

EXERCÍCIOS COM AUXÍLIO DO COMPUTADOR

Para realizar os exercícios descritos neste texto você deve baixar o arquivo de dados [amostraEnem2019](#) e basear-se no texto [Análise Estatística no Ambiente Computacional R / R-Commander](#).

Exercício 1. (Com base na matéria do Capítulo 3.)

Da amostra aleatória simples de 2.000 avaliados no Enem 2019, fazer:

- a) Tabela de frequência da *cor/raça* do candidato.
- b) Classifique o nível socioeconômico (NSE) em dois níveis: nível 1 (*até cinco*) e nível 2 (*acima de cinco*). Faça uma tabela de frequências para essa variável categorizada.
- c) Tabela de contingência entre *cor/raça* e a classificação do NSE que você fez no item anterior. Refaça a tabela calculando a porcentagem de cada categoria de *cor/raça* para os níveis 1 e 2 de NSE. Interprete os resultados.
- d) Histograma de frequências para a nota em Língua Portuguesa (NOTA_LP) e para a nota em Matemática (NOTA_MT). O que você diz sobre a simetria dessas distribuições?
- e) Média, desvio padrão, mediana e quartis da nota em Língua Portuguesa para os níveis 1 e 2 de NSE. Interprete os resultados.
- f) Diagrama em caixas da nota em Ciências da Natureza (NOTA_CN) para cada *cor/raça* dos candidatos. Interprete o gráfico.

Exercício 2. Com base na matéria do Capítulo 5.

- a) Fazer gráficos das distribuições de probabilidades binomial com $n = 10$ e $\pi = 0,5$ e binomial com $n = 10$ e $\pi = 0,2$. Discutir sobre a simetria dessas distribuições.
- b) Uma urna contém 500 bolas brancas e 500 bolas pretas. Retira-se, aleatoriamente, dez bolas da urna. Seja X a quantidade de bolas brancas na amostra. Apresente a distribuição de probabilidades de X pela distribuição exata (hipergeométrica) e pela aproximação binomial. Compare os resultados.
- c) O número de ocorrências de falhas por dia num sistema eletrônico foi observado durante cem dias, resultando na seguinte distribuição de

frequências:

Número de ocorrências	0	1	2	3	4	5	6
Frequência	15	26	27	20	8	3	1

Obtenha as probabilidades de uma distribuição de Poisson com taxa média de ocorrências igual a média da distribuição de frequências do número de ocorrências de falhas diária. As distribuições são parecidas?

Exercício 3. Com base na matéria do Capítulo 6.

- a) Fazer histograma e gráfico de probabilidade normal para as notas (proficiências) obtidas pelos candidatos em cada área. Quais são mais parecidas com a forma de uma distribuição normal?
- b) Fazer gráficos para as funções de probabilidade binomial com parâmetros:
- i. $n = 20, \pi = 0,5$

- ii. $n = 20, \pi = 0,8$
- iii. $n = 500, \pi = 0,8$

Em quais casos a distribuição tem forma parecida com a normal?

Exercício 4. Com base na matéria do Capítulo 7.

- a) Gere 10.000 observações de uma distribuição uniforme em $[0, 1]$. Faça um histograma e observe a sua forma. Calcule a média e a variância dessa amostra e compare com os resultados teóricos do valor esperado e da variância de uma distribuição de probabilidades uniforme em $[0, 1]$.
- b) Gere 1.000 amostras de tamanho $n = 100$ de uma distribuição uniforme em $[0, 1]$. Calcule a média de cada amostra. Faça um histograma dessas médias. A forma desse histograma parece com distribuição uniforme ou normal? Calcule o desvio padrão dessas 1.000 médias e compare com o valor teórico do erro padrão da média.
- c) Para este exercício você deve usar o arquivo *amostraEnem2019*, disponível em Materiais Suplementares. Apresente e interprete intervalos de 95% de confiança para a média dos candidatos do Enem 2019 em cada uma das cinco provas.

Exercício 5. Com base na matéria do Capítulo 8.

Para cada área do Enem, as notas são padronizadas com média 500 e desvio padrão 100 tendo como referência de candidatos concluintes de escolas regulares do Enem de 2009. Baseado na amostra de 2.000 candidatos do Enem de 2019, teste se a média das notas de cada área dos candidatos do Enem 2019 pode ser considerada igual a 500.

Exercício 6. Com base na matéria do Capítulo 9.

- a) Para cada uma das cinco notas do Enem 2019, teste, ao nível de significância de 0,01, se há diferença na nota média de candidatos do sexo feminino e do sexo masculino.
- b) Considerando a nota em Língua Portuguesa (NOTA_LP), teste se há diferença da nota média entre as diferentes cores/raças dos candidatos.

Exercício 7. Com base na matéria do Capítulo 10.

Para cada uma das cinco notas do Enem 2019, teste, usando abordagem não paramétrica, se há diferença na nota mediana entre candidatos do sexo feminino e candidatos do sexo masculino. Use nível de significância de 0,01.

Exercício 8. Com base na matéria do Capítulo 11.

- a) No Enem há quatro notas de acordo com a área do conhecimento e a nota de redação. Com os dados do arquivo *amostraEnem2019* faça um diagrama de dispersão e obtenha o coeficiente de correlação para cada par de notas. Interprete.

- b) Faça uma regressão linear simples para “explicar” a nota em Língua Portuguesa (NOTA_LP) em função do nível socioeconômico (NSE) do candidato. O coeficiente angular é diferente de zero ao nível de significância de 1%? Interprete esse coeficiente. Interprete o R^2 dessa regressão. Faça uma análise de resíduos e avalie a adequação do modelo.
- c) Faça uma regressão para “explicar” a nota em Matemática (NOTA_MT) em função do NSE do candidato. Faça tudo que foi pedido no item anterior para essa regressão.
- d) Faça uma regressão múltipla para “explicar” a nota em Matemática (NOTA_MT) em função da *cor/raça* do candidato. Interprete os coeficientes.
- e) Na regressão do item anterior, inclua também o NSE do candidato como variável explicativa. Os coeficientes das variáveis indicadoras associadas à *cor/raça* alteraram? Explique o porquê disto.