Comparação de Curvas de Sobrevivência

Adaptado de material do
Prof. Vicente G. Cancho (ICMC/USP)

Departamento de Matemática Aplicada e Estatística
Universidade de São Paulo

2023

Dados não censurados

1. Teste de $H_0 : \mu_1 = \mu_2$ *versus* $H_1 : \mu_1 \neq \mu_2$.
   - Pode ser que $F_1(t) \neq F_2(t)$ mesmo que $\mu_1 = \mu_2$.
   - Exemplo: $N(\mu_1, \sigma_1^2)$ *versus* $N(\mu_2, \sigma_2^2)$, com $\sigma_1^2 \neq \sigma_2^2$.

Dados não censurados

1. Teste de $H_0 : \mu_1 = \mu_2$ *versus* $H_1 : \mu_1 \neq \mu_2$.
    - Pode ser que $F_1(t) \neq F_2(t)$ mesmo que $\mu_1 = \mu_2$.
    - Exemplo: $N(\mu_1, \sigma_1^2)$ *versus* $N(\mu_2, \sigma_2^2)$, com $\sigma_1^2 \neq \sigma_2^2$.
2. Kolmogorov-Smirnov: $H_0 : F_1(t) = F_2(t)$ para todo t *versus* $H_1 : F_1(t) \neq F_2(t)$ para pelos menos um t.

Dados não censurados

1. Teste de $H_0 : \mu_1 = \mu_2$ *versus* $H_1 : \mu_1 \neq \mu_2$.
   - Pode ser que $F_1(t) \neq F_2(t)$ mesmo que $\mu_1 = \mu_2$.
   - Exemplo: $N(\mu_1, \sigma_1^2)$ *versus* $N(\mu_2, \sigma_2^2)$, com $\sigma_1^2 \neq \sigma_2^2$.
2. Kolmogorov-Smirnov: $H_0 : F_1(t) = F_2(t)$ para todo t *versus* $H_1 : F_1(t) \neq F_2(t)$ para pelos menos um t.
   - Estatística de teste: $\sup_t |\widehat{F}_1(t) - \widehat{F}_2(t)|$, em que $\widehat{F}_j$, $j = 1, 2$, é a função distribuição empírica.

Dados não censurados

1. Teste de $H_0 : \mu_1 = \mu_2$ *versus* $H_1 : \mu_1 \neq \mu_2$.
    - Pode ser que $F_1(t) \neq F_2(t)$ mesmo que $\mu_1 = \mu_2$.
    - Exemplo: $N(\mu_1, \sigma_1^2)$ *versus* $N(\mu_2, \sigma_2^2)$, com $\sigma_1^2 \neq \sigma_2^2$.
2. Kolmogorov-Smirnov: $H_0 : F_1(t) = F_2(t)$ para todo t *versus* $H_1 : F_1(t) \neq F_2(t)$ para pelos menos um t.
    - Estatística de teste: $\sup_t |\widehat{F}_1(t) - \widehat{F}_2(t)|$, em que $\widehat{F}_j$, $j = 1, 2$, é a função distribuição empírica.
    - Outras estatísticas.

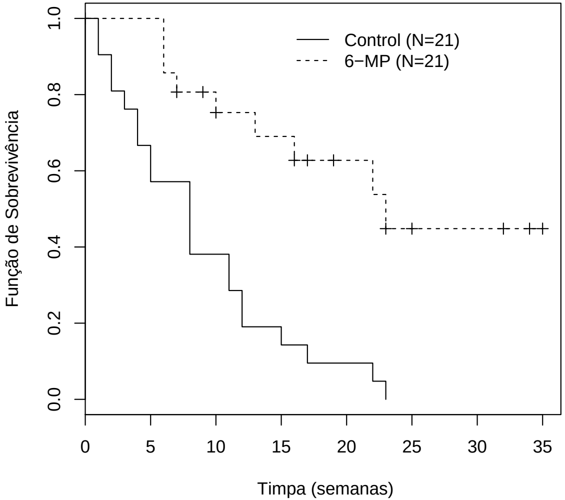
Exemplo. Dados de leucemia

Os dados abaixo representam os tempos de recorrência (em meses) de câncer (leucemia aguda) em 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^+	

Obs. "+" indica tempo censurado.

