

script_projeto

AUTHOR
nome aqui

meu script, meu projeto

Porque desta atividade?

Praticar as habilidades ensinadas no curso em projeto com dados reais de pesquisa na área de saúde, neste exemplo sobre Câncer de Pâncreas.

Nesta pesquisa queremos averiguar as associações entre o desfecho Câncer de Pâncreas e os hábitos que foram levantados (idade, sexo, bebe café, fuma) na coorte de indivíduos.

Dicas aulas -

O que fazer com um banco de dados de pesquisa?

Realizar transformações versáteis que facilitam análises de dados para diferentes estudos:

- Manipulação
- Análise descritiva
- Avaliar e testar modelos estatísticos

Como

Seguir as instruções abaixo:

- Cada discente terá seu arquivo individualizado, você precisa **baixar o arquivo com seu nome da página do curso** ;
- **Criar um projeto no Rstudio**, contendo o arquivo de dados do exercício com seu nome e
- **Fazer um script/quarto** realizando as tarefas abaixo
 - Vamos seguir o fluxo de ideias com perguntas...

Manipulando os dados

Seus dados estão adequados? Tem “sujeira”?

Umas das funções mais frequentes em projetos é transformar os dados:

1. Seus dados já estão limpos, campos categóricos foram transformados em numéricos para análises estatísticas?

1. Você tivessem que fazer esse processo, como fariam?

Dicas aulas - 10, 33.

- Como esta?

| | id | sexo | idade | cafe | cigarro | cancerpancreas | idademdecenios |
|----|----|-----------|---------|------|---------|----------------|----------------|
| 1 | 1 | Masculino | 34.9855 | Sim | Sim | Sim | 3.49855 |
| 2 | 2 | Feminino | 85.7051 | Sim | Sim | Sim | 8.57051 |
| 3 | 3 | Feminino | 64.5085 | Sim | Sim | Sim | 6.45085 |
| 4 | 4 | Masculino | 57.7650 | Sim | Sim | Sim | 5.77650 |
| 5 | 5 | Feminino | 83.5876 | Sim | Sim | Sim | 8.35876 |
| 6 | 6 | Masculino | 37.2958 | Sim | Sim | Sim | 3.72958 |
| 7 | 7 | Feminino | 79.8193 | Sim | Sim | Sim | 7.98193 |
| 8 | 8 | Feminino | 91.2272 | Sim | Sim | Sim | 9.12272 |
| 9 | 9 | Masculino | 76.1706 | Sim | Sim | Sim | 7.61706 |
| 10 | 10 | Masculino | 35.1937 | Sim | Sim | Sim | 3.51938 |
| 11 | 11 | Feminino | 46.6645 | Sim | Sim | Sim | 4.66645 |
| 12 | 12 | Masculino | 95.2634 | Sim | Sim | Sim | 9.52634 |
| 13 | 13 | Feminino | 80.7825 | Sim | Sim | Sim | 8.07825 |
| 14 | 14 | Feminino | 78.0244 | Sim | Sim | Sim | 7.80244 |
| 15 | 15 | Feminino | 88.5654 | Sim | Sim | Sim | 8.85653 |
| 16 | 16 | Feminino | 93.9592 | Sim | Sim | Não | 9.39592 |
| 17 | 17 | Masculino | 52.5677 | Sim | Sim | Sim | 5.25677 |
| 18 | 18 | Feminino | 62.9983 | Sim | Sim | Sim | 6.29983 |
| 19 | 19 | Masculino | 43.2603 | Sim | Sim | Sim | 4.32603 |
| 20 | 20 | Feminino | 58.3441 | Sim | Sim | Sim | 5.83441 |

- Como deve ficar?

| | id | sexo | idade | cafe | cigarro | cancerpancreas | idademdecenios |
|----|----|------|-------|------|---------|----------------|----------------|
| 1 | 1 | 0 | 34 | 1 | 1 | 1 | 3 |
| 2 | 2 | 1 | 85 | 1 | 1 | 1 | 8 |
| 3 | 3 | 1 | 64 | 1 | 1 | 1 | 6 |
| 4 | 4 | 0 | 57 | 1 | 1 | 1 | 5 |
| 5 | 5 | 1 | 83 | 1 | 1 | 1 | 8 |
| 6 | 6 | 0 | 37 | 1 | 1 | 1 | 3 |
| 7 | 7 | 1 | 79 | 1 | 1 | 1 | 7 |
| 8 | 8 | 1 | 91 | 1 | 1 | 1 | 9 |
| 9 | 9 | 0 | 76 | 1 | 1 | 1 | 7 |
| 10 | 10 | 0 | 35 | 1 | 1 | 1 | 3 |
| 11 | 11 | 1 | 46 | 1 | 1 | 1 | 4 |
| 12 | 12 | 0 | 95 | 1 | 1 | 1 | 9 |
| 13 | 13 | 1 | 80 | 1 | 1 | 1 | 8 |
| 14 | 14 | 1 | 78 | 1 | 1 | 1 | 7 |
| 15 | 15 | 1 | 88 | 1 | 1 | 1 | 8 |
| 16 | 16 | 1 | 93 | 1 | 1 | 0 | 9 |
| 17 | 17 | 0 | 52 | 1 | 1 | 1 | 5 |
| 18 | 18 | 1 | 62 | 1 | 1 | 1 | 6 |
| 19 | 19 | 0 | 43 | 1 | 1 | 1 | 4 |
| 20 | 20 | 1 | 58 | 1 | 1 | 1 | 5 |

