

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	3
Exercício 17	4
Exercício 19	4
Exercício 23	4
Exercício 28	5
Exercício 30	5
Exercício 33	6

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.

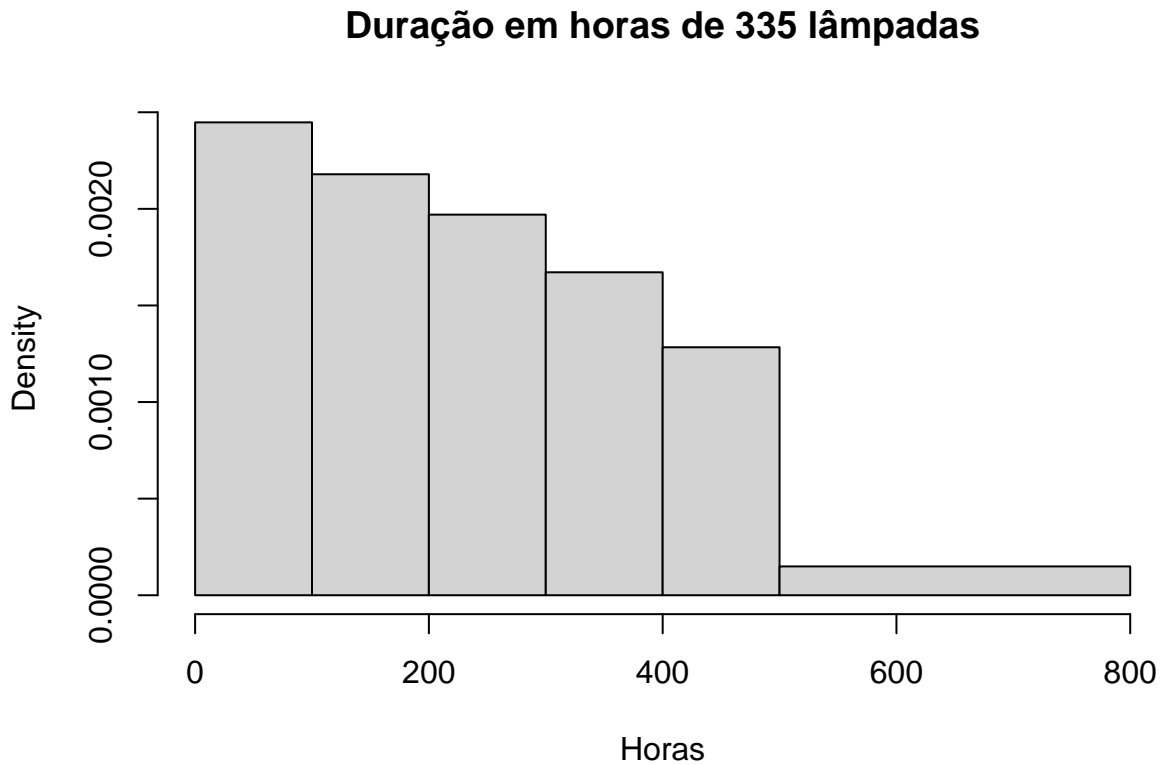
Exercício 12

Exercício 14

Na tabela abaixo estão indicadas as durações de 335 lâmpadas.

Duração(horas)	Número de Lâmpadas
0-100	82
100-200	71
200-300	68
300-400	56
400-500	43
500-800	15

- a) Esboce o histograma correspondente.



b) Calcule os quantis de ordem $p=0,1; 0,3; 0,5; 0,7$ e $0,9$

Exercício 15

Os dados apresentados na Tabela 2 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 2: Planilha com instantes de realização de chamados solicitando auxílio em estradas.

a) Construa um histograma para a distribuição de frequências dos instantes de chamados em cada uma das estradas.

- b) 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?
- c) 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 17

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:

- a) 5% e 50% dos estudantes com atividade física intensa e fraca, respectivamente, tiveram pulsação inferior a 70 .
- b) 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.
- c) A atividade física não tem efeito na média da pulsação dos estudantes.
- d) Mais da metade dos estudantes com atividade física intensa têm pulsação maior que 82 .

Exercício 19

Os histogramas apresentados na Figura 3.35 mostram a distribuição das temperaturas ($^{\circ}\text{C}$) ao longo de vários dias de investigação para duas regiões (R1 e R2). Indique se as afirmações abaixo estão corretas, justificando as respostas:

- a) As temperaturas das regiões R1 e R2 têm mesma média e mesma variância.
- b) Não é possível comparar as variâncias.
- c) A temperatura média da regiões R2 é maior que a de R1.
- d) As temperaturas das regiões R1 e R2 têm mesma média e variância diferentes

Exercício 23

A tabela abaixo representa a distribuição do número de dependentes por empregado de uma determinada empresa.

Dependentes	Frequência
1	40
2	50
3	30
4	20
5	10
Total	150

Nenhuma das alternativas. De fato, a media é igual a 2.4 enquanto a mediana = 2 e moda = 2.

```
x <- x %>%
  mutate(freq=`Frequência`/sum(`Frequência`))

# média
x %>% summarise(media = sum(Dependentes * freq)) %>% pull
```

```
## [1] 2.4
```

```
# mediana
x <- x %>% mutate(freqacum = cumsum(freq))
x %>% summarise(mediana = Dependentes[findInterval(0.5, freqacum)+1]) %>% pull
```

```
## [1] 2
```

```
# moda
x %>% summarise(modas = Dependentes[which.max(freq)]) %>% pull
```

```
## [1] 2
```

Exercício 28

Exercício 30

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.

$$\bar{Z} = 1/n \sum_1^n Z_i$$

$$\bar{Z} = 1/n \sum_1^n (X_i - \bar{X})/S$$

$$\bar{Z} = \frac{1}{S} (1/n \sum_1^n X_i - 1/n \sum_1^n \bar{X})$$

$$\bar{X} = 1/n \sum_1^n X_i \quad n\bar{X} = \sum_1^n n\bar{X}$$

$$\bar{Z} = \frac{1}{S}(\bar{X} - \frac{n\bar{X}}{n})$$

$$\bar{Z} = 0$$

$$\text{dp}(Z) = \sqrt{\text{var}(Z)}$$

$$\text{dp}(Z) = \sqrt{1/n \sum_1^n (Z_i - \bar{Z})^2}$$

$$\bar{Z} = 0$$

$$\text{dp}(Z) = \sqrt{\frac{1}{n} \sum_1^n \frac{X_i - \bar{X}}{S}}$$

$$\text{dp}(Z) = \sqrt{\frac{1}{S^2} \frac{1}{n} \sum_1^n X_i^2 - 2X_i\bar{X} + \bar{X}^2}$$

$$\bar{X}^2 = \frac{1}{n} \sum_1^n X_i^2 \quad \bar{X} = \frac{1}{n} \sum_1^n X_i \quad n\bar{X}^2 = \frac{1}{n} \sum_1^n \bar{X}^2$$

$$\text{dp}(Z) = \sqrt{\frac{1}{S^2} (X_i^2 - 2\bar{X}^2 + \bar{X}^2)}$$

$$\text{dp}(Z) = \sqrt{\frac{1}{S^2} (\bar{X}^2 - \bar{X}^2)}$$

$$\text{dp}(Z) = \sqrt{\frac{S^2}{S^2}}$$

$$\text{dp}(Z) = 1$$

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).

- c) 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.
- d) Repita os passos a) - c) simulando amostras de uma distribuição qui-quadrado com 3 graus de liberdade.