Exemplo. Dados de leucemia



Comparação das curvas de sobrevivência no tempo t_0

- Pergunta de interesse: O tratamento 6-MP tem um efeito significativo sobre o tempo de recorrência da leucemia aguda?

Comparação das curvas de sobrevivência no tempo t_0

- Pergunta de interesse: O tratamento 6-MP tem um efeito significativo sobre o tempo de recorrência da leucemia aguda?

Comparação das funções sobrevivência no tempo t_0

- Usamos o estimador de Kaplan-Meier $\widehat{S}_0(t_0) - \widehat{S}_1(t_0)$ para estimar $S_0(t_0) - S_1(t_0)$, em que os índices "0" e "1" denotam as duas curvas.
- Usamos a normalidade assintótica do estimador de Kaplan-Meier para obter a distribuição aproximada de $\widehat{S}_0(t_0) - \widehat{S}_1(t_0)$.
- Teste de $H_0 : S_0(t_0) = S_1(t_0)$ versus $H_1 : S_0(t_0) \neq S_1(t_0)$. Estatística de teste (distribuição se H_0 for verdadeira):

$$Z = \frac{\widehat{S}_0(t_0) - \widehat{S}_1(t_0)}{\sqrt{Var(\widehat{S}_0(t_0)) + Var(\widehat{S}_1(t_0))}} \xrightarrow{D} N(0,1).$$

Comparação das curvas de sobrevivência no tempo t_0

Problemas

1. Clinicamente, pode não ser apropriado comparar as funções sobrevivência em um único ponto.

Comparação das curvas de sobrevivência no tempo t_0

Problemas

1. Clinicamente, pode não ser apropriado comparar as funções sobrevivência em um único ponto.
2. Ineficiente, pois descartamos a informação sobre o restante das curvas.

Objetivo

Comparar as curvas de sobrevivência inteiras.

Teste logrank (Mantel, 1966)

- Considere que temos interesse em testar a igualdade de duas funções de sobrevivência.

$$H_0 : S_0(t) = S_1(t) \text{ versus } H_1 : S_0(t) = [S_1(t)]^{\phi}$$

ou

$$\begin{cases} H_0 : h_0(t) = h_1(t) \text{ versus} & \Leftrightarrow H_0 : \frac{h_0(t)}{h_1(t)} = 1 \\ H_1 : h_0(t) = \phi h_1(t) & \Leftrightarrow H_1 : \frac{h_0(t)}{h_1(t)} = \phi, \end{cases}$$

para todo $t > 0$ e $\phi > 0$. H_1 significa riscos proporcionais.
De outra forma,

$$H_0 : \phi = 1 \text{ versus } H_1 : \phi \neq 1.$$

Teste logrank

- Dados: amostras de cada um dos grupos.
- Sejam $t_1 < t_2 < \cdots < t_k$ os tempos de falha distintos da amostra formada pela combinação das duas amostras.
- Para cada t_j, $j = 1, \ldots, k$, considere uma tabela de contingências 2×2 da forma

	Falha		
	Sim	Não	Em risco
Grupo 0 (G0)	d_{0j}	$n_{0j} - d_{0j}$	n_{0j}
Grupo 1 (G1)	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
Total	d_j	$n_j - d_j$	n_j

- n_{ij} : número de indivíduos em risco no grupo i no tempo t_j.
- d_{ij} : número de falhas no grupo i no tempo t_j.
- d_j : número total de indivíduos em risco na amostra combinada no tempo t_j (n_j é análogo).

Teste logrank

- Fixando os totais marginais na tabela acima, a distribuição do número de falhas na amostra do grupo 1, D_{1j}, sob H_0, é a distribuição hipergeométrica, ou seja,

$$D_{1j} \sim \text{Hipergeométrica}(n_j, n_{1j}, d_j).$$

- Para a falha no tempo t_j , calculamos o número observado de falhas no grupo 1 e o número esperado de falhas sob H_0:

$$\begin{aligned} O_{1j} = d_{1j} : & \quad \text{observado.} \\ E_{1j} = E(D_{1j}) = n_{1j}\frac{d_j}{n_j} : & \quad \text{esperado (sob } H_0\text{).} \end{aligned}$$

Sob H_0, o número de falhas no tempo t_j no grupo 1 é proporcional ao número de indivíduos em risco no grupo 1 (proporção $= n_{1j}/n_j$).

