EPI10 - Análise de Sobrevivência

Comparação de funções de sobrevivência

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

Universidade Federal do Rio Grande do Sul
Faculdade de Medicina
Programa de Pós-Graduação em Epidemiologia

Porto Alegre, 2021

Relembrando

Estudo de Hepatite

- Um estudo clínico aleatorizado foi realizado para investigar o efeito da terapia com esteroide no tratamento de hepatite viral aguda.
- **Vinte e nove pacientes com esta doença** foram aleatorizados para receber um placebo ou o tratamento com esteroide.
- Cada paciente foi acompanhado por 16 semanas ou até o óbito (evento de interesse) ou até a perda de acompanhamento.

Estudo de Hepatite

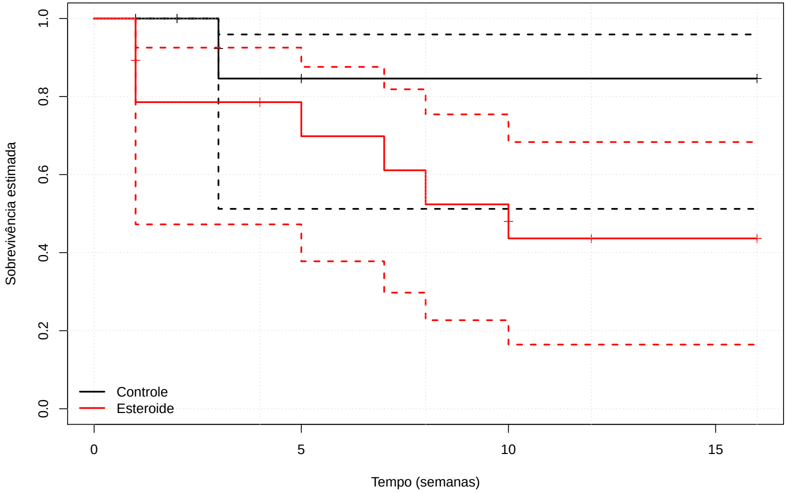
- Os tempos de sobrevivência observados, em semanas, para os dois grupos são apresentados na tabela a seguir (+ indica censura).

Grupo	Tempo de sobrevivência em semanas
Controle	1+, 2+, 3, 3, 3+, 5+, 5+, 16+, 16+, 16+, 16+, 16+, 16+, 16+, 16+
Esteroide	1, 1, 1, 1+, 4+, 5, 7, 8, 10, 10+, 12+, 16+, 16+, 16+

Estudo de Hepatite

```
## Call: survfit(formula = Surv(time = tempo, event = cens) ~ grupo, data = df.
##     conf.type = "log-log")
##
##                  n events median 0.95LCL 0.95UCL
## grupo=Controle  15      2     NA      NA      NA
## grupo=Esteroide 14      7     10       1      NA
```

Estudo de Hepatite



Estudo de Hepatite

Considerações

- Aparentemente, o grupo Esteroide apresenta uma sobrevivência menor que o grupo Controle.
  - O **tempo mediano de sobrevivência** para o grupo Esteroide é estimado em 10 semanas; para o grupo Controle, o tempo mediano de sobrevivência é maior que 16 semanas (último tempo de acompanhamento).
  - A probabilidade de um indivíduo do grupo Esteroide sobreviver a 12 semanas é estimada em 0,437 (IC 95% 0,164-0,683); no grupo Controle, esta probabilidade é estimada em 0,846 (IC 95% 0,512-0,959).
  - As curvas de sobrevivência dos dois grupos não atingem o valor zero; isto sempre ocorre quando o maior tempo observado na amostra é uma censura.

Comparação de funções de sobrevivência

Comparação de funções de sobrevivência

- **Pergunta:** as funções de sobrevivência do grupo Controle e Esteroide diferem?
- Em outras palavras, como podemos testar $S_1(t) = S_2(t)$.
  - Os intervalos de confiança construídos anteriormente $\{$são pontuais$\}$. Ou seja, para cada ponto t temos um intervalo de $100(1-\alpha)\%$ confiança para $S(t)$. Mas, este coeficiente de confiança não é garantido quando olhamos para a "curva toda".
- O teste de ***log-rank***[1] pode responder esta questão adequadamente.

[1] Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*. 1966 Mar;50(3):163-70.

