Tópicos de análise de sobrevivência

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

A distribuição de probabilidade da variaável 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

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

b) A fdp de T é definida por

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

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

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

de modo que

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

Assim,

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

Interpretação:

$$\lambda(t)\Delta t \approx P[t \leq T < t + \Delta t | T > t]$$

$$= P[Morte \quad em \quad [t, t + \Delta t) | Sobreviver \quad ate \quad o \quad tempo \quad t].$$

Observamos que

$$\int_0^t \lambda(u) du = \int_0^t \frac{f(u)}{1 - F(u)} du = -\log[1 - F(u)]|_0^t$$

$$= -\log[1 - F(t)] = -\log(S(t)),$$

onde $F(t) = 1 - S(t)$ é a f.d.

Temos então a relação:

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

de modo que

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

Tipos de censuras

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

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 $T_1, \ldots, T_n$, 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 $T_{(1)}, \ldots, T_{(n)}$ as estatisticas de ordem correspondentes a $T_1, \ldots, T_n$. O experimento termina depois da ocorrencia da r-ésima falha, de modo que a amostra observada é $T_{(1)}, \ldots, T_{(r)}$. Portanto, observamos

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

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

Censura aleatória

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

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

where

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

com

$$\delta_i = I(T_i \leq L_i), \qquad i = 1, \ldots, n.$$

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? Exponecial, 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

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

de modo que

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

Também

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

Também é possível trabalhar com a parametrização

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

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

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

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

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

Para construir IC e testes pode-se usar também

$$V = \frac{2\sum_{i=1}^{n} T_i}{\lambda} \sim \chi^2_{2n} \qquad (Verificar).$$

Dadas $X \sim Exp(\lambda_1)$ e $X \sim Exp(\lambda_2)$ indepedentes, temos

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

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

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

Temos

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

Portanto, a fdp é dada por

$$f(t) = \alpha\lambda(\lambda t)^{\alpha-1}\mathrm{e}xp^{-(\lambda t)^{\alpha}}, \qquad t > 0.$$

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)

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

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

Distribuição log-normal

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

Função de Sobrevivência:

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

onde $\Phi(.)$ é a fdp da $N(0, 1)$.

A fdp é dada por

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

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

Modelo Pareto

Para este modelo temos

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

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 esrita 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

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

e

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

temos que

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

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 continuas com funções de distribuição $F_\theta(.)$ e $H_\eta(.)$, e que esta última não envolve θ. A densidade conjunta de (t_i, δ_i), denotada por $g_{\delta_i}(t_i)$ será dada por

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

$i = 1, \ldots, n.$

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

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

Pelas condições de regularidade do EMV,

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

e pelo método delta

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

onde

$$I(\theta) = \int_0^\infty (\frac{\partial \log S_\theta(t)}{\partial \theta})^2 S_\theta(t) h(u) du$$

$$+ \int_0^\infty (\frac{\partial \log f_\theta(u)}{\partial \theta})^2 f_\theta(u)[1 - H(u)] du.$$

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

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

temos a função de verossimilhança

$$L(\theta|\mathbf{t}, \delta) = \prod_{i=1}^{n} L(t_i, \delta_i)$$

$$= \prod_{i \in C} f(t_i) \prod_{i \in C^0} S(t_i),$$

onde C^0 corresponde a sobservações censuradas e C as não censuradas.

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

O modelo tobit

o modelo tobit foi proposto por Tobin (1956) e corresponde a um modelo censurado a esuqerda de zero (ou algum limite de deteção 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 $w_i = \mathbf{x}_i'\beta + \epsilon_i$, com $\epsilon_i \sim N(0, \sigma^2)$, $i = 1, \ldots, n$. 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 $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$ e o i-ésimo termo residual por ϵ_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 ϕ e Φ 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 ϵ_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 verossimlhança $L(\mathbf{t}, \delta)$ se sob condições de regularidade temos

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

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

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

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

1. Wald:

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

sob H_0.

2. Razão de verossimilhanças.

$$-2\log(\frac{L(\hat{\theta})}{L(\hat{\theta})}) \sim \chi^2_p,$$

under H_0.