Teste logrank

- Para cada tempo de falha calculamos

$$U_j = O_{1j} - E_{1j}.$$

- Sob H_0,

$$\begin{aligned} E(U_j) &= 0 \text{ e} \\ Var(U_j) &= n_{1j}(\frac{d_j}{n_j})(1 - \frac{d_j}{n_j})(\frac{n_j - n_{1j}}{n_j - 1}) = V_j. \end{aligned}$$

- A estatística logrank é dada por (distribuição sob H_0)

$$Z^2 = \frac{\left[\sum_{j=1}^{k}(O_{1j} - E_{1j})\right]^2}{\sum_{j=1}^{k} V_j} = \frac{\left[\sum_{j=1}^{k} U_j\right]^2}{\sum_{j=1}^{k} V_j} \xrightarrow{D} \chi_1^2.$$

Exemplo

- Grupo 0 : 3.1, 6.8^+,9,9,11.3^+, 16.2
- Grupo 1 : 8.7,9,10.1^+,12.1^+,18.7, 23.1^+
- Tempos de falha ordenados: 3.1, 8.7, 9, 16.2, 18.7.
- $k = 5$ (tempos de falha distintos).

	$t_1 = 3,1$		Em risco
G0	1	5	6
G1	0	6	6
	1	11	12

Exemplo

- Grupo 0 : 3.1, 6.8^+,9,9,11.3^+, 16.2
- Grupo 1 : 8.7,9,10.1^+,12.1^+,18.7, 23.1^+
- Tempos de falha ordenados: 3.1, 8.7, 9, 16.2, 18.7.
- $k = 5$ (tempos de falha distintos).

	$t_1 = 3,1$		Em risco	$t_2 = 8,7$		Em risco
G0	1	5	6	0	4	4
G1	0	6	6	1	5	6
	1	11	12	1	9	10

Exemplo

- Grupo 0 : 3.1, 6.8^{+},9,9,11.3^{+}, 16.2
- Grupo 1 : 8.7,9,10.1^{+},12.1^{+},18.7, 23.1^{+}
- Tempos de falha ordenados: 3.1, 8.7, 9, 16.2, 18.7.
- $k = 5$ (tempos de falha distintos).

	$t_1 = 3,1$		Em risco
G0	1	5	6
G1	0	6	6
	1	11	12

$t_2 = 8,7$		Em risco
0	4	4
1	5	6
1	9	10

$t_3 = 9$		Em risco
2	2	4
1	4	5
3	6	9

Exemplo

- Grupo 0 : 3.1, 6.8^+,9,9,11.3^+, 16.2
- Grupo 1 : 8.7,9,10.1^+,12.1^+,18.7, 23.1^+
- Tempos de falha ordenados: 3.1, 8.7, 9, 16.2, 18.7.
- $k = 5$ (tempos de falha distintos).

	$t_1 = 3,1$		Em risco	$t_2 = 8,7$		Em risco
G0	1	5	6	0	4	4
G1	0	6	6	1	5	6
	1	11	12	1	9	10

$t_3 = 9$		Em risco	$t_4 = 16.2$		Em risco
2	2	4	1	0	1
1	4	5	0	2	2
3	6	9	1	2	3

Exemplo

- Grupo 0 : 3.1, 6.8^+,9,9,11.3^+, 16.2
- Grupo 1 : 8.7,9,10.1^+,12.1^+,18.7, 23.1^+
- Tempos de falha ordenados: 3.1, 8.7, 9, 16.2, 18.7.
- $k = 5$ (tempos de falha distintos).

	$t_1 = 3,1$		Em risco	$t_2 = 8,7$		Em risco
G0	1	5	6	0	4	4
G1	0	6	6	1	5	6
	1	11	12	1	9	10

$t_3 = 9$		Em risco	$t_4 = 16.2$		Em risco	$t_5 = 18.7$		Em risco
2	2	4	1	0	1	0	0	0
1	4	5	0	2	2	1	1	2
3	6	9	1	2	3	1	1	2

Exemplo

- Grupo 0 : 3.1, 6.8^{+},9,9,11.3^{+}, 16.2
- Grupo 1 : 8.7,9,10.1^{+},12.1^{+},18.7, 23.1^{+}
- Tempos de falha ordenados: 3.1, 8.7, 9, 16.2, 18.7.
- $k = 5$ (tempos de falha distintos).

