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.

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?

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

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?

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

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.

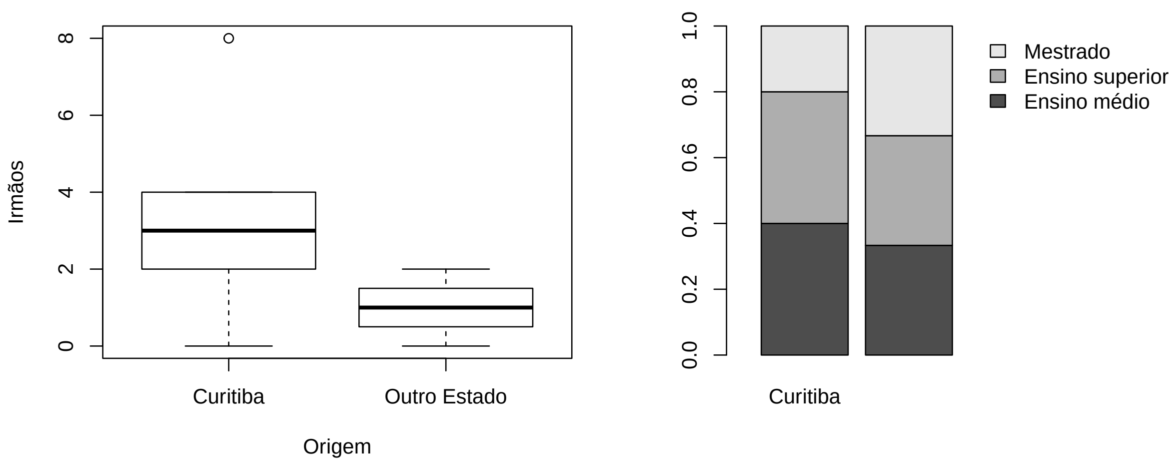
7) Número de irmãos e tempo até a universidade são variáveis em diferentes escalas, qual delas apresenta maior variabilidade? Utiliza uma medida de comparação adequada.

8) Com base na tabela do item (4), esboce o histograma da variável pesos. O que você conclui a respeito da simetria?

9) 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?

10) 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?

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



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

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

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad \overline{y} = \frac{\sum_{i=1}^{k} f_i \cdot y_i}{\sum_{i=1}^{k} f_i} \qquad A = max(y) - min(y)$$

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

$$s^2 = \mathrm{Var}(y) = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}\right) \qquad s = \sqrt{s^2}$$

$$\mathrm{CV} = 100 \cdot \frac{s}{\overline{y}} \qquad z = \frac{y_i - \bar{y}}{s} \qquad H = -\sum_{i=1}^{S} f_r \times ln(f_r) \qquad Q = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

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