	0.999
7	8	1	0.611	0.138	0.392	0.952
8	7	1	0.524	0.143	0.306	0.896
10	6	1	0.437	0.144	0.229	0.832



Para comparar os grupos temos a tabela  
 $\text{survdiff}(\text{formula} = \text{Surv}(\text{ts}, \text{cs}) \sim \text{grupos}, \text{rho} = 0)$

	N	Observed	Expected	$(O - E)^2/E$	$(O - E)^2/V$
grupos=1	15	2	4.81	1.64	3.67
grupos=2	14	7	4.19	1.89	3.67

Conclusão: Chisq= 3.7 com 1 degrees of freedom, p= 0.0555

Resultado do enfoque Bayesiano Para comparar dois grupos com modelo Exponencial

Parâmetro *bt*

mean	sd	<i>MCerror</i>	<i>val2.5pc</i>	median	<i>val97.5pc</i>	start	sample
1.795	0.8977	0.03451	0.1785	1.751	3.695	11001	110000

De acordo com o enfoque Bayesiano existe diferença significativa entre os grupos (5%). Note que o teste log-rank não deteta diferença.

Teste "logrank" para comparar funções de sobrevivência  $S_1(t)$  e  $S_2(t)$ .

Considere tempos de falhas conjuntos

$$t_1 < \cdots < t_k.$$

Sejam

$d_j$ : número de falhas em  $t_j$ .

$n_j$ : número de indivíduos em risco em  $t_j$ .

Similarmente define-se  $n_{ij}$  e  $d_{ij}$ , na amostra  $i = 1, 2$ .

Distribuição de  $d_{2j}$ :

$$\frac{\binom{n_{1j}}{d_{1j}} \binom{n_{2j}}{d_{2j}}}{\binom{n_j}{d_j}}.$$

Define-se aproximação normal (sob  $H_0$ ) para distribuição normalizada de  $d_{2j}$ ,  $j = 1, \dots, k$ .

Os resultados a seguir são apresentados em Lima (tese de mestrado).

Considere

$$X \sim \text{Exp}(\mu) \text{ e } L \sim \text{Exp}(\mu).$$

Assim,

$$S(t) = 1 - F(t) = e^{-\theta t} \text{ and } H(t) = e^{-\mu t}.$$

Conforme estabelecido em Lima,

$$\begin{aligned} \text{Var}[\hat{S}_{KM}] &= \int_0^t \frac{f(u)}{[1 - H(u)]S_T(u)^2} du \\ &= \frac{1}{n} e^{-2\theta t} (e^{-(\theta+\mu)t} - 1) \frac{\theta}{\theta + \mu}. \end{aligned}$$

Para o EMV,

$$Var[\hat{S}_{MV}] = \frac{1}{n} I_F^{-1}(\theta) \left\{ \frac{\partial S(t)}{\partial \theta} \right\} = \frac{1}{n} \theta t^2 (\theta + \mu) e^{-2\theta t},$$

de modo que

$$\begin{aligned} EAR_{KM,EM} &= (\theta + \mu)^2 t^2 (e^{(\theta+\mu)t} - 1)^{-1} \\ &= (1 + \rho)^2 (\theta t)^2 (e^{(1+\rho)\theta t} - 1)^{-1}, \end{aligned}$$

onde  $\rho = \mu/\theta$ . Eficiências variam de 0.19 a 0.64.