	$t_1 = 3,1$		Em risco	$t_2 = 8,7$		Em risco
G0	1	5	6	0	4	4
G1	0	6	6	1	5	6
	1	11	12	1	9	10

$t_3 = 9$		Em risco	$t_4 = 16.2$		Em risco	$t_5 = 18.7$		Em risco
2	2	4	1	0	1	0	0	0
1	4	5	0	2	2	1	1	2
3	6	9	1	2	3	1	1	2

$$O_1 = d_{11} = 0$$

$$E_1 = n_{11}\left(\frac{d_1}{n_1}\right) = 6\left(\frac{1}{12}\right) = \frac{1}{2}$$

$$V_1 = n_{11}\left(\frac{d_1}{n_1}\right)\left(1 - \frac{d_1}{n_1}\right)\left(\frac{n_1 - n_{11}}{n_1 - 1}\right) = 6\left(\frac{1}{12}\right)\left(1 - \frac{1}{12}\right)\left(\frac{12 - 6}{12 - 1}\right) = \frac{1}{4}$$

Exemplo

- Grupo 0 : 3.1, 6.8^{+},9,9,11.3^{+}, 16.2
- Grupo 1 : 8.7,9,10.1^{+},12.1^{+},18.7, 23.1^{+}
- Tempos de falha ordenados: 3.1, 8.7, 9, 16.2, 18.7.
- $k = 5$ (tempos de falha distintos).

	$t_1 = 3,1$		Em risco	$t_2 = 8,7$		Em risco
G0	1	5	6	0	4	4
G1	0	6	6	1	5	6
	1	11	12	1	9	10

$t_3 = 9$		Em risco	$t_4 = 16.2$		Em risco	$t_5 = 18.7$		Em risco
2	2	4	1	0	1	0	0	0
1	4	5	0	2	2	1	1	2
3	6	9	1	2	3	1	1	2

$$O_1 = d_{11} = 0$$

$$E_1 = n_{11}(\frac{d_1}{n_1}) = 6(\frac{1}{12}) = \frac{1}{2}$$

$$V_1 = n_{11}(\frac{d_1}{n_1})(1 - \frac{d_1}{n_1})(\frac{n_1 - n_{11}}{n_1 - 1}) = 6(\frac{1}{12})(1 - \frac{1}{12})(\frac{12 - 6}{12 - 1}) = \frac{1}{4}$$

$$O_2 = d_{12} = 1$$

$$E_2 = n_{12}(\frac{d_2}{n_2}) = 6 \times \frac{1}{10}$$

$$V_2 = n_{12}(\frac{d_2}{n_2})(1 - \frac{d_2}{n_2})(\frac{n_2 - n_{12}}{n_2 - 1}) = 6(\frac{1}{10})(1 - \frac{1}{10})(\frac{10 - 6}{10 - 1}) = \frac{6}{25}$$

$$
\begin{aligned}
O_3 = d_{13} &= 1 \\
E_3 = n_{13}\left(\frac{d_3}{n_3}\right) &= 5 \times \tfrac{3}{9} \\
V_3 = n_{13}\left(\frac{d_3}{n_3}\right)\left(1 - \frac{d_3}{n_3}\right)\left(\frac{n_3 - n_{13}}{n_3 - 1}\right) = 5\left(\frac{3}{9}\right)\left(1 - \frac{3}{9}\right)\left(\frac{9 - 5}{9 - 1}\right) &= \tfrac{5}{9}
\end{aligned}
$$

$$\begin{aligned} O_3 &= d_{13} &&= 1 \\ E_3 &= n_{13}(\frac{d_3}{n_3}) &&= 5 \times \tfrac{3}{9} \\ V_3 = n_{13}(\frac{d_3}{n_3})(1-\frac{d_3}{n_3})(\frac{n_3-n_{13}}{n_3-1}) &= 5(\frac{3}{9})(1-\frac{3}{9})(\frac{9-5}{9-1}) &&= \tfrac{5}{9} \end{aligned}$$

