

# Lição de Casa 2

James Hunter, Ph.D.

2 de outubro de 2020

Nesta lição da casa, vamos trabalhar com algumas problemas verdadeiras e alguns simplificados. As respostas precisam ser submetidas antes de **16 de outubro** por email: jameshunterbr@gmail.com.

Os dados ficam no arquivo “melanoma\_raw.rds” no GitHub.

## Resumo dos Dados

Dados são medidas feitas em pacientes com melanoma maligno. Cada paciente teve o tumor cirurgicamente removido no Departamento de Cirurgia Plástica no Hospital Universitário de Odense, Dinamarca durante 1962 até 1977. Entre as medidas foram a espessura do tumor se uma úlcera for presente ou não. Os pesquisadores quiseram determinar se essas duas características (espessura grande e presença de ulceração) aumentou a probabilidade de morte por causa de melanoma. Pacientes foram seguidos até o fim de 1977.

### Fonte:

Angelo Canty and Brian Ripley (2019). boot: Bootstrap R (S-Plus) Functions. R package version 1.3-23. Dados vêm de: Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993) **Statistical Models Based on Counting Processes**. Springer-Verlag.

## Dicionário dos Dados do Arquivo melanoma\_raw.rds

- **time** (tempo) (num)
  - Tempo de sobrevivência em dias desde a cirurgia, possivelmente censurado.
- **status** (estado) (num)
  - O estado dos pacientes ao final do estudo. “1” - morreu da melanoma “2” - vivo “3” - morreu de uma causa não relacionada à melanoma.
- **sex** (gênero) (num)
  - “1” - masculino
  - “2” - feminino
- **age** (idade) (num)
  - Idade em anos na data da cirurgia
- **year** (ano) (num)
  - Ano da cirurgia
- **thickness** (espessura) (num)
  - Espessura do tumor em mm
- **ulcer** (úlcer) (num)
  - “1” - úlcera presente
  - “2” - úlcera ausente

## Trabalho Preliminar

Antes de montar os gráficos, vai precisar fazer um pouco de limpeza de dados. Quais variáveis são realmente categóricas e devem estar traduzidas aos factors?

## Problema 1

### Parte A - Gráficos

#### Gráfico 1

Qual é a distribuição das idades dos pacientes no estudo? Mostre através de uma histograma construída com `ggplot` e a `geom_histogram()`.

#### Gráfico 2

Qual é a relação entre tempo de sobrevivência (`tempo`) e idade? Existe uma associação? Mostre usando um scatterplot construído com `ggplot` e `geom_point()`.

#### Gráfico 3

Existe uma diferença entre as espessuras dos tumores dos homens e mulheres? Faça um boxplot mostrando essa diferença, incluindo os pontos dos pacientes. Utilize `ggpubr` e sua função `ggboxplot`.

#### Gráfico 4

Existe diferenças a sobrevivência das pessoas baseada no ano em que os pacientes fizeram a cirurgia? Mostre isso com scatterplot que inclui uma linha de tendência (`geom_smooth(method = "lm")`). Também, deve incluir o gênero do paciente como cor. Pode usar qualquer umas das funções que aprenderam. NB, este gráfico é mais complicado que os outros. Planeje ele bem antes de sentar em frente do RStudio.

### Parte B - Estatística Descritiva

Use as funções `freq()` e `descr()` de `summarytools` para fazer um resumo dos variáveis de `melanoma`.

## Problema 2

Existe uma diferença entre a sobrevivência das mulheres e homens depois da cirurgia? Também pode usar um t-test.

## Problema 3

Reorganize a variável `status` para diferenciar entre óbito por melanoma e outros resultados. A pergunta é se mais pessoas têm óbito se tiverem tumores ulcerados.

## Problema 4

Conduzir uma regressão linear utilizando `lm()` da relação possível entre idade e tempo de sobrevivência. Mostrar o resultado com `summary()`. Extrair os coeficientes com `broom::tidy()`.