

# Exercício 9 - Emerson Adam

2023-04-09

## Análise do Dataset Lung

Este arquivo mostra a análise estatística da database LUNG do NCCTG(North Central Cancer Treatment Group), que demonstra dados de pacientes com câncer de pulmão em estágio avançado.

Este dataset contém 228 linhas(registros) e 10 colunas (variáveis). Algumas variáveis de exemplo como tempo de sobrevivência, sexo, peso e calorias ingeridas. Neste estudo iremos utilizar somente idade e sexo para identificar fatores de risco relacionados ao óbito do paciente.

Nota-se que o tempo médio de vida é de 310 dias.

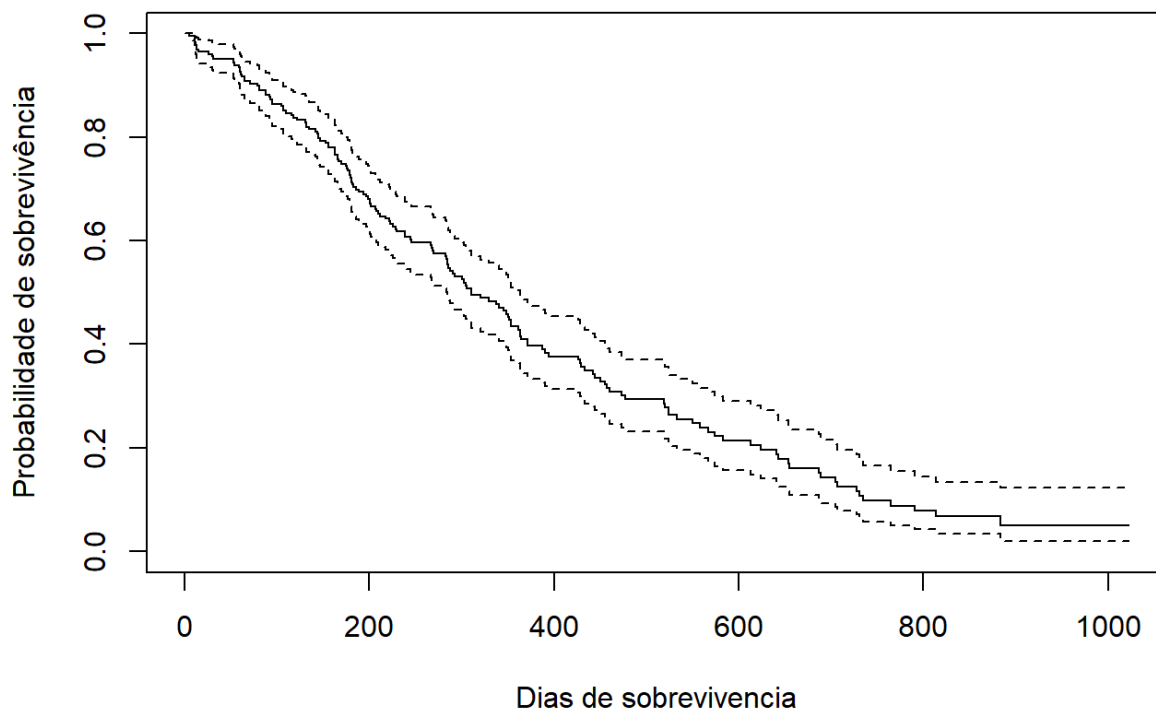
km

```
## Call: survfit(formula = Surv(time, status_adj) ~ 1, data = dados)
##
##          n events median 0.95LCL 0.95UCL
## [1,] 228      165      310      285      363
```

## Etapa exploratória do dataset

Iremos utilizar o método Kaplan Meier neste estudo, pois é o mais indicado para análise não paramétrica para estimar a taxa/função de sobrevivência a partir de dados de vida.

No gráfico abaixo vemos que há queda brusca na taxa de sobrevivência a partir de 180 dias.



Ao verificarmos a proporção entre homens e mulheres que vieram a óbito neste dataset, encontramos que 68% dos registros são homens e 32% são mulheres. Logo espera-se uma taxa de óbito maior para homens.

```
##
##           0           1
## Homens    0.4126984 0.6787879
## Mulheres  0.5873016 0.3212121
```

