2026/1° SEMESTRE

UNIVERSIDADE FEDERAL DO PARANÁ

DEPARTAMENTO DE ESTATÍSTICA

CE301 - ESTATÍSTICA BÁSICA

MINI TESTE 1 - ESTATÍSTICA DESCRITIVA E EXPLORATÓRIA

1. Uma pesquisa foi realizada com estudantes universitários, coletando as seguintes variáveis: curso (A ou B); gênero (F ou M); tempo de estudo por dia (em horas); número de disciplinas cursadas. Os dados observados foram:

Curso	Gênero	Tempo (h)	Disciplinas
A	F	2,0	5
B	M	3,5	6
A	M	1,5	4
B	F	2,5	5
A	F	2,0	5
B	M	4,0	7
A	M	3,0	6
B	F	2,0	4

a) Suponha que os alunos foram selecionados separadamente dentro de cada curso, por sorteio, e depois os dados foram reunidos. Qual o tipo de amostragem? É probabilística? Justifique.

b) Monte uma tabela de frequência (absoluta e relativa) para a variável Curso.

c) Considere a variável tempo de estudo. Calcule sua média e mediana. Existe algum valor extremo? Justifique.

d) Monte uma tabela de dupla entrada entre Curso e Gênero. O que você observa sobre a distribuição?

e) Obtenha uma medida de associação entre Curso e Gênero. O que você conclui?

2. O número diário de solicitações em um serviço de atendimento online foi registrado por um período de 200 dias, e os resultados foram resumidos na tabela a seguir.

Concentração	Número de amostras
[0, 200)	50
[200, 400)	65
[400, 800)	70
[800, 1200)	10
[1200, 2000]	5
Total	200

Com base nos dados apresentados, responda:

a) Construa um histograma para representar esses dados. Atenção para a largura das classes.

b) Calcule o número médio de solicitações.

c) Determine o número mediano de solicitações.

d) Determine o terceiro quartil do número de solicitações.

e) Calcule o coeficiente de variação do número de solicitações.

Respostas

---

1. (a) **Amostragem probabilística do tipo estratificada**. Justificativa:

   i. A coleta de dados configura-se como sendo uma amostragem pois seleciona-se um **subconjunto da população**;

   ii. a amostragem é probabilística pois a escolha de cada elemento a compor a amostra é feita por **sorteio** garantindo, assim, **aleatoriedade**;

   iii. a amostragem probabilística é do tipo estratificada pois:

   - a população de interesse possui **estratificações naturais** (cursos universitários A e B);
   - **amostra de cada um desses estratos** são obtidas;
   - **uma única amostra final** é composta com as amostras dos grupos.

   (b)

Curso	f_a	f_r	Total
A	4	0.5	4
B	4	0.5	4
Total	8	1.0	8

Tabela 1: Frequência absoluta e relativa por curso

   (c) Seja T a variável 'tempo de estudo (h)' e t_i as observações de T para todo i = 1, ..., 8. Então, t = {2.0, 3.5, 1.5, 2.5, 2.0, 4.0, 3.0, 2.0}

   i. Calculando a média:

   - Calculando a média:

$$\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i = \frac{1}{8}\sum_{i=1}^{8} t_i =$$

$$= \frac{1}{8}(2.0 + 3.5 + 1.5 + 2.5 + 2.0 + 4.0 + 3.0 + 2.0) =$$

$$= \frac{20.5}{8} = \mathbf{2.5625}$$

   ii. Calculando a mediana:

- Para calcular a mediana, precisamos primeiro ordenar os dados:

$$t_{(i)} = \{1.5, 2.0, 2.0, 2.0, 2.5, 3.0, 3.5, 4.0\}$$

- Como $n = 8$ (par), a mediana é dada por:

$$m_d(t) = \frac{t_{(\frac{n}{2})} + t_{(\frac{n}{2}+1)}}{2} = \frac{t_{(\frac{8}{2})} + t_{(\frac{8}{2}+1)}}{2} = \frac{t_{(4)} + t_{(5)}}{2}$$

- Substituindo os valores:

$$m_d(t) = \frac{2.0 + 2.5}{2} = \frac{4.5}{2} = \mathbf{2.25}$$

iii. Encontrando os valores extremos (outliers):

- Para encontrarmos os outliers, precisamos calcular a amplitude do gráfico boxplot e avaliar se existem valores de t fora desse intervalo...