3. Escore

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

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

Inferência no modelo exponencial

$$L(\theta) = \prod_{i \in C} f(t_i) \prod_{i \in C^0} S(t_i)$$

$$= \theta^{n_u} \mathrm{e}xp\{-\theta \sum_{i=1}^{n} t_i\},$$

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,

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

de modo que

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

de modo que

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

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

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

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

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

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

Verossimilhança:

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

Métodos iterativos necessários para maximização.

Condicionais determinam conjunta (marginais)
Let

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

onde

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

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

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

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

$S_n(t) = 1 - F_n(t)$: função de sobreivência empírica.
Represente S_n para a amostra acima!!!!
$S_n(5) = 1 - F_n(5) = 1 - 2/4 = 2/4.$

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

Estimador de Kaplan-Meier:

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

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

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

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

Portanto, sendo
n_i: número de indivíduos em risco em (imediatamente antes de) t_i, temos

$$S_n = \prod_{i, t_i < t} (1 - \frac{1}{n_i}).$$

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

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

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

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

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

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

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”)

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

Comparação entre Empírico e paramétrico

1. Modelo Exponencial:

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

de modo que

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

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

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

de modo que

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

Para $t = 1$, temos as eficiências

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

Considere $t_{(1)}, \ldots, t_{(k)}$, 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

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

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

Note que $S(3) = (1 - 2/13) = 1 - 0.154 = .846$..

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

Note que

$S(1) = (1 - 3/14) = 0.786.$
$S(5) = (0.786)(1 - 1/9) = 0.698, \ldots$

Para comparar os grupos temos a tabela
survdiff(formula = Surv(ts, cs) $\sim$ grupos, 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	$MCerror$	$val2.5pc$	median	$val97.5pc$	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 funçõesde 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 $i = 1, 2$.
Define-se uma estatistica baseada na média e variancia de d_{2j}

$$E[d_{2j}] = n_{2j}\frac{d_j}{n_j},$$

$$Var[d_{2j}] = n_{2j}(n_j - n_{2j})d_j(n_j - d_j)n_j^{-2}(n_j - 1)^{-1}.$$

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}, $j = 1, \ldots, k$.

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

Considere

$X \sim Exp(\theta)$ e $L \sim Exp(\mu)$.

Assim,

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

Conforme estabelecido em Lima,

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

Para o EMV,

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

de modo que

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

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

Modelo de Regressão de Cox

Seja

$$\lambda(t, \mathbf{x}) = e^{\mathbf{x}'\beta}\lambda_0(t),$$

onde $\beta = (\beta_1, \ldots, \beta_p)'$ é o vetor de coeficientes de regressão. Note que a função de risco baseline $\lambda_0(t)$, é constante para todos os pacientes.

Para o modelo Weibull, temos

$$\lambda_0(t) = \alpha\lambda^\alpha t^{\alpha-1},$$

que é constante para todos os pacientes. Ou seja o modelo Weibull tem a propriedade dos riscos proporcionais.

Considerando os tempos de falha: $t_1, \ldots, t_k$, temos

$$P[morte \quad em[t_i, t_i + \Delta t | R_i] = \sum_{j \in R_i} \lambda_0(t_i) \mathrm{e}^{\beta' \mathbf{x}_j} \Delta_t.$$

e

$$P[morte \quad de \quad (i) \quad em \quad t_i] = \lambda_0(t_i) \mathrm{e}^{\beta' \mathbf{x}_j}$$

Assim,

$$P[morte \quad de \quad (i) \quad no \quad tempo \quad t_i | morte \quad em \quad t_i]$$

$$= \frac{\mathrm{e}^{\beta' \mathbf{x}_i}}{\sum_{j \in R_i} \mathrm{e}^{\beta' \mathbf{x}_j}}.$$

Temos então a função de verossimilhança

$$L_c(\beta) = \prod_{i=1}^{n} \left(\frac{e^{\beta' \mathbf{x}_j}}{\sum_{j \in R_i} e^{\beta' \mathbf{x}_j}}\right)^{\delta_i},$$

que pode ser considerada como uma verossimilhança usual, com derivada primeira levando ao escore e segunda a observacao observada.

Tomando $l(\beta) = \log(L(\beta))$ e derivando com respeito a β, temos o escore

$$\mathbf{U}(\beta) = \sum_{i=1}^{n} \delta_i [\mathbf{x}_i - \frac{\sum_{j \in R_i} \mathbf{x}_j \mathrm{e}^{\beta' \mathbf{x}_j}}{\sum_{j \in R_i} \mathrm{e}^{\beta' \mathbf{x}_j}}],$$

que deve ser resolvida por métodos numéricos.