$$\begin{aligned} O_4 &= d_{14} &&= 0 \\ E_4 &= n_{14}(\frac{d_4}{n_4}) &&= 2 \times \tfrac{1}{3} \\ V_4 = n_{14}(\frac{d_4}{n_4})(1-\frac{d_4}{n_4})(\frac{n_4-n_{14}}{n_4-1}) &= 2(\frac{1}{3})(1-\frac{1}{3})(\frac{3-2}{3-1}) &&= \tfrac{2}{9} \end{aligned}$$

$$
\begin{aligned}
O_3 &= d_{13} && = 1 \\
E_3 &= n_{13}\left(\frac{d_3}{n_3}\right) && = 5 \times \tfrac{3}{9} \\
V_3 &= n_{13}\left(\frac{d_3}{n_3}\right)\left(1 - \frac{d_3}{n_3}\right)\left(\frac{n_3 - n_{13}}{n_3 - 1}\right) = 5\left(\frac{3}{9}\right)\left(1 - \frac{3}{9}\right)\left(\frac{9-5}{9-1}\right) && = \tfrac{5}{9}
\end{aligned}
$$

$$
\begin{aligned}
O_4 &= d_{14} && = 0 \\
E_4 &= n_{14}\left(\frac{d_4}{n_4}\right) && = 2 \times \tfrac{1}{3} \\
V_4 &= n_{14}\left(\frac{d_4}{n_4}\right)\left(1 - \frac{d_4}{n_4}\right)\left(\frac{n_4 - n_{14}}{n_4 - 1}\right) = 2\left(\frac{1}{3}\right)\left(1 - \frac{1}{3}\right)\left(\frac{3-2}{3-1}\right) && = \tfrac{2}{9}
\end{aligned}
$$

$$
\begin{aligned}
O_5 &= d_{15} && = 1 \\
E_5 &= n_{15}\left(\frac{d_5}{n_5}\right) && = 2 \times \tfrac{1}{2} \\
V_5 &= n_{15}\left(\frac{d_5}{n_5}\right)\left(1 - \frac{d_5}{n_5}\right)\left(\frac{n_5 - n_{15}}{n_5 - 1}\right) = 2\left(\frac{1}{2}\right)\left(1 - \frac{1}{2}\right)\left(\frac{2-2}{2-1}\right) && = 0
\end{aligned}
$$

Em resumo, temos

j	1	2	3	4	5
O_j	0	1	1	0	1
E_j	1/2	3/5	15/9	2/3	1
V_j	1/4	6/25	5/9	2/9	0

Calculamos a estatística

$$Z^2 = \frac{[\sum_{j=1}^{k}(O_{1j} - E_{1j})]^2}{\sum_{j=1}^{k} V_j} = \frac{(3 - 4,43)^2}{1,27} = 1,62,$$

com valor-p $=P(Z_0^2 > 1,62) = 0,2027$, em que $Z_0^2 \sim \chi_1^2$.

- Conclusão: Não há diferença significativa entres as curvas de sobrevivência.

Exemplo em R

- t=c(3.1, 6.8, 9.0, 9.0, 11.3, 16.2, 8.7, 9.0 ,10.1, 12.1, 18.3, 23.1)
- d=c(1, 0 ,1, 1, 0, 1, 1, 1, 0, 0 ,1, 0)
- grupo=c(1,1,1,1,1,1,2,2,2,2,2,2)
- fit=survdiff(Surv(t,d)~grupo)
- fit

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
grupo=1	6	4	2.57	0.800	1.62
grupo=2	6	3	4.43	0.463	1.62

Chisq= 1.6 on 1 degrees of freedom, p= 0.2

Exemplo com dados de leucemia em R

- library(KMsurv)
- data(drug6mp)
- drug6mp[1:7,]

	pair	remstat	t1	t2	relapse
1	1	1	1	10	1
2	2	2	22	7	1
3	3	2	3	32	0
4	4	2	12	23	1
5	5	2	8	22	1
6	6	1	17	6	1
7	7	2	2	16	1

- time=c(drug6mp$t1,drug6mp$t2)
- irelapse=c(rep(1,21),drug6mp$relapse)
- grupo=c(rep(1,21),rep(2,21))
- fit=survdiff(Surv(time, irelapse)~grupo)

Exemplo com dados de leucemia em R

- fit=survdiff(Surv(time, irelapse)~grupo)

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
grupo=1	21	21	10.7	9.77	16.8
grupo=2	21	9	19.3	5.46	16.8

  Chisq= 16.8 on 1 degrees of freedom, p= 4e-05

- Conclusão: Existe diferença estatisticamente significativa entre as curvas de sobrevivência.

