

MAE0217 - Estatística Descritiva - Lista 2

Natalia Koza¹
Rafael Gonçalves Pereira da Silva²
Ricardo Geraldês Tolesano³
Rubens Kushimizo Rodrigues Xavier⁴
Rubens Gomes Neto⁵
Rubens Santos Andrade Filho⁶
Thamires dos Santos Matos⁷

Maio de 2021

Sumário

Exercício 1	2
Exercício 12	2
Exercício 14	2
Exercício 15	2
Exercício 17	2
Exercício 19	3
Exercício 23	3
Exercício 28	4
Exercício 30	4
Exercício 33	4

¹Número USP: 10698432

²Número USP: 9009600

³Número USP: 10734557

⁴Número USP: 8626718

⁵Número USP: 9318484

⁶Número USP: 10370336

⁷Número USP: 9402940

Exercício 1

O arquivo **rehabcardio** contém informações sobre um estudo de reabilitação de pacientes cardíacos. Elabore um relatório indicando possíveis inconsistências na matriz de dados e faça uma análise descritiva de todas as variáveis do estudo, construindo distribuições de frequências para as variáveis qualitativas e obtendo medidas resumo para as variáveis qualitativas.

Exercício 12

Construa gráficos de quantis e de simetria para os dados de manchas solares disponíveis no arquivo **manchas**.

Exercício 14

Uma outra medida de assimetria é

$$A = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$$

que é igual a zero no caso de distribuições simétricas. Calcule-a para os dados do Exercício 6 .

Exercício 15

Considere os valores X_1, \dots, X_n de uma variável X , com média \bar{X} desvio padrão S . Mostre que a variável Z , cujos valores são $Z_i = (X_i - \bar{X})/S, i = 1, \dots, n$ tem média 0 e desvio padrão 1.

Exercício 17

Os dados apresentados na Tabela 1 referem-se aos instantes nos quais o centro de controle operacional de estradas rodoviárias recebeu chamados solicitando algum tipo de auxílio em duas estradas num determinado dia.

Estrada 1	12 : 07 : 00AM	12 : 58 : 00AM	01 : 24 : 00AM	01 : 35 : 00AM	02 : 05 : 00AM
	03 : 14 : 00AM	03 : 25 : 00AM	03 : 46 : 00AM	05 : 44 : 00AM	05 : 56 : 00AM
	06 : 36 : 00AM	07 : 26 : 00AM	07 : 48 : 00AM	09 : 13 : 00AM	12 : 05 : 00PM
	12 : 48 : 00PM	01 : 21 : 00PM	02 : 22 : 00PM	05 : 30 : 00PM	06 : 00 : 00PM
	07 : 53 : 00PM	09 : 15 : 00PM	09 : 49 : 00PM	09 : 59 : 00PM	10 : 53 : 00PM
	11 : 27 : 00PM	11 : 49 : 00PM	11 : 57 : 00PM		
Estrada 2	12 : 03 : 00AM	01 : 18 : 00AM	04 : 35 : 00AM	06 : 13 : 00AM	06 : 59 : 00AM
	08 : 03 : 00AM	10 : 07 : 00AM	12 : 24 : 00PM	01 : 45 : 00PM	02 : 07 : 00PM
	03 : 23 : 00PM	06 : 34 : 00PM	07 : 19 : 00PM	09 : 44 : 00PM	10 : 27 : 00PM
	10 : 52 : 00PM	11 : 19 : 00PM	11 : 29 : 00PM	11 : 44 : 00PM	

Tabela 1: Planilha com instantes de realização de chamados solicitando auxílio em estradas.