Exemplo: Dados sobre 4 pacientes:

PACIENTE	t	Status	x (fumante)
P1	2	1	1
P2	3	1	0
P3	5	0	0
P4	8	1	1

Modelo de Cox:

$$\lambda(t, x) = \lambda_0 \mathrm{e}^{\beta_1 FUMANTE}.$$

Temos então o risco de cada paciente:

PACIENTE	risco
P1	$\lambda_0(t)\mathrm{e}^{\beta_1}$
P2	$\lambda_0(t)\mathrm{e}^{0}$
P3	$\lambda_0(t)\mathrm{e}^{0}$
P4	$\lambda_0(t)\mathrm{e}^{\beta_1}$

Verossimilhança:

$$L(\beta) = [\frac{\lambda_0(t)\mathrm{e}^{\beta_1}}{\lambda_0(t)\mathrm{e}^{\beta_1} + \lambda_0(t)\mathrm{e}^{0} + \lambda_0(t)\mathrm{e}^{0} + \lambda_0(t)\mathrm{e}^{\beta_1}}]$$

$$\times [\frac{\lambda_0(t)\mathrm{e}^{0}}{\lambda_0(t)\mathrm{e}^{0} + \lambda_0(t)\mathrm{e}^{0} + \lambda_0(t)\mathrm{e}^{\beta_1}}]$$

$$\times [\frac{\lambda_0(t)\mathrm{e}^{\beta_1}}{\lambda_0(t)\mathrm{e}^{\beta_1}}].$$

Então temos a função de verossimilhança

$$L(\beta) = \frac{e_1^{\beta}}{2e_1^{\beta} + 2} \times \frac{1}{e_1^{\beta} + 2},$$

Tomando log, obtemos

$$l(\beta) = \log(L(\beta)) = \beta_1 - \log(2e_1^{\beta} + 2) - \log(2 + e^{\beta_1}).$$

Derivando com respeito a beta, obtemos

$$1 - \frac{2e_1^{\beta}}{2e_1^{\beta} + 2} - \frac{e_1^{\beta}}{e_1^{\beta} + 2} = 0,$$

que leva a equação $(y = e_1^{\beta})$,

$$\frac{y}{1+y} + \frac{y}{2+y} = 1,$$

Ou seja,

$$y^2 - 2 = 0, \longrightarrow y = \sqrt{2} \longrightarrow \hat{\beta}_1 = \log(\sqrt{2}).$$

Uma outra configuração:

PACIENTE	t	Status	x (fumante)
P1	2	1	1
P2	3	1	0
P3	5	0	1
P4	8	0	0

Modelo de Cox:

$$\lambda(t, x) = \lambda_0 \mathrm{e}^{\beta_1 FUMANTE}.$$

Verossimilhança:

$$L(\beta) = [\frac{\lambda_0(t)\mathrm{e}^{\beta_1}}{\lambda_0(t)\mathrm{e}^{\beta_1} + \lambda_0(t)\mathrm{e}^{0} + \lambda_0(t)\mathrm{e}^{0} + \lambda_0(t)\mathrm{e}^{\beta_1}}]$$

$$\times[\frac{\lambda_0(t)\mathrm{e}^{0}}{\lambda_0(t)\mathrm{e}^{0} + \lambda_0(t)\mathrm{e}^{0} + \lambda_0(t)\mathrm{e}^{\beta_1}}]$$

$$= \frac{\mathrm{e}^{\beta_1}}{2 + 2\mathrm{e}^{\beta_1}} \times \frac{1}{\mathrm{e}^{\beta_1}}.$$