Teste log-rank

- Sejam $t_1 < t_2 < \ldots < t_k$ os tempos de falha distintos da amostra formada pela combinação das duas amostras individuais (ou seja, os tempos de ocorrência dos eventos dos dois grupos combinados).
- Suponha que no tempo t_j acontecem $d_j = d_{1j} + d_{2j}$ eventos e $n_j = n_{1j} + n_{2j}$ indivíduos estão sob risco em um tempo imediatamente inferior a t_j na amostra combinada, em que d_{ij} e n_{ij} são o número de eventos e indivíduos em risco, respectivamente, na amostra i, para $i = 1, 2$ **(1: Controle e 2: Esteroide, por exemplo)** e $j = 1, \ldots, k$.

Teste log-rank

- Em **cada tempo de ocorrência do evento** t_j, os dados podem ser organizados em uma tabela de contingência 2×2 com d_{ij} eventos e $n_{ij} - d_{ij}$ não eventos na **coluna** i

	Grupo		
	1	2	
Evento	d_{1j}	d_{2j}	d_j
Não evento	$n_{1j} - d_{1j}$	$n_{2j} - d_{2j}$	$n_j - d_j$
	n_{1j}	n_{2j}	n_j

Teste log-rank

- Da mesma forma que na análise de muitas tabelas 2×2, os indivíduos em risco são classificados nessas tabelas para responder à pergunta:
  - **o fator de risco (grupo de tratamento, ou exeposição) está associado à sobrevivência?**
- Para tratar de forma eficaz esta questão com uma única medida de associação, a medida escolhida deve ser constante com respeito ao tempo de sobrevivência.
- Para criar um único resumo abrangente, os dados são estratificados pelo tempo de ocorrência do evento.
- Uma medida de risco não é influenciada pelo tempo de sobrevivência quando é calculada **dentro de cada estrato** (tabela) e combinada em todos os estratos para resumir a associação entre o fator de risco e o desfecho.

Teste log-rank

- Retomando o exemplo do **estudo de hepatite**, o tempo do óbito $t_1 = 1$, gera a primeira tabela 2×2, em que

	Grupo		
	Controle	Esteroide	
Óbito	0	3	3
Sobreviveu	15	11	26
	15	14	29

Teste log-rank

- Note que, condicional à experiência de falha e censura até o tempo t_j (fixando as marginais da coluna) e ao número de eventos no tempo t_j (fixando as marginais de linha), o valor observado d_{2j} é então, sob H_0 **(as variáveis Grupo e Desfecho são independentes)**, a realização de uma variável aleatória **hipergeométrica**, D_{2j}, com distribuição de probabilidade

$$\Pr(D_{2j} = d_{2j} | H_0) = \frac{\binom{n_{1j}}{d_{1j}}\binom{n_{2j}}{d_{2j}}}{\binom{n_j}{d_j}}, \; \max(0, d_j - n_{1j}) \leq d_{2j} \leq min(d_j, n_{2j}).$$

Teste log-rank

- Sob H_0, é possível mostrar que a **média** e a **variância** de D_{2j} são, respectivamente

$$\overline{D}_{2j} = \frac{d_j \times n_{2j}}{n_j} \quad \text{e} \quad V(D_{2j}) = \frac{n_{1j} \times n_{2j} \times d_j \times (n_j - d_j)}{n_j^2(n_j - 1)}.$$

- O valor $\overline{D}_{2j}$ pode ser visto como o **número esperado de eventos no grupo tratamento** (ou exposto), sob a hipótese nula de independência entre tratamento e desfecho, em t_j.

Teste log-rank

- Uma estatística de teste poderia considerar a comparação entre o número observado e esperado de eventos.
- Um teste para grandes amostras de H_0 (as variáveis Grupo e Desfecho são independentes) envolve a estatística

$$Z = \frac{d_{2j} - \overline{D}_{2j}}{\sqrt{V(D_{2j})}} \overset{a}{\sim} N(0,1),$$

ou, equivalentemente,

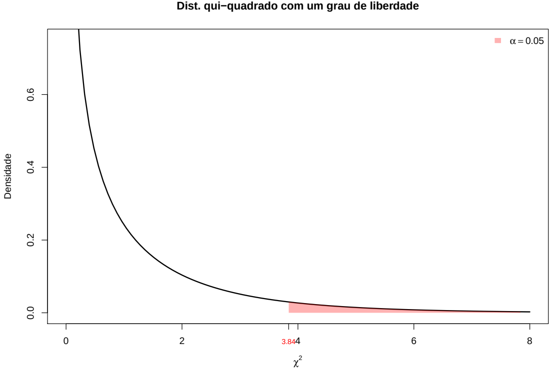
$$\chi^2 = Z^2 = \frac{(d_{2j} - \overline{D}_{2j})^2}{V(D_{2j})} \overset{a}{\sim} \chi^2(1).$$

- Esta estatística χ^2 é conhecido como a **estatística qui-quadro de Mantel-Haenszel**[2].