Outros testes

- No cálculo de Z^2, considere que a ponderação de (Obs-Esp) varia ao longo do tempo.
- A estatística de teste é dada por

$$\frac{\left[\sum_{j=1}^{k} w_j(O_{1j} - E_{1j})\right]^2}{\sum_{j=1}^{k} w_j V_j} = \frac{\left[\sum_{j=1}^{k} U_j\right]^2}{\sum_{j=1}^{k} w_j^2 V_j} \sim \chi^2_{(1)}.$$

  com $U_i = w_j(O_{1j} - E_{1j})$.
- Permite inflar diferenças precoces ou tardias.
- Aumento do poder do teste sob riscos não proporcionais.

Outros testes

Propostas

- Se $w_j = n_j$, obtemos o teste de Gehan-Breslow.
  - Pesos iguais ao número total de indivíduos em risco.
  - Maior peso aos tempos de falha iniciais.
- Se $w_j = \widehat{S}(t_j^-)$, obtemos o teste generalizado de Wilcoxon.
  - Pesos iguais à estimativa combinada de sobrevivência imediatamente anterior ao tempo t_j,
  - Maior peso aos tempos de falha iniciais.
- Se $w_j = [\widehat{S}(t_j^-)]^\rho[1 - \widehat{S}(t_j^-)]^\gamma$, obtemos a família de estatísticas $G^{\rho,\gamma}$ (Harrington-Fleming, 1982).
  - $\rho = \gamma = 0$ leva ao teste logrank.
  - $\rho = 1$ e $\gamma = 0$ leva ao teste generalizado de Wilcoxon.

Outros testes

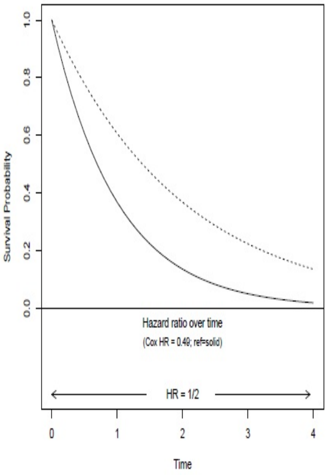
Propostas

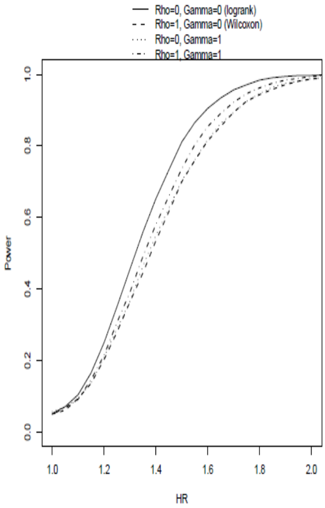
- Se $w_j = n_j$, obtemos o teste de Gehan-Breslow.
  - Pesos iguais ao número total de indivíduos em risco.
  - Maior peso aos tempos de falha iniciais.
- Se $w_j = \widehat{S}(t_j^-)$, obtemos o teste generalizado de Wilcoxon.
  - Pesos iguais à estimativa combinada de sobrevivência imediatamente anterior ao tempo t_j,
  - Maior peso aos tempos de falha iniciais.
- Se $w_j = [\widehat{S}(t_j^-)]^\rho[1 - \widehat{S}(t_j^-)]^\gamma$, obtemos a família de estatísticas $G^{\rho,\gamma}$ (Harrington-Fleming, 1982).
  - $\rho = \gamma = 0$ leva ao teste logrank.
  - $\rho = 1$ e $\gamma = 0$ leva ao teste generalizado de Wilcoxon.

Obs. Algumas opções estão implementadas na função `survdiff` do pacote `survival` em R.