$$t \notin [Q_1 - 1.5 \cdot AIQ,\ Q_3 + 1.5 \cdot AIQ]$$

- Amplitude interquartílica:

$$AIQ = Q_3 - Q_1$$

- Como $n = 8$ (par), Q1 é dada por:

$$Q_1(t) = \frac{t_{(\frac{n}{4})} + t_{(\frac{n}{4}+1)}}{2} = \frac{t_{(\frac{8}{4})} + t_{(\frac{8}{4}+1)}}{2} = \frac{t_{(2)} + t_{(3)}}{2} =$$

$$= \frac{2.0 + 2.0}{2} = \frac{4.0}{2} = 2.0$$

- Como $n = 8$ (par), Q3 é dada por:

$$Q_3(t) = \frac{t_{(\frac{3*n}{4})} + t_{(\frac{3*n}{4}+1)}}{2} = \frac{t_{(\frac{3*8}{4})} + t_{(\frac{3*8}{4}+1)}}{2} = \frac{t_{(\frac{24}{4})} + t_{(\frac{24}{4}+1)}}{2} = \frac{t_{(6)} + t_{(7)}}{2} =$$

$$= \frac{3.0 + 3.5}{2} = \frac{6.5}{2} = 3.25$$

- AIQ:

$$AIQ = Q_3 - Q_1 = 3.25 - 2.0 = 1.25$$

- Amplitude:

$$[2.0 - 1.5 \cdot 1.25,\ 3.25 + 1.5 \cdot 1.25] =$$

$$[2.0 - 1.875,\ 3.25 + 1.875] =$$

$$\mathbf{[0.125,\ \ 5.125]}$$

Como não há valores de t_i fora do intervalo $[0.125,\ \ 5.125]$, concluimos não haver pontos extremos (outliers).

(d)

	Masculino	Feminino	Total
Curso A	2	2	4
Curso B	2	2	4
Total	4	4	8

Tabela 2: Tabela de dupla entrada: Curso x Gênero

Observa-se claramente que as classes são uniformemente frequentes.

(e) Vamos avaliar o grau de associação entre as variáveis Curso e Sexo. Para isso, podemos utilizar a medida Qui-Quadrado:

$$Q = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}},$$

Os o_{ij}'s são as frequências absolutas observadas para cada uma das combinações de valores das variáveis Curso e Gênero, que já calculamos no item anterior com a tabela de dupla entrada. Já os e_{ij}'s são os valores esperados para cada mesma combinação, e são obtidos da seguinte forma:

$$e_{ij} = \frac{(t_{i\cdot})(t_{\cdot j})}{n}$$

Trocando pelos dados do exercício, temos:
Agora, calculando de fato a medida Qui-Quadrado, temos que:

$$Q = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

	Masculino	Feminino	Total
Curso A	$o_{11} = 2$	$o_{12} = 2$	4
Curso B	$o_{21} = 2$	$o_{22} = 2$	4
Total	4	4	8

Tabela 3: Frequências absolutas observadas

	Masculino	Feminino	Total
Curso A	$e_{11} = \frac{t_{1.}\, t_{.1}}{n} = \frac{4 \cdot 4}{8} = 2$	$e_{12} = \frac{t_{1.}\, t_{.2}}{n} = \frac{4 \cdot 4}{8} = 2$	4
Curso B	$e_{21} = \frac{t_{2.}\, t_{.1}}{n} = \frac{4 \cdot 4}{8} = 2$	$e_{22} = \frac{t_{2.}\, t_{.2}}{n} = \frac{4 \cdot 4}{8} = 2$	4
Total	4	4	8

Tabela 4: Frequências absolutas esperadas

	Masculino	Feminino
Curso A	$\frac{(o_{11}-e_{11})^2}{E_{11}} = \frac{(2-2)^2}{2} = 0$	$\frac{(o_{12}-e_{12})^2}{E_{12}} = \frac{(2-2)^2}{2} = 0$
Curso B	$\frac{(o_{21}-e_{21})^2}{E_{21}} = \frac{(2-2)^2}{2} = 0$	$\frac{(o_{22}-e_{22})^2}{E_{22}} = \frac{(2-2)^2}{2} = 0$

