

Introdução à Estatística

Prof. Lineu Alberto Cavazani de Freitas

Considere o conjunto de dados que representa uma amostra de alunos de primeiro ano de uma turma de graduação. As variáveis coletadas foram: grau de escolaridade (ensino médio, ensino superior, mestrado, doutorado), origem, número de irmãos e tempo que o indivíduo leva para chegar à universidade em minutos. Os dados da amostra estão descritos na tabela.

Escolaridade	Origem	Irmãos	Tempo até universidade
Ensino superior	Curitiba	2	30
Ensino médio	Região metropolitana	3	110
Ensino médio	Curitiba	8	150
Mestrado	Curitiba	3	25
Ensino superior	Curitiba	4	45
Ensino médio	Curitiba	0	30
Ensino superior	Outro Estado	2	60
Ensino superior	Interior do Paraná	2	60
Ensino médio	Outro Estado	1	10
Mestrado	Outro Estado	0	30

Com base nos dados, responda as questões.

Nas respostas use pelo menos 2 casas decimais.

Nos gráficos atente-se para a legenda e escalas.

-
- 1) Quais são os tipos das variáveis coletadas? Classifique-as em qualitativa nominal, qualitativa ordinal, quantitativa discreta ou quantitativa contínua.

Escolaridade	Qualitativa ordinal
Origem	Qualitativa nominal
Irmãos	Quantitativa discreta
Tempo	Quantitativa contínua

- 2) Considere que para gerar a amostra existia um cadastro de alunos. Decidiu-se por selecionar 1 a cada 50 alunos deste cadastro. Para isso, sorteou-se um número aleatório entre 1 e 50 para decidir a unidade de partida, para as demais unidades incrementou-se o número da unidade de partida de 50 em 50 até que o tamanho da amostra desejada fosse atingido. Qual o nome deste método de amostragem? Este plano de amostragem corresponde a um método probabilístico ou não probabilístico?

Amostragem.sistemática	Método probabilístico
------------------------	-----------------------

- 3) Monte uma tabela de frequências para a variável origem. Use frequências absolutas e relativas. Qual é a classe modal?

Origem	Frequência	Frequência relativa
Curitiba	5	0.5
Interior do Paraná	1	0.1
Outro Estado	3	0.3
Região metropolitana	1	0.1

R Curitiba é a classe modal, ou seja, a maior parte dos indivíduos vieram de Curitiba.

- 4) Monte uma tabela de frequências para a variável tempo até a universidade. Use faixas de tamanho 20, partindo de 0 até 160. Qual é a faixa modal?

Tempo	Frequência	Frequência relativa
[0,20)	1	0.1
[20,40)	4	0.4
[40,60)	1	0.1
[60,80)	2	0.2
[80,100)	0	0.0
[100,120)	1	0.1
[120,140)	0	0.0
[140,160]	1	0.1

R [20, 40) é a faixa modal.

- 5) Obtenha média, mediana, desvio padrão para as variáveis número de irmãos e tempo até a universidade.

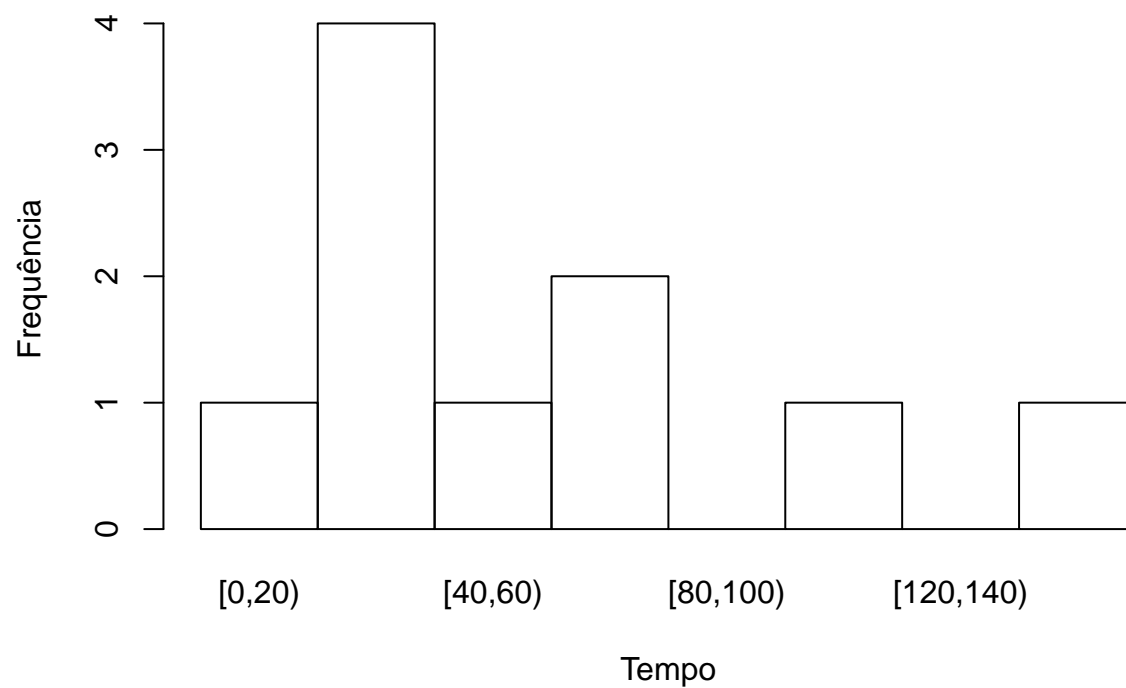
	Média	Variância	Desvio padrão
Irmãos	2.5	5.39	2.32
Tempo	55.0	1888.89	43.46

