

Lição de Casa 3

James R. Hunter

26-11-2024

Nesta lição da casa, vamos trabalhar com gráficos para ajudar entender o conjunto dos dados “melanoma”. Os dados ficam no arquivo “melanoma_raw.rds” no GitHub/Classroom.

Resumo dos Dados

Dados são medidas feitas em pacientes com melanoma maligna. Cada paciente teve o tumor cirurgicamente removido no Departamento de Cirurgia Plástica no Hospital Universitário de Odense, Dinamarca de 1962 até 1977. As medidas principais eram a espessura do tumor e 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** (úlceras) (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/Gráfico 1a

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

Problema/Gráfico 1b

Mostre a distribuição das idades agrupado por sex através de um gráfico de densidade e depois um de violino. Qual comunica melhor a informação sobre idade?

Problema/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()`.

Problema/Gráfico 3

Existe uma diferença entre as espessuras dos tumores dos homens e mulheres? Faça um plot mostrando essa diferença, incluindo os pontos dos pacientes. Utilize `ggstatsplot` e sua função `ggbetweenstats()`.

Problema/Gráfico 4

Existe diferenças entre 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.