[2] Mantel N., Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959 Apr;22(4):719-48.

Teste log-rank

Lembrando



- Tabelas e *softwares* podem ser utilizados para avaliação da estatística de teste.

Teste log-rank

Retornando ao exemplo do **estudo de hepatite** no tempo do óbito t_1, temos

- Número observado de óbitos no grupo Esteroide: $d_{21} = 3$;
- Número esperado de óbitos no grupo Esteroide (sob H_0): $\frac{d_1 \times n_{21}}{n_1} = (3 \times 14)/29 \approx 1.45$.

Teste log-rank

- O tempo do óbito $t_2 = 3$, gera uma segunda tabela 2×2, em que

	Grupo		
	Controle	Esteroide	
Óbito	2	0	3
Sobreviveu	11	10	26
	13	10	23

- Número observado de óbitos no grupo Esteroide: $d_{22} = 0$;
- Número esperado de óbitos no grupo Esteroide (sob H_0): $\frac{d_2 \times n_{22}}{n_2} = (2 \times 10)/23 \approx 0.87$.

Teste log-rank

Considerando os k tempos distintos de falha, poderíamos organizar os dados referentes as k tabelas 2×2 na seguinte tabela

t_j	n_j	d_j	n_{1j}	d_{1j}	n_{2j}	d_{2j}	$\overline{D}_{2j}$	$d_{2j} - \overline{D}_{2j}$	$V(D_{2j})$
1	29	3	15	0	14	3	1,448	1,552	0,696
3	23	2	13	2	10	0	0,870	-0,870	0,469
5	19	1	10	0	9	1	0,474	0,526	0,249
7	16	1	8	0	8	1	0,500	0,500	0,250
8	15	1	8	0	7	1	0,467	0,533	0,249
10	14	1	8	0	6	1	0,429	0,571	0,245
Total	-	9	-	2	-	7	4,187	2,813	2,158

Teste log-rank

Teste log-rank

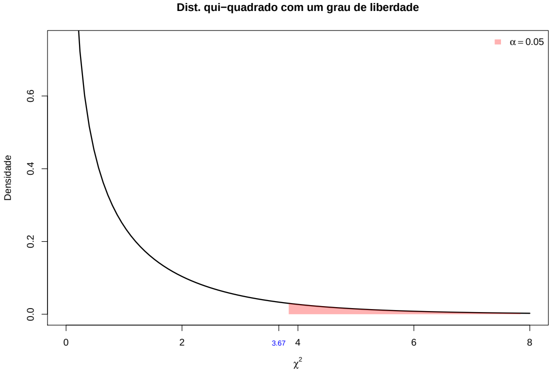
- Se as k tabelas de contingência forem independentes, um teste aproximado para a igualdade das duas funções de sobrevivência pode ser baseado na estatística

$$\chi^2_{LR} = \frac{\left[\sum_{j=1}^{k} (d_{2j} - \overline{D}_{2j})\right]^2}{\sum_{j=1}^{k} V(D_{2j})}$$

- Sob a hipótese nula $H_0 : S_1(t) = S_2(t)$ **para todo o** t, em grandes amostras, tem uma **distribuição aproximada qui-quadrado com 1 grau de liberdade**.

Teste log-rank

- Em nosso exemplo, $\chi^2_{LR} = (2,813)^2/2,158 \approx 3,67$.



- O valor desta estatística corresponde ao valor $p = \Pr(\chi^2_{LR} \geq 3,67) \approx 0,0555$.

Código R

```
survdiff(Surv(time = tempo, event = cens) ~ grupo,
         data = df.hep)
```

```
## Call:
## survdiff(formula = Surv(time = tempo, event = cens) ~ grupo,
##     data = df.hep)
##
##                  N Observed Expected (O-E)^2/E (O-E)^2/V
## grupo=Controle  15        2     4.81      1.64      3.67
## grupo=Esteroide 14        7     4.19      1.89      3.67
##
##  Chisq= 3.7  on 1 degrees of freedom, p= 0.06
```

- Ao nível de 5% de significância não há evidências contra H_0.
  - Portanto, as curvas de sobrevivência dos grupos Controle e Esteroide podem ser iguais.