1 = obito 0 = censurado

## Teste de Hipótese

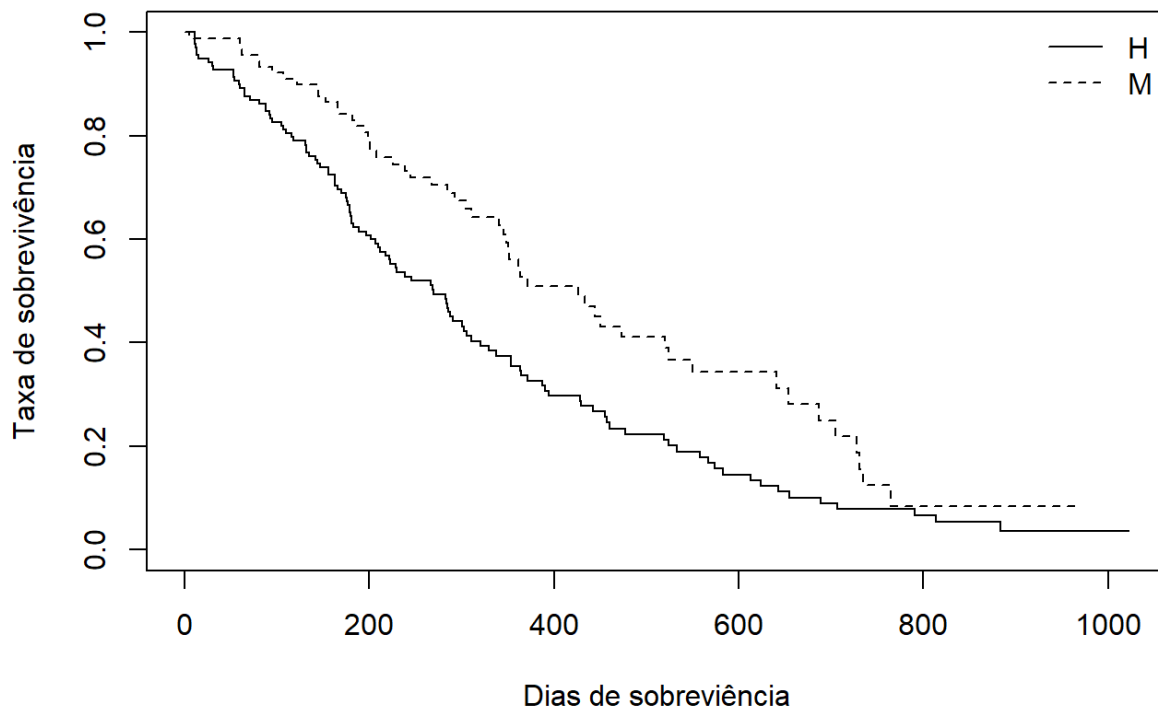
#H0 a taxa de sobrevivência entre H e M é =

#H1 a taxa de sobrevivência entre H e M é !=

Ao olharmos o tempo de sobrevivência em um ano, separando os sexos, temos que a taxa de sobrevivência das mulheres é maior do que a dos homens, sendo 53% e 34% respectivamente.

```
## Call: survfit(formula = Surv(time, status_adj) ~ sex_cat, data = dados)
##
##           sex_cat=Homens
##      time      n.risk    n.event  survival   std.err lower 95% CI
## 365.0000     35.0000    85.0000    0.3361    0.0434    0.2609
## upper 95% CI
##      0.4329
##
##           sex_cat=Mulheres
##      time      n.risk    n.event  survival   std.err lower 95% CI
## 365.0000     30.0000    36.0000    0.5265    0.0597    0.4215
## upper 95% CI
##      0.6576
```

Colocando isto em um grafico, fica ainda mais evidente que a taxa de sobrevivência das mulheres é maior do que a dos homens. Logo podemos rejeitar h0 sendo que a taxa entre os 2 grupos é diferente



Ao realizarmos o teste de Log-Rank, temos um valor de p menor que 0,05 e que podemos rejeitar h0, ou seja, a taxa de sobrevivência é diferente entre os 2 grupos.

```
## Call:
## survival::survdiff(formula = Surv(time, status_adj) ~ sex_cat,
##   data = dados)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## sex_cat=Homens 138      112    91.6      4.55     10.3
## sex_cat=Mulheres 90       53    73.4      5.68     10.3
##
##   Chisq= 10.3  on 1 degrees of freedom, p= 0.001
```

## Ajuste do modelo

Nesta etapa, iremos ajustar o modelo avaliando os fatores de risco relacionados ao óbito. Neste estudo iremos usar somente a idade e sexo, porém outros fatores como peso e calorias ingeridas poderiam ser avaliadas.

Ao utilizarmos o modelo de Cox para explicarmos o tempo de sobrevivência em função de idade e sexo, temos o seguinte resultado;