- Construa um histograma para a distribuição de frequências dos instantes de chamados em cada uma das estradas.
- Calcule os intervalos de tempo entre as sucessivas chamadas e descreva-os, para cada uma das estradas, utilizando medidas resumo e gráficos do tipo boxplot. Existe alguma relação entre o tipo de estrada e o intervalo de tempo entre as chamadas?
- Por intermédio de um gráfico do tipo QQ, verifique se a distribuição da variável “Intervalo de tempo entre as chamadas” em cada estrada é compatível com um modelo normal. Faça o mesmo para um modelo exponencial. Compare as distribuições de frequências correspondentes às duas estradas.

Exercício 19

Considere o seguinte resumo descritivo da pulsação de estudantes com atividade física intensa e fraca:

Atividade	N	Média	Mediana	DP	Min	Max	Q1	Q3
Intensa	30	79,6	82	10,5	62	90	70	85
Fraca	30	73,1	70	9,6	58	92	63	77

DP: desvio padrão, Q1: primeiro quartil, Q3: terceiro quartil

Indique se as seguintes afirmações estão corretas, justificando a sua respostas:

- 5% e 50% dos estudantes com atividade física intensa e fraca, respectivamente, tiveram pulsação inferior a 70 .
- A proporção de estudantes com fraca atividade física com pulsação inferior a 63 é menor que a proporção de estudantes com atividade física intensa com pulsação inferior a 70.
- A atividade física não tem efeito na média da pulsação dos estudantes.
- Mais da metade dos estudantes com atividade física intensa têm pulsação maior que 82 .

Exercício 23

Num conjunto de dados, o primeiro quartil é 10, a mediana é 15 e o terceiro quartil é 20. Indique quais das seguintes afirmativas são verdadeiras, justificando sua resposta:

- A distância interquartis é 5 .
- O valor 32 seria considerado outlier segundo o critério utilizado na construção do boxplot.
- A mediana ficaria alterada de 2 unidades se um ponto com valor acima do terceiro quartil fosse substituído por outro 2 vezes maior.
- O valor mínimo é maior do que zero.

Exercício 28

Os dados encontrados no arquivo **esforco** são provenientes de um estudo sobre teste de esforço cardiopulmonar em pacientes com insuficiência cardíaca. As variáveis medidas durante a realização do teste foram observadas em quatro momentos distintos: repouso (REP), limiar anaeróbio (LAN), ponto de compensação respiratório (PCR) e pico (PICO). As demais variáveis são referentes às características demográficas e clínicas dos pacientes e foram registradas uma única vez.

- Descreva a distribuição da variável consumo de oxigênio (VO_2) em cada um dos quatro momentos de avaliação utilizando medidas resumo (mínimo, máximo, quartis, mediana, média, desvio padrão, etc.), boxplots e histogramas. Você identifica algum paciente com valores de consumo de oxigênio discrepantes? Interprete os resultados.
- Descreva a distribuição da classe funcional NYHA por meio de uma tabela de frequências. Utilize um método gráfico para representar essa tabela.

Exercício 30

Utilizando o mesmo enunciado da questão anterior (29), indique a resposta correta:

- Não é possível estimar nem a média nem a variância com esses dados.
- A variância é menor que 30 .
- A média estimada é 12,8 .
- Em apenas 35% dos casos, o número de vasos é maior que 10.
- Nenhuma das anteriores.

Exercício 33

Com a finalidade de entender a diferença entre “desvio padrão” e “erro padrão”,

- Simule 10000 dados de uma distribuição normal com média 12 e desvio padrão 4. Construa o histograma correspondente, calcule a média e o desvio padrão amostrais e compare os valores obtidos com aqueles utilizados na geração dos dados.
- Simule 500 amostras de tamanho $n = 4$ dessa população. Calcule a média amostral de cada amostra, construa o histograma dessas médias e estime o correspondente desvio padrão (que é o erro padrão da média).
- Repita os passos a) e b) com amostras de tamanhos $n = 9$ e $n = 100$. Comente os resultados comparando-os com aqueles preconizados pela teoria.
- Repita os passos a) - c) simulando amostras de uma distribuição qui-quadrado com 3 graus de liberdade.