Teste log-rank

- A generalização do teste **log-rank** avaliar a hipótese de igualdade entre $S_1(t), S_2(t), \ldots, S_r(t)$, $r > 2$.
  - O desenvolvimento não será demonstrado, mas salienta-se que esta estatística de teste **log-rank** generalizado para r funções tem **distribuição aproximada qui-quadrado com $r-1$ graus de liberdade**.
- Neste caso, se H_0 é rejeitada, concluímos que pelo menos um grupo difere dos demais em relação à função de sobrevivência.
- Para identificarmos quais grupos diferem uns dos outros, uma possibilidade é realizar comparações dos grupos, **dois a dois**, por meio do teste de **log-rank** para dois grupos.
  - O **método de Bonferroni** (α/[número de comparações múltiplas]) pode ser utilizado para controlar as taxas de erro tipo I.

Exemplo

Estudo de Malária

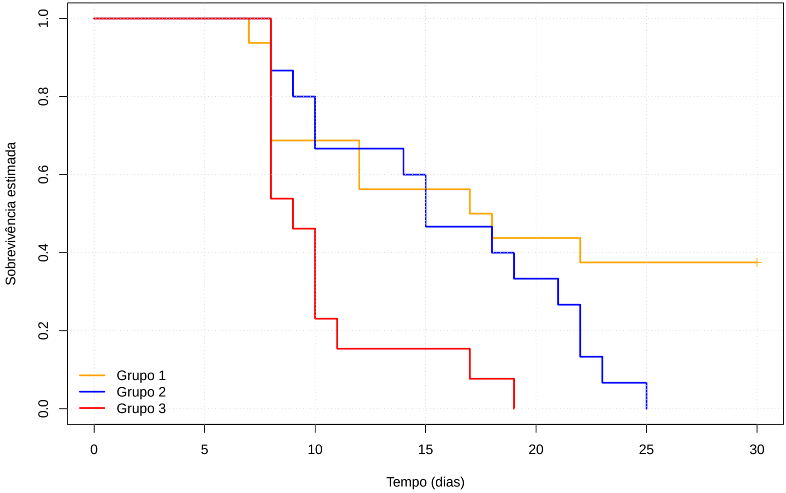
- Um estudo experimental realizado com camundongos para verificar a eficácia da imunização pela malária foi conduzido no Centro de Pesquisas Renee Rachou, Fiocruz, Minas Gerais.
- Nesse estudo, quarenta e quatro camundongos foram infectados pela malária.
  - Os camundongos do grupo 1 foram imunizados 30 dias antes da infecção.
  - Além da infecção pela malária, os camundongos dos grupos 1 e 3 foram, também, infectados pela esquistossomose.
- O desfecho de interesse nesse estudo foi o tempo (em dias) decorrido desde a infecção pela malária até a morte do camundongo.
  - O estudo teve duração de 30 dias.

Estudo de Malária

```
df.mala <- read.table("../dados/malaria.csv",
                      sep = ";",
                      header = TRUE)
head(df.mala)
```

```
##   id grupo tempo cens
## 1  1     1     7    1
## 2  2     1     8    1
## 3  3     1     8    1
## 4  4     1     8    1
## 5  5     1     8    1
## 6  6     1    12    1
```

Estudo de Malária



Estudo de Malária

```
survdiff(Surv(time = tempo, event = cens) ~ grupo,
         data = df.mala)
```

```
## Call:
## survdiff(formula = Surv(time = tempo, event = cens) ~ grupo,
##     data = df.mala)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## grupo=1 16       10    17.00    2.8816    6.4111
## grupo=2 15       15    14.51    0.0167    0.0317
## grupo=3 13       13     6.49    6.5190   10.4447
##
##  Chisq= 12.6  on 2 degrees of freedom, p= 0.002
```

Estudo de Malária

- Constatada a diferença entre os grupos **($p = 0,002$)**, existe a necessidade de identificar quais curvas diferem entre si.
- Se realizarmos comparações dois a dois, o método de Bonferroni ajusta o nível de significância de acordo com o número de comparações múltiplas.
  - Como temos três grupos, três comparações dois a dois são possíveis de se realizar.
  - Utilizando o nível de 5% de significância, o nível de significância ajustado por Bonferroni é $\alpha^* = \alpha/3 = 0,05/3 = 0,017$ para cada um dos testes.