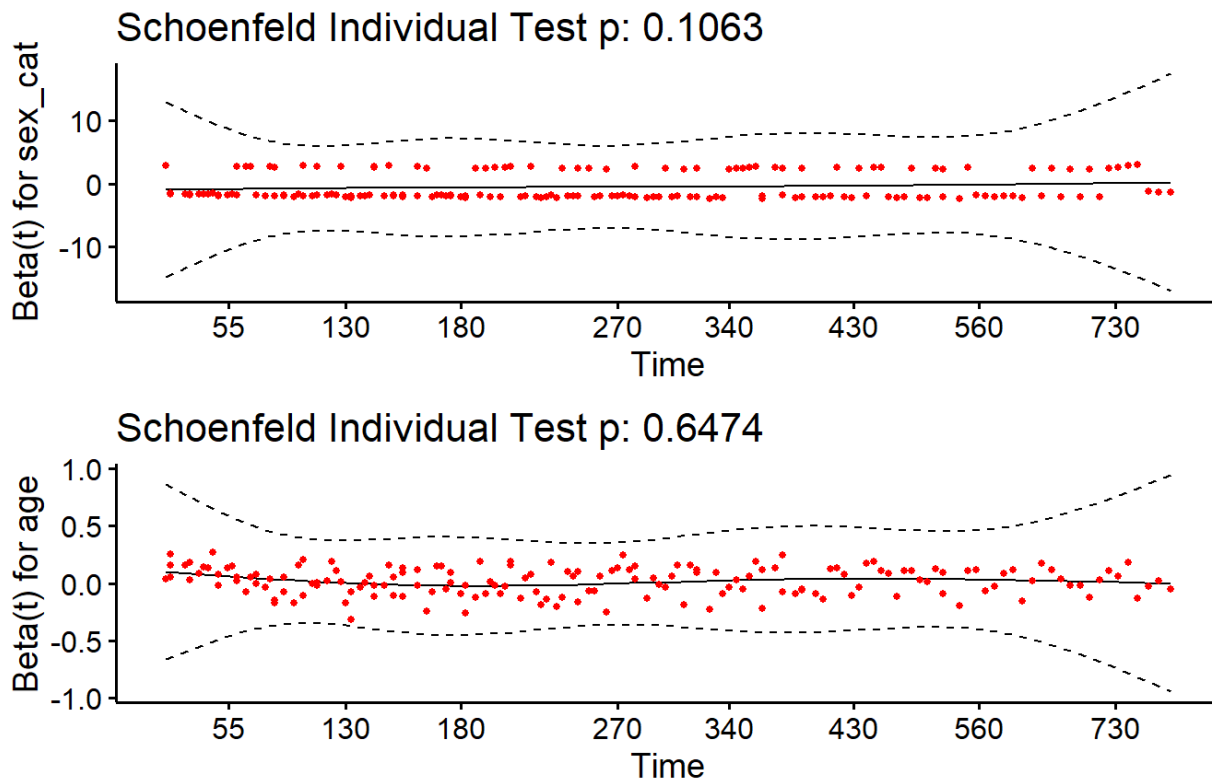
1. O aumento de 1 ano de idade aumenta o risco de morte em 2
2. O sexo feminino possuem um risco 40% (1-0,6) menor quando comparamos com pacientes masculinos

```
## Call:
## survival::coxph(formula = Surv(time, status_adj) ~ sex_cat +
##   age, data = dados)
##
##   n= 228, number of events= 165
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## sex_catMulheres -0.513219  0.598566  0.167458 -3.065  0.00218 **
## age              0.017045  1.017191  0.009223  1.848  0.06459 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## sex_catMulheres    0.5986    1.6707    0.4311    0.8311
## age                1.0172    0.9831    0.9990    1.0357
##
## Concordance= 0.603 (se = 0.025 )
## Likelihood ratio test= 14.12 on 2 df,  p=9e-04
## Wald test              = 13.47 on 2 df,  p=0.001
## Score (logrank) test = 13.72 on 2 df,  p=0.001
```

Precisamos agora validar o modelo, garantindo que ele está correto e que podemos confiar nos dados, ou seja, o modelo não viola os pressupostos de validação do método

No gráfico abaixo, nota-se que os dados estão dentro da faixa de confiança, indicando que o modelo está bem ajustado.

Global Schoenfeld Test p: 0.2502



E por último um resumo do modelo, onde esperamos encontrar um valor de concordância maior que 0.6. Então temos um valor de concordância do modelo de 0.63, o que indica que o modelo está correto.

```
## Call:
## survival::coxph(formula = Surv(time, status_adj) ~ sex_cat +
##   age, data = dados)
##
##   n= 228, number of events= 165
##
##               coef exp(coef)  se(coef)      z Pr(>|z|)
## sex_catMulheres -0.513219  0.598566  0.167458 -3.065  0.00218 **
## age              0.017045  1.017191  0.009223  1.848  0.06459 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## sex_catMulheres    0.5986    1.6707    0.4311    0.8311
## age                1.0172    0.9831    0.9990    1.0357
##
## Concordance= 0.603 (se = 0.025 )
## Likelihood ratio test= 14.12 on 2 df,  p=9e-04
## Wald test               = 13.47 on 2 df,  p=0.001
## Score (logrank) test = 13.72 on 2 df,  p=0.001
```

Portanto é válido e correto dizer que a taxa de sobrevivência entre os grupos é diferente, sendo que a maior taxa é do grupo das mulheres.