2. Desafio: criar coluna com faixas etárias e agrupar

1. Adicionar colunas novas?

Dicas aulas - 24

| | id | sexo | idade | cafe | cigarro | cancerpancreas | idademdecenios | faixa_etaria |
|----|----|------|-------|------|---------|----------------|----------------|--------------|
| 1 | 1 | 0 | 34 | 1 | 1 | 1 | 3 | 30-39 |
| 2 | 2 | 1 | 85 | 1 | 1 | 1 | 8 | 80-89 |
| 3 | 3 | 1 | 64 | 1 | 1 | 1 | 6 | 60-69 |
| 4 | 4 | 0 | 57 | 1 | 1 | 1 | 5 | 50-59 |
| 5 | 5 | 1 | 83 | 1 | 1 | 1 | 8 | 80-89 |
| 6 | 6 | 0 | 37 | 1 | 1 | 1 | 3 | 30-39 |
| 7 | 7 | 1 | 79 | 1 | 1 | 1 | 7 | 70-79 |
| 8 | 8 | 1 | 91 | 1 | 1 | 1 | 9 | 90-99 |
| 9 | 9 | 0 | 76 | 1 | 1 | 1 | 7 | 70-79 |
| 10 | 10 | 0 | 35 | 1 | 1 | 1 | 3 | 30-39 |
| 11 | 11 | 1 | 46 | 1 | 1 | 1 | 4 | 40-49 |
| 12 | 12 | 0 | 95 | 1 | 1 | 1 | 9 | 90-99 |
| 13 | 13 | 1 | 80 | 1 | 1 | 1 | 8 | 70-79 |
| 14 | 14 | 1 | 78 | 1 | 1 | 1 | 7 | 70-79 |
| 15 | 15 | 1 | 88 | 1 | 1 | 1 | 8 | 80-89 |
| 16 | 16 | 1 | 93 | 1 | 1 | 0 | 9 | 90-99 |
| 17 | 17 | 0 | 52 | 1 | 1 | 1 | 5 | 50-59 |
| 18 | 18 | 1 | 62 | 1 | 1 | 1 | 6 | 60-69 |
| 19 | 19 | 0 | 43 | 1 | 1 | 1 | 4 | 40-49 |
| 20 | 20 | 1 | 58 | 1 | 1 | 1 | 5 | 50-59 |

Analises descritivas

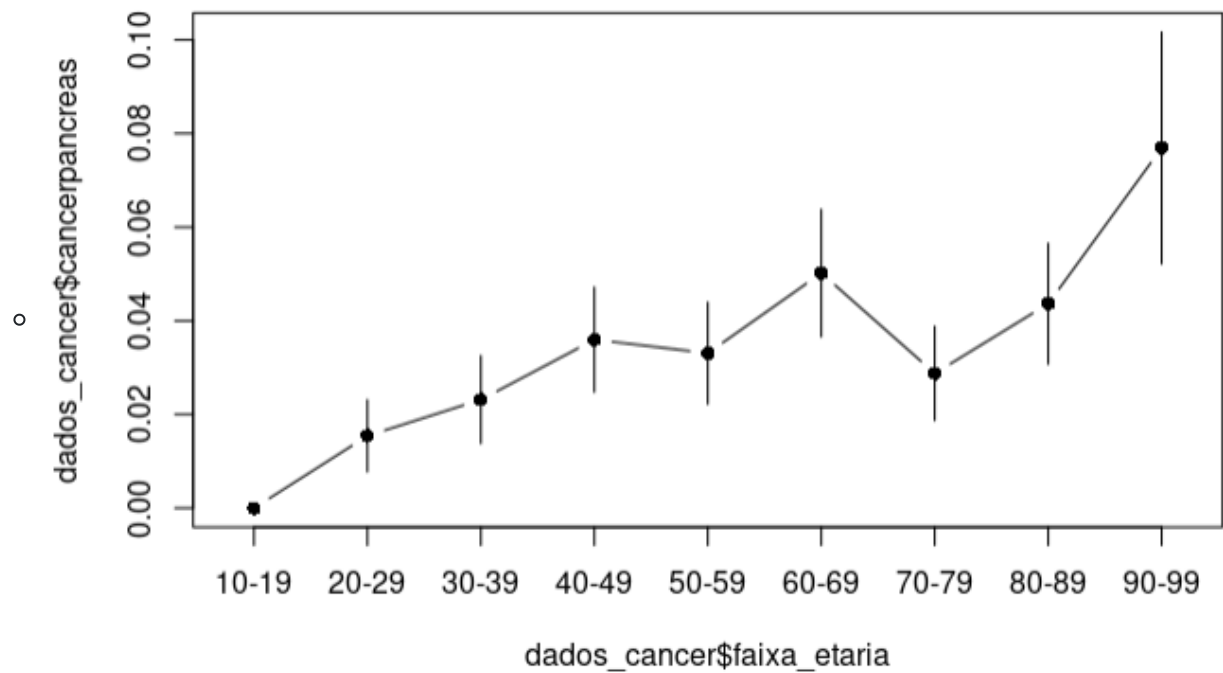
Já abordamos diversas bibliotecas para análises exploratórias no módulo 1 e após limpar os dados este é um bom momento para fazer este tipo de abordagem.