Estudo de Malária

Grupo 1 vs. Grupo 2

```r
# grupo 1 vs grupo 2
survdiff(Surv(time = tempo, event = cens) ~ grupo,
         data = df.mala,
         subset = grupo != 3)
```

```
## Call:
## survdiff(formula = Surv(time = tempo, event = cens) ~ grupo,
##     data = df.mala, subset = grupo != 3)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## grupo=1 16       10     13.7      1.01      2.53
## grupo=2 15       15     11.3      1.23      2.53
##
##  Chisq= 2.5  on 1 degrees of freedom, p= 0.1
```

Estudo de Malária

Grupo 1 vs. Grupo 3

```
# grupo 1 vs grupo 3
survdiff(Surv(time = tempo, event = cens) ~ grupo,
         data = df.mala,
         subset = grupo != 2)
```

```
## Call:
## survdiff(formula = Surv(time = tempo, event = cens) ~ grupo,
##     data = df.mala, subset = grupo != 2)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## grupo=1 16       10    15.34      1.86      7.86
## grupo=3 13       13     7.66      3.72      7.86
##
##  Chisq= 7.9  on 1 degrees of freedom, p= 0.005
```

Estudo de Malária

Grupo 2 vs. Grupo 3

```
# grupo 2 vs grupo 3
survdiff(Surv(time = tempo, event = cens) ~ grupo,
         data = df.mala,
         subset = grupo != 1)
```

```
## Call:
## survdiff(formula = Surv(time = tempo, event = cens) ~ grupo,
##     data = df.mala, subset = grupo != 1)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## grupo=2 15       15    20.53      1.49      7.98
## grupo=3 13       13     7.47      4.08      7.98
##
##  Chisq= 8  on 1 degrees of freedom, p= 0.005
```

Estudo de Malária

- **Conclusão:** ao nível de 5% de significância,
  - entre os grupos 1 e 2, não foram encontradas evidências de diferenças;
  - a diferença entre os grupos 1 e 3 atesta a eficácia da imunização pela malária na presença de infecções pela malária e pela equistossomose;
  - por outro lado, a diferença entre os grupos 2 e 3 mostra o impacto na mortalidade dos camundongos devido à infecção pela esquistossomose.

Considerações

- Outros testes, alternativos ao (ou generalizações do) *log-rank*, foram propostos na literatura:
  - generalização para a estatística de Wilcoxon[3];
  - Peto e Peto (1972)[4];
  - Tarone e Ware (1977)[5];
  - Prentice (1978)[6].
- Foi dado ênfase ao teste de *log-rank*, pois este possui boas propriedades estatísticas, além de ser um dos testes mais utilizados para comparar curvas de sobrevivência.

---

[3] Gehan, E. A. (1965). A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples. *Biometrika*, 52(1/2), 203–223.

[4] Peto, R., & Peto, J. (1972). Asymptotically Efficient Rank Invariant Test Procedures. *Journal of the Royal Statistical Society. Series A (General)*, 135(2), 185–207.

[5] Tarone, R. E., & Ware, J. (1977). On Distribution-Free Tests for Equality of Survival Distributions. *Biometrika*, 64(1), 156–160.

[6] Prentice, R. L. (1978). Linear Rank Tests with Right Censored Data. *Biometrika*, 65(1), 167–179.

Considerações

- Com o estimador de **Kaplan-Meier** e o teste de **log-rank** é possível:
  - descrever dados de sobrevivência;
  - comparar funções de sobrevivência entre grupos.
- No entanto, estamos limitados a avaliar a influência de covariáveis (exposições ou tratamentos) discretas (categóricas ou categorizadas) na função de sobevivência em análises não ajustadas.
- Como avaliar o efeito, na função de sobrevivência, de covariáveis contínuas, e ajustando para potenciais vairáveis de confusão?
  - Uma possibilidade é proposição de modelos estatísticos com uma estrutura de regressão.
  - Um modelo muito utilizado é o **modelo de Cox**.

Para casa

1. Leia o capítulo 2 do livro **Análise de sobrevivência aplicada**[7].
2. Leia os capítulo 4 do livro **Análise de sobrevivência: teoria e aplicações em saúde**[8].

[7] Colosimo, E. A. e Giolo, S. R. **Análise de sobrevivência aplicada**, Blucher, 2006.

[8] Carvalho, M. S., Andreozzi, V. L., Codeço, C. T., Campos, D. P., Barbosa, M. T. S. e Shimakura, E. S. **Análise de sobrevivência: teoria e aplicações em saúde**, 2ª ed. Editora Fiocruz, 2011.

Próxima aula

- Modelo de Cox.

Por hoje é só!

Bons estudos!