

Proval Parcial 1 - Parte 2 - 23/10/2022

Amanda Freitas Carnaiba - NUSP 13485660

Para responder esta parte, deve-se tomar o banco de dados que está junto com este arquivo de enunciado para a prova (Base_prova_parte2.csv).

```
base <- read.csv("base_prova_parte2.csv")
head(base)
```

```
##      votos   financ idade n_part      fpart votos_t1
## 1 296395.2 1603.920   47      2 0.27491429 96987.30
## 2 194729.0 13618.653   31      1 0.02034933 59863.14
## 3 233566.1  7171.793   52      1 0.16806297 55564.86
## 4 283144.1 24071.725   53      1 0.12419878 72222.40
## 5 289448.0  6372.322   52      1 0.29304743 70674.89
## 6 103730.4  5841.561   36      1 0.30477020  4096.82
```

A variável dependente é votos e as demais colunas são todas variáveis explicativas.

Deve-se selecionar uma amostra aleatória de tamanho igual a 500 para esse exercício. Isto deve ser feito utilizando dois comandos no R: `set.seed()` e o `sample_n()`.

Para o primeiro comando, deve-se novamente utilizar o número USP como o seed. Para um número USP como o do exemplo anterior, deve-se digitar no início do código: `set.seed(12345678)` e em seguida `sample_n(500)`.

```
set.seed(13485660)
base_nova <-
  base %>%
  sample_n(500)
```

```
## sample_n: removed 9,500 rows (95%), 500 rows remaining
```

```
dim(base_nova)
```

```
## [1] 500    6
```

De posse deste novo banco de dados, deve-se buscar encontrar um modelo de bom ajuste para a previsão dos votos. O livro de código das variáveis está resumido a seguir:

- Votos: Votação total recebida por um candidato a deputado federal na eleição no ano t.
- financ: Valor (em milhares) declarado pelo candidato da arrecadação da campanha,
- idade: Idade do candidato (em anos) no dia 31/07 do ano t.
- Fpart: Proporção dos recursos de campanha do candidato recebida do partido.
- Votos_t1: Votação total recebida por um candidato a deputado federal na eleição no ano t-1.

Em sua resposta, você deve apresentar o modelo de MQO que melhor se adéqua aos dados selecionados. Se for necessário, considere a situação de apresentar mais de um modelo. Apresente também gráficos e estatísticas descritivas das variáveis que contribuam na sua decisão sobre qual o melhor modelo. Teste para a validade das hipóteses do modelo de MQO e mostre os resultados obtidos. Apresente ao final as primeiras 25 linhas do seu banco de dados também.