- 6) Considere o indivíduo da terceira linha da tabela. Qual o valor do Z-escore para número de irmãos e tempo até a universidade deste indivíduo? Interprete o resultado.

	Z
Irmãos	2.37
Tempo	2.19

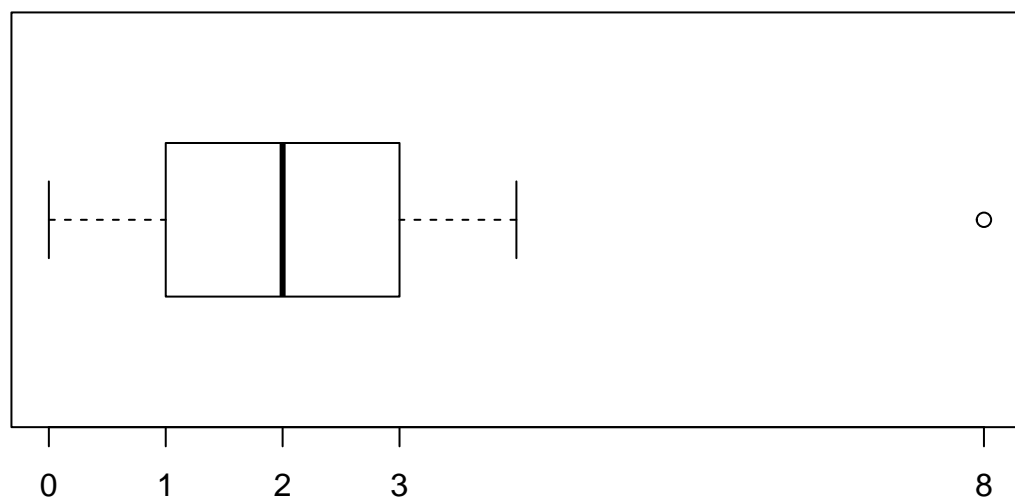
R O indivíduo está mais de 2 desvios padrões distante da média para ambas as variáveis.

- 7) Número de irmãos e tempo até a universidade são variáveis em diferentes escalas, qual delas apresenta maior variabilidade? Utiliza uma medida de comparação adequada.
- 8) Com base na tabela do item (4), esboce o histograma da variável pesos. O que você conclui a respeito da simetria?



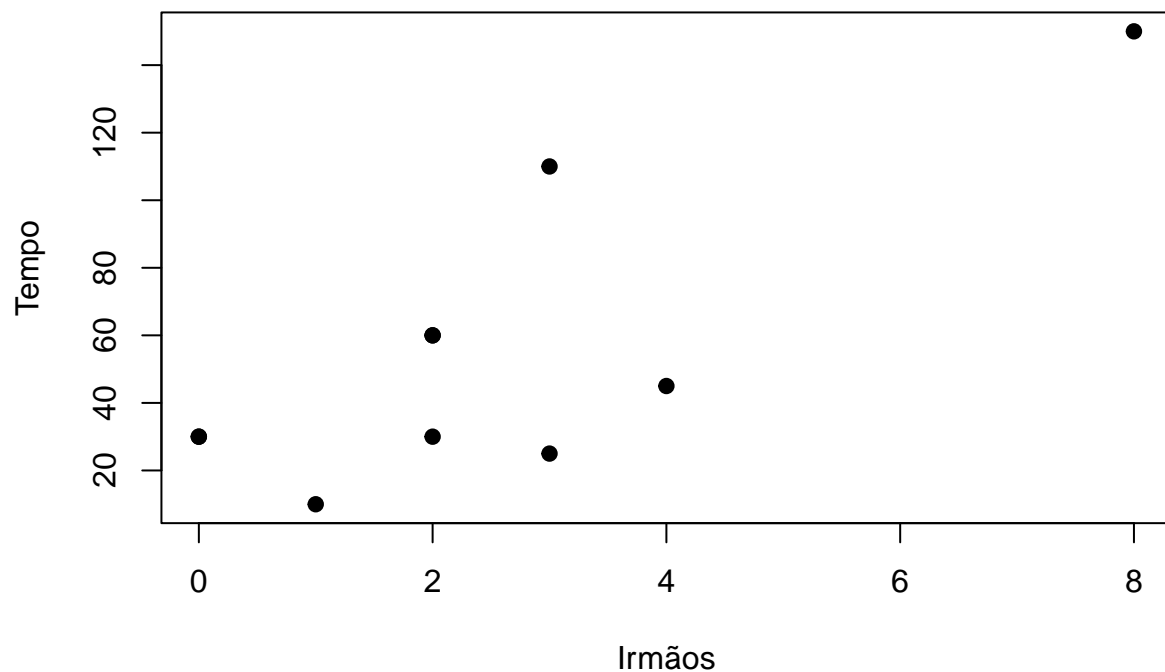
- 8) Obtenha as quantidades necessárias e esboce o box-plot da variável número de irmãos. Coloque nos eixos os valores utilizados para o esboço. O que você conclui a respeito da simetria e da presença de valores atípicos?