Tabela 5: Contribuições individuais para medida Qui-Quadrado

$$Q = \frac{(o_{11} - e_{11})^2}{e_{11}} + \frac{(o_{12} - e_{12})^2}{e_{12}} + \frac{(o_{21} - e_{21})^2}{e_{21}} + \frac{(o_{22} - e_{22})^2}{e_{22}}$$

$$Q = \frac{(2-2)^2}{2} + \frac{(2-2)^2}{2} + \frac{(2-2)^2}{2} + \frac{(2-2)^2}{2}$$

$$Q = \frac{0^2}{2} + \frac{0^2}{2} + \frac{0^2}{2} + \frac{0^2}{2}$$

$$Q = 0 + 0 + 0 + 0$$

$$\mathbf{Q = 0}$$

Quanto mais próxima a medida Qui-Quadrado de 0, menor a associação/dependência entre as variáveis, pois os valores observados são muito próximos dos valores esperados. No contexto do exercício, as variáveis Curso e Gênero tem absolutamente nenhuma associação ou dependência uma da outra, pois suas frequencias quando condicionadas uma a outra são iguais às frequências marginais.

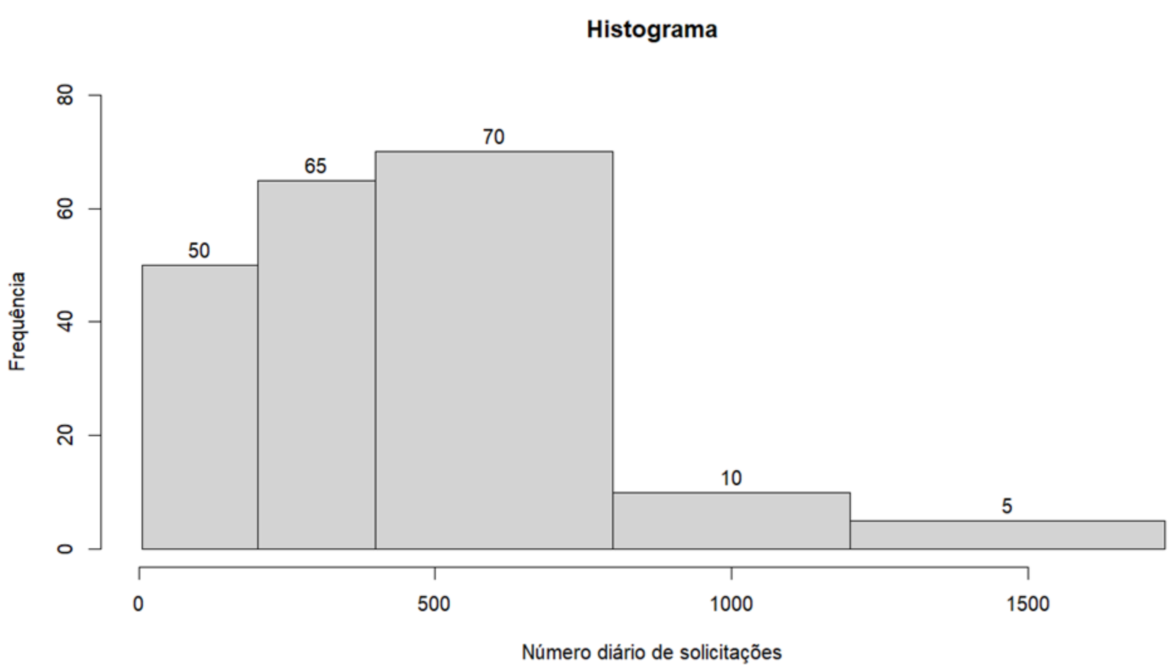
(Obs.: no Livro Estatística Básica de Bussab, na seção 4.4 Medidas de Associação entre Variáveis Qualitativas do capítulo 4 Análise Bidimensional, o autor apresenta uma boa intuição por trás da medida de associação Qui-Quadrado!)

2.

(a)

```r
# Simula amostra
x <- sample(1:199, 50, replace = TRUE)
x <- c(x, sample(200:399, 65, replace = TRUE))
x <- c(x, sample(400:799, 70, replace = TRUE))
x <- c(x, sample(800:1199, 10, replace = TRUE))
x <- c(x, sample(1200:2000, 5, replace = TRUE))

# Plota histograma
hist(
    x
    ,breaks = c(min(x), 200, 400, 800, 1200, max(x))
    ,freq = T
    ,labels = T
    ,ylim = c(0, 80)
    ,main = "Histograma"
    ,xlab = "Número diário de solicitações"
    ,ylab = "Frequência"
)
```