Nosso objetivo neste trabalho é seguir além, para esta segunda etapa **comecem a verificar a interação entre as variáveis:**

(sexo e cancer | café e cancer | faixas etárias e cancer)

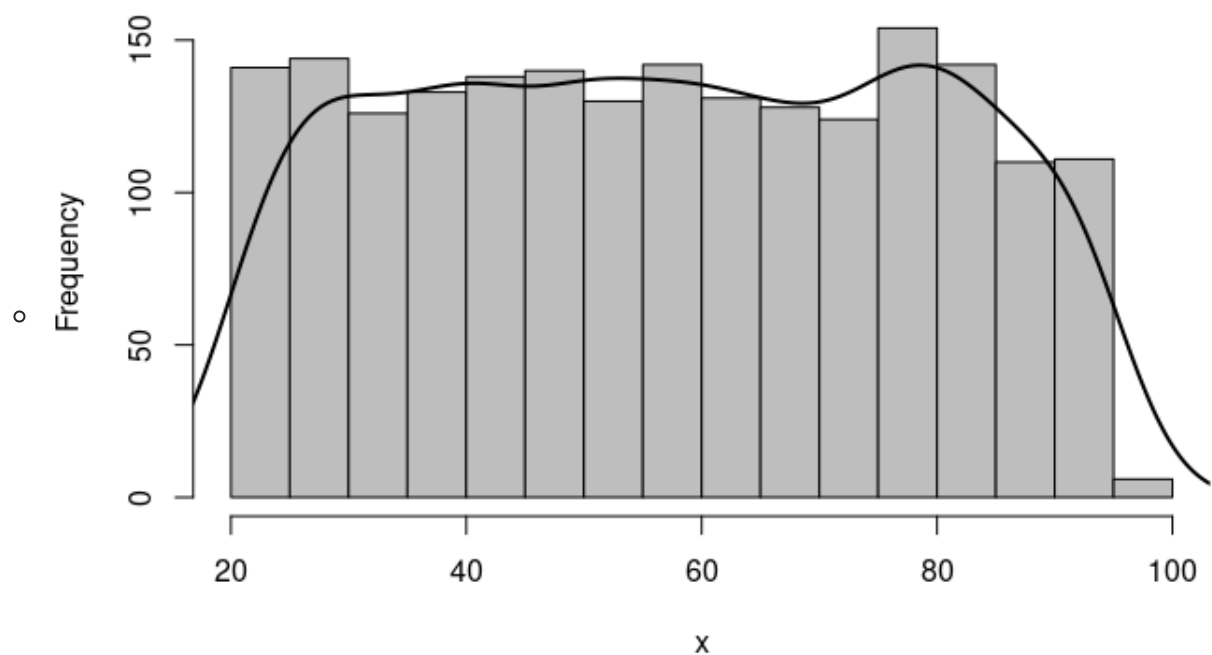
Dicas Aulas - 14, 15, 19-1 e 19-2

- Seus dados podem ficar diferentes, apenas um exemplo de referência:



- Será que a sua população amostral é uniforme?
responda qual o tipo de população quanto a idade e sexo

Dicas aulas - 17, 18



Risco Relativo (RR)

Dicas aulas - 25, 26, 27, 28, 29

1. Criar tabelas de contingências com as variáveis de interesse

Dica extra - nestas horas podemos mesclar ou voltar a usar as tabelas originais para facilitar a criação das tabelas novas de contingência entre os dados inteiros e categóricos

criação das tabelas novas de correlação entre os dados numéricos e categóricos.

Avaliar modelos logísticos para câncer no pâncreas

Dicas aulas -30, 35, 37, 38, 41,

1. Qual das variáveis tem maior correlação com o câncer de pâncreas?

podem usar as aulas de regressão `lm()`

Dicas aulas - 48, 49 e 50

2. Apresentar cálculo de regressão logística

podem usar exemplo da aula de regressão `glm()`

Colocar um parágrafo de conclusão no seu relatório, usando seus dados, com as correlações, ou regressões ou valor-p.

Não estamos esperando um modelo de resposta ideal, o objetivo é vocês praticarem o método científico e aplicar os aprendizados no R, gerando um relatório ao final (no RStudio clicar no botão Render).

Devem usar este arquivo como referência.

Entregar o arquivo com seu script “script_projeto_Maria”, chamando os dados com seu Maria.csv, por exemplo. Aproveitem as ‘dicas das aulas’ que dão as referências para os exercícios.

Bom trabalho e semana.