Mínimo	$Q1 - 1.5AIQ$	Q1	Q2	Q3	$Q3 + 1.5AIQ$	Máximo
0	-2	1	2	3	6	8

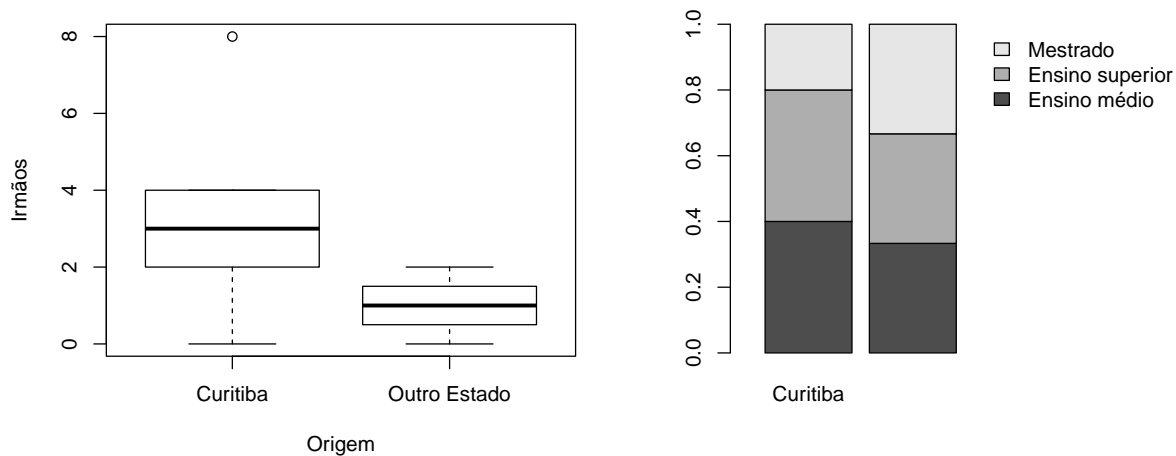


R A variável é razoavelmente simétrica mas apresenta um valor atípico.

- 9) Esboce a representação gráfica adequada que permita avaliar a relação entre as variáveis número de irmãos e tempo até a universidade. O que você conclui?



10) Avalie os gráficos abaixo. O que você conclui?



-
- R1 Aparentemente indivíduos de Curitiba tem mais irmãos que indivíduos de outros estados.
 R2 A proporção de indivíduos de outro estado com mestrado é superior à proporção de indivíduos com mestrado de Curitiba.
-

12) Obtenha uma medida de associação entre número de irmãos e tempo. O que você conclui?

$$\text{Correlação} = 0.79$$

R O valor calculado é maior que 0, indicando relação linear crescente entre as variáveis, assim como o diagrama de dispersão da questão 9.

- 13) Obtenha uma tabela de dupla entrada e uma medida de associação entre escolaridade e origem. O que você conclui?

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

$$\bar{y} = \frac{\sum_{i=1}^k f_i \cdot y_i}{\sum_{i=1}^k f_i}$$

$$A = \max(y) - \min(y)$$

$$DAM_{média} = \frac{1}{n} \sum_{i=1}^n |(y_i - \bar{y})|$$

$$DAM_{mediana} = \frac{1}{n} \sum_{i=1}^n |(y_i - md)|$$

$$s^2 = \text{Var}(y) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} \right)$$

$$s = \sqrt{s^2}$$

$$CV = 100 \cdot \frac{s}{\bar{y}}$$

$$z = \frac{y_i - \bar{y}}{s}$$

$$H = - \sum_{i=1}^S f_r \times \ln(f_r)$$

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

$$\text{Cov}(y_1, y_2) = \frac{1}{n-1} \sum_{i=1}^n (y_{1i} - \bar{y}_1) \cdot (y_{2i} - \bar{y}_2)$$

$$r = \frac{\sum_{i=1}^n (y_{1i} - \bar{y}_1) \cdot (y_{2i} - \bar{y}_2)}{\sqrt{\sum_{i=1}^n (y_{1i} - \bar{y}_1)^2} \cdot \sqrt{\sum_{i=1}^n (y_{2i} - \bar{y}_2)^2}} = \frac{\text{Cov}(y_1, y_2)}{\sqrt{V(y_1) \cdot V(y_2)}}$$
