

Métodos Quantitativos IV - Lista 6 (resolução)

Departamento de Ciência Política da FFLCH-USP

Luiz Henrique da Silva Batista (Número USP: 12687228)

2023-12-19

```
# Material de apoio para esta lista:  
# https://jonnyphillips.github.io/Analise\_de\_Dados\_2022/
```

Carregando pacotes

```
library(tidyverse)  
library(janitor)  
library(stargazer)  
library(vroom)  
library(knitr)
```

Exercício 1

Nesta lista, utilizaremos os microdados do ENEM relativos a 2022. Faça o download dos dados do ENEM de 2022 e do dicionário de dados. Leia o dicionário de dados

```
# Link: https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/  
# microdados/enem  
  
# Selecione as seguintes variáveis:  
## NU_NOTA_CH  
## NU_NOTA_MT  
## NU_NOTA_CN  
## NU_NOTA_LC  
## TP_SEXO  
## TP_COR_RACA  
## TP_ESCOLA
```

Limpe os dados para a análise. Para resolver as questões, utilize uma amostra aleatória com 10.000 observações. Lembre-se de utilizar a função `set.seed()`. Verifique como as variáveis estão codificadas e faça os ajustes necessários para a análise.

```
set.seed(123)

# setwd(dir = "C:\\sua\\estrutura\\de\\pastas\\no\\windows")
# setwd(dir = "C:/sua/estrutura/de/pastas/no/linux/ou/mac")

# Importando os dados
df <- vroom("datasets/microdados_enem_2022/DADOS/MICRODADOS_ENEM_2022.csv",
            delim = ";")

# Limpando os dados
dados <- df |>
  # Limpa o nome das colunas
  janitor::clean_names() |>
  # Seleciona somente as colunas de interesse
  select(nu_notas_ch, nu_notas_mt, nu_notas_cn, nu_notas_lc,
         tp_sexo, tp_cor_raca, tp_escola) |>
  # Remove valores ausentes
  drop_na() |>
  # Seleciona 10 mil casos aleatorios
  sample_n(size = 10000)

# Verificando a codificação das variáveis
# glimpse(dados)
```