Assim, sendo $y = \mathrm{e}^{\beta_1}$, temos

$$l = \log(L(y) = y - \log(y+1) - log(y+2),$$

que, derivando com relação a y, leva a equação

$$\frac{1}{1+y} + \frac{1}{2+y} = 1,$$

ou seja,

$$y^2 + y - 1 = 0.$$

ou

$$y = \frac{-1+\sqrt{5}}{2}.$$

G1 (T)	G1 (X)	G2(T)	G2(X)	G1(T)	G1(X)	G2(T)	G2(X)
6	2.31	1	2.8	10+	2.7	8	2.32
6	2.31	1	2.8	10+	2.7	8	2.32
6	4.06	1	5	11+	2.6	8	3.26
6	2.38	2	4.91	17+	2.16	11	3.49
7	4.43	2	4.91	19+	2.05	11	2.12
10	2.96	3	4.01	20	2.01	12	1.50
13	2.88	4	4.36	25+	1.78	12	3.06
16	3.6	4	2.42	32+	2.2	15	2.3
22	2.32	5	3.49	32+	2.53	17	2.95
23	2.57	5	3.97	34+	1.47	22	2.73
6+	3.2	8	3.52	35+	1.45	23	1.97
9+	2.8	8	3.05				

Birnbaum and Saunders (1969) fizeram algumas suposições sobre o processo de fadiga, que são elas:

1. Material é submetido a um ciclo de tensào e força;
2. Crescimento de fiddura que excede cero nível de resistência denotado por w;
3. A sequência de tensão imposta ao material é uma variável aleatória que só depende da fissural atual causada pela tensão no repectivo ciclo;

4. A extensão incremental da fissura representada por X_i resultante da aplicação da i-ésima oscilação de carga é uma variável aleatória que só depende da fissura atual causada pela tensão neste ciclo.
5. a extensão da fissura durante o ciclo j, é

$$Y_{J+1} = X_{jm+1} + \cdots + X_{jm+m}, \quad j = 0, 1, \ldots,$$

onde X_{jm+i} é a extensão da fissura após a i-ésima oscilação de carga do ciclo $(j+1)$.
6. A extensão total da fissura, Y_j devido após n ciclos pela seguinte variável aleatória

$$W_n = \sum_{j=1}^{n} = \sum_{j=1}^{n} Y_j,$$

com fda

$$H_n(w) = P[W_n \leq w], \quad n = 1, 2, \ldots.$$

Definiremos como N o número de ciclos necessários até que ocorra uma falha. A fda da v.a. N pode ser expressa por

$$P[N \leq n] P[\sum_{j=1}^{n} Y_j >< w] = P[W_n > w] = 1 - H_n(w).$$

Considerando Y_j's v.a.i.i.d. pelo uso do TLC temos

$$P[N \leq n] = P[\sum_{j=1}^{n} \frac{Y_j - \mu}{\sigma\sqrt{n}} > \frac{w - n\mu}{\sigma\sqrt{n}}]$$

$$= 1 - P[\sum_{j=1}^{n} \frac{Y_j - \mu}{\sigma\sqrt{n}} \leq \frac{w - n\mu}{\sigma\sqrt{n}}]$$

$$\approx 1 - \Phi(\frac{w - n\mu}{\sigma\sqrt{n}}) = \Phi(\frac{\mu\sqrt{n}}{\sigma} - \frac{w}{\sigma\sqrt{n}}).$$

Assim, substituindo n por t onde T é a v.a. tempo de falha e definindo

$$\alpha = \frac{\sigma}{\sqrt{\mu w}} \quad e \quad \beta = \frac{w}{n},$$

temos que a fda de T é dada por

$$F(t; \alpha, \beta) = \Phi[\frac{1}{\alpha}(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}),$$

$t > 0, \alpha > 0, \beta > 0.$
Note que β é a mediana da distribuição, isto é,

$$F(\beta; \alpha, \beta) = \Phi(0) = 0.5.$$

Derivando a FDA acima com respeito a t temos a fdp

$$f(t;\alpha,\beta) = \frac{1}{\sqrt{2\pi}} \frac{t^{3/2}(t+\beta)}{2\alpha\sqrt{\beta}} exp[-\frac{1}{2\alpha^2}(\frac{t}{\beta}+\frac{\beta}{t}-2)], \quad t > 0.$$

Pode-se mostrar que se $T \sim BS(\alpha,\beta)$, então

$$Z = \frac{1}{\alpha}[\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}] \sim N(0,1).$$

Reciprocamente, se $Z \sim N(0,1)$, então

$$T = \frac{\beta}{4}[\alpha Z + \sqrt{\alpha^2 Z^2 + 4}]^2 \sim BS(\alpha,\beta).$$