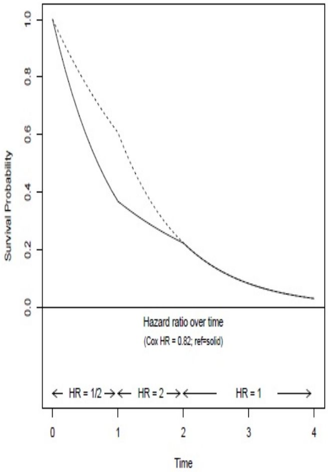
Comparação do poder

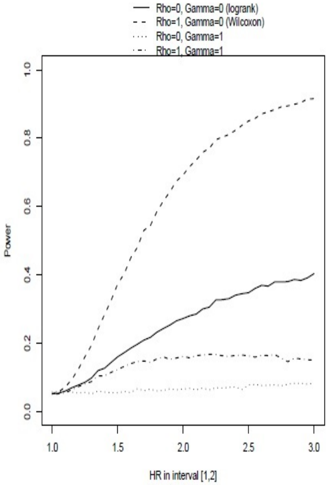
Obs. HR (*hazard ratio*) ou RR é a razão de riscos dada por $HR(t) = h_0(t)/h_1(t)$.



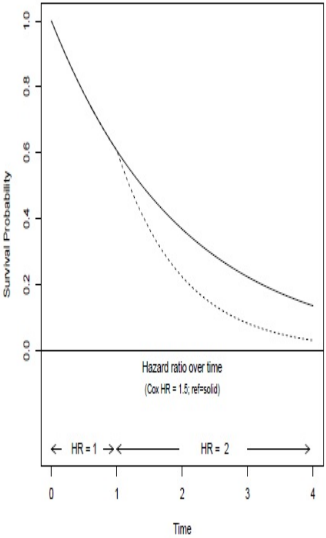


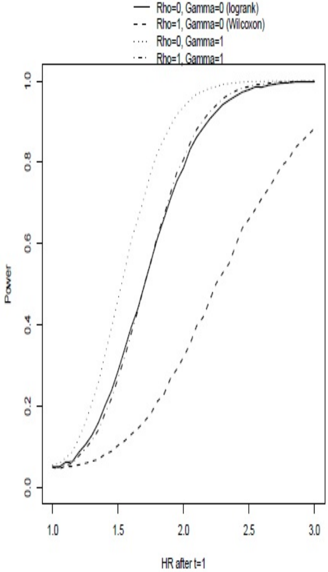
Comparação do poder com sobrevivência divergente precoce





Comparação do poder com sobrevivência divergentes tardia





Quando devemos utilizar o teste logrank ponderado?

- O teste logrank (não ponderado) é mais poderoso sob riscos proporcionais.
- Como podemos (informalmente) verificar a suposição de riscos proporcionais?
  - Se temos riscos proporcionais, então

    $$h_0(t) = \phi h_1(t), \text{ ou seja, } RR(t) = \frac{h_0(t)}{h_1(t)} = \phi,$$

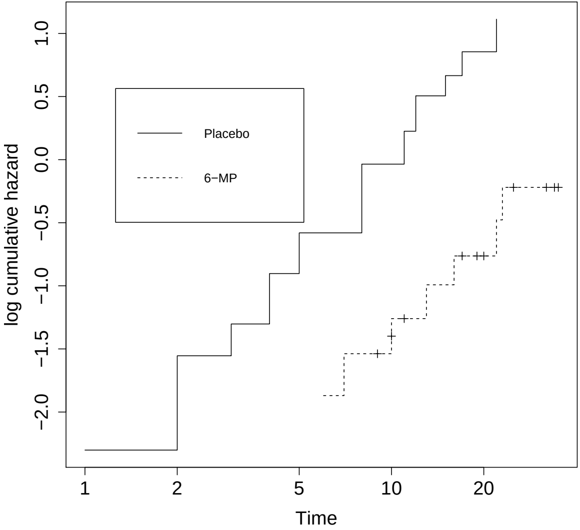
    de modo que

    $$\log(H_0(t)) = \log(\phi) + \log(H_1(t)).$$

- Assim, se as estimativas dos logaritmos dos riscos acumulados são aproximadamente paralelas, então o teste de logrank tenderá a ser mais poderoso.

Obs. $\log(H(t)) = \log(-\log(S(t)))$.

Riscos acumulados para os dados de leucemia



Comparação de R curvas de sobrevivência

- Suponha que temos $R > 2$ grupos e queremos compará-los simultaneamente com relação à sobrevivência (ou risco).

Comparação de R curvas de sobrevivência

- Suponha que temos $R > 2$ grupos e queremos compará-los simultaneamente com relação à sobrevivência (ou risco).

$$H_0 : h_1(t) = h_2(t) = \cdots = h_R(t),$$

ou seja, as curvas de sobrevivência para todos os grupos são as mesmas para todo t.