(b) Para calcular o número médio de solicitações, devemos realizar uma média ponderada. Para isso, consideramos o ponto médio e a frequência de cada intervalo:

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i \cdot f_i}{\sum_{i=1}^{k} f_i}$$

$$m_i = \frac{L_i + U_i}{2}$$

Usando os dados do exercício, temos:

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i \cdot f_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} \frac{L_i+U_i}{2} \cdot f_i}{\sum_{i=1}^{k} f_i}$$

$$= \frac{\frac{0+200}{2} \cdot 50 + \frac{200+400}{2} \cdot 65 + \frac{400+800}{2} \cdot 70 + \frac{800+1200}{2} \cdot 10 + \frac{1200+2000}{2} \cdot 5}{50 + 65 + 70 + 10 + 5}$$

$$= \frac{\frac{200}{2} \cdot 50 + \frac{600}{2} \cdot 65 + \frac{1200}{2} \cdot 70 + \frac{2000}{2} \cdot 10 + \frac{3200}{2} \cdot 5}{50 + 65 + 70 + 10 + 5}$$

$$= \frac{100 \cdot 50 + 300 \cdot 65 + 600 \cdot 70 + 1000 \cdot 10 + 1600 \cdot 5}{50 + 65 + 70 + 10 + 5}$$

$$= \frac{5000 + 19500 + 42000 + 10000 + 8000}{50 + 65 + 70 + 10 + 5}$$

$$= \frac{81500}{200} = \mathbf{422{,}5}$$

(c)

Para calcularmos o número mediano de solicitações, vamos fazer uma tabela de frequência absoluta, relatIva e acumulada, pois assim conseguiremos identificar qual intervalo de valores contém a mediana:

Concentração	f_a	f_r	f_{rcum}
[0, 200)	50	0.250	0.250
[200, 400)	65	0.325	0.575
[400, 800)	70	0.350	0.925
[800, 1200)	10	0.050	0.975
[1200, 2000]	5	0.025	1.000
Total	200	1.000	1.000

Tabela 6: Frequências absoluta, relativa e relativa acumulada.

A partir da tabela acima, podemos pereceber que a mediana está no segundo intervalo de valores [200, 400), visto que a frequência acumulada no primeiro intervalo anterior é menor que 50% e no segundo é maior que 50%. Para encontrar a mediana, usamos de uma regra de três:

$$\frac{400 - 200}{0.325} = \frac{m_d - 200}{0.500 - 0.250}$$

$$= = \frac{200}{0.325} = \frac{m_d - 200}{0.250} =>$$

$$=> m_d = \frac{200 \cdot 0.250}{0.325} + 200 = \frac{50}{0.325} + 200 <=>$$

$$<=> m_d = 153.8462 + 200 = \mathbf{353.8462}$$

(d)

Para calcularmos o terceiro quartil do número de solicitações, precisamos saber onde se concentram os 75% menores valores (ou os 25% maiores). Pela tabela 6 de frequencias acumuladas, conseguimos identificar que o Q3 está no intervalo [400, 800), visto que a frequência acumulada do intervalo [200, 400) é menor que 75% e a do intervalo [400, 800) é maior. Procedemos como no item anterior:

$$\frac{800 - 400}{0.350} = \frac{Q_3 - 400}{0.750 - 0.575}$$

$$= \frac{400}{0.350} = \frac{Q_3 - 400}{0.175} =>$$

$$=> Q_3 = \frac{400 \cdot 0.175}{0.350} + 400 = \frac{70}{0.350} + 400 <=>$$

$$<=> Q_3 = 200 + 400 = \mathbf{600}$$

(e) Coeficiente de variação

$$cv = \frac{s}{\bar{x}} <=> cv = \frac{\sqrt{\sum_{i=1}^{k} \frac{n_i \cdot (x_i - \bar{x})^2}{k}}}{\bar{x}} <=> cv = \frac{\sqrt{\sum_{i=1}^{k} f_i \cdot (x_i - \bar{x})^2}}{\bar{x}} <=>$$

$$cv = \frac{\sqrt{\sum_{i=1}^{k} 0.25 \cdot (100 - 422.5)^2 + ... + 0.025 \cdot (1600 - 422.5)^2}}{422.5} <=>$$

$$cv = \frac{\sqrt{26001.56 + 4877.03 + 11027.19 + 16675.31 + 34662.66}}{422,5} <=>$$

$$cv = \frac{305.36}{422.5} = \mathbf{0.723\ (72.3\%)}$$