

Untitled

2023-11-28

Essa a lista compreende questões sobre a verificação dos pressupostos do modelo de regressão linear.

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

Carregando pacotes

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

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/aceso-a-informacao/dados-abertos/microdados/enem  
# Selecione as variáveis NU_NOTA_CH e NU_NOTA_MT
```

Limpe os dados para a análise.

```
enem_2022 <- vroom::vroom("datasets/microdados_enem_2022/DADOS/MICRODADOS_ENEM_2022.csv",  
                          quote = "", delim = ";") |>  
  # Coloca em letras minúsculas os nomes das colunas  
  janitor::clean_names() |>  
  # Seleciona somente as variáveis de interesse  
  dplyr::select(nu_inscricao, nu_notas_ch, nu_notas_mt)
```

Para a leitura da nossa base de dados utilizamos a função `vroom()`, do pacote `{vroom}`. A motivação se deve ao fato de que a velocidade de leitura dessa função é de 1,23GB/s, tornando-se excelente para importar bases grandes. Para mais informações sobre esse pacote, acesse este link.

Exercício 2

Utilize a nota do ENEM em ciências humanas para prever a nota em matemática por meio de uma regressão linear simples. Interprete os resultados.

```
reg <- lm(nu_nota_mt ~ nu_nota_ch, data = enem_2022)

stargazer::stargazer(reg,
  type = "latex",
  style = "ajps",
  title = "Regressão linear",
  single.row = T,
  keep.stat = c("n"),
  dep.var.labels = "Nota em matemática",
  header = F,
  dep.var.caption= "")
```

Table 1: Regressão linear

	Nota em matemática
nu_nota_ch	0.905*** (0.001)
Constant	62.975*** (0.399)
N	2344823

***p < .01; **p < .05; *p < .1

Formalmente, o modelo pode ser representado pela seguinte expressão:

$$Nota\ em\ matemática = \alpha + \beta \cdot X_{Nota\ em\ ciências\ humanas}$$

Nosso modelo de regressão estima que a nota de matemática aumenta, em média, cerca de 0.905 pontos para cada 1 (um) ponto a mais obtido na prova de ciências humanas. Além disso, quando a nota em ciências humanas é zero temos uma nota de matemática esperada de 62.975.

Exercício 3

Calcule os intervalos de confiança dos coeficientes.

Intervalo de confiança de 95% para α	Intervalo de confiança de 95% para β
(62.177; 62.975 + 2 * 0.399)	(0.903; 0.907)

Por que o R utiliza a distribuição t (e não a distribuição normal) para as inferências?

Exercício 4

Utilize os resíduos para testar graficamente o seguinte: $E[\hat{\epsilon}|X = x]$. Explique a sua conclusão.

Qual é a motivação desse teste? Em outras palavras, qual pressuposto do modelo linear gostaríamos de testar?

Qual é a consequência da violação desse pressuposto?

Exercício 5

Com base nos resíduos, utilize um gráfico para testar a hipótese de homocedasticidade dos erros. Explique a sua conclusão.

Qual é a consequência da violação desse pressuposto?

Exercício 6

Com base nos resíduos, utilize um gráfico para testar a hipótese de normalidade dos erros. Explique a sua conclusão.

Qual é a consequência da violação desse pressuposto?

Exercício 7

Apresente seus resultados em um arquivo PDF. Garanta que seu arquivo esteja limpo, contendo as respostas, os gráficos e as tabelas, mas não eventuais mensagens e erros. O arquivo PDF pode ser gerado diretamente a partir do R por meio do RMarkdown ou do RSweave. Para os alunos de graduação, isso é recomendado, mas não obrigatório. Adicionalmente, forneça o script para replicação.