- Particularmente, temos interesse como hipótese alternativa

$$H_1 : \text{Ao menos um } h_i(t) \neq h_j(t), \text{ para algum}, i \neq j = 1, \ldots, R.$$

Comparação de R curvas de sobrevivência

- A estatística de teste é uma generalização da estatística para duas amostras, que depende da covariância entres as diferenças (Obs - Esp) em cada grupo,
- Consideramos os dados no j-ésimo tempo de falha t_j, $j = 1, \ldots, k$, na amostra combinada formando uma tabela $2 \times R$.

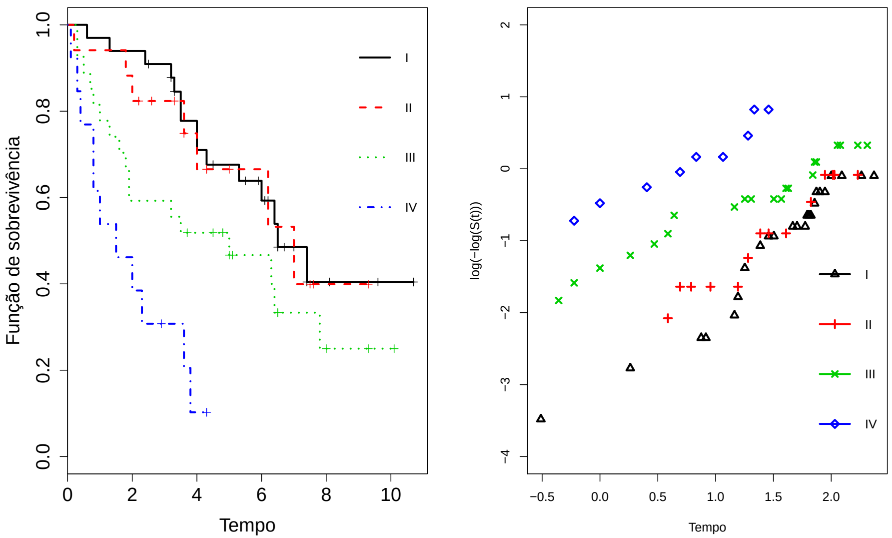
1	2	...	m	...	R
d_{1j}	d_{2j}	...	d_{mj}	...	d_{Rj}
$n_{1j} - d_{1j}$	$n_{2j} - d_{2j}$	...	$n_{mj} - d_{mi}$	...	$n_{Rj} - d_{Rj}$

- Sob H_0, a estatística de teste tem distribuição assintótica χ^2_{R-1}.

Exemplo. Câncer de laringe

- Dados de sobrevivência em pacientes com câncer de laringe (em meses) do pacote `KMsurv` em R.
- Tempo inicial: diagnóstico de câncer.
- Evento de interesse: óbito devido ao câncer de laringe.
- Questão de interesse: Como o tempo de sobrevivência desde o diagnóstico até o óbito varia em função do estágio da doença no momento do diagnóstico (I,II,III,IV)?

Exemplo. Câncer de laringe



Exemplo. Câncer de laringe

- library(KMsurv)
- survdiff(Surv(time,delta)∼ stage, data=larynx)

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
stage=1	33	15	22.57	2.537	4.741
stage=2	17	7	10.01	0.906	1.152
stage=3	27	17	14.08	0.603	0.856
stage=4	13	11	3.34	17.590	19.827

Chisq= 22.8 on 3 degrees of freedom, p= 5e-05

- Conclusão: Há diferença significativa entre as curvas de sobrevivência.

Exemplo. Câncer de laringe

- library(KMsurv)
- survdiff(Surv(time,delta)∼ stage, data=larynx)

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
stage=1	33	15	22.57	2.537	4.741
stage=2	17	7	10.01	0.906	1.152
stage=3	27	17	14.08	0.603	0.856
stage=4	13	11	3.34	17.590	19.827

  Chisq= 22.8 on 3 degrees of freedom, p= 5e-05
- Conclusão: Há diferença significativa entre as curvas de sobrevivência.

Exercício

- Faça uma análise estatística e responda à questão de interesse.
- Verifique entre quais grupos de pacientes a diferença é significativa.