

Prof Dr. Paulo Canas Rodriques

Respostas da primeira lista de exercício

Questão 01) Um inquérito a 100 pessoas em Salvador sobre as suas idades resultou no quadro seguinte:

| Escalão Etário | Frequência absoluta |
|----------------|---------------------|
| [0, 20] | 20 |
|]20, 40] | 25 |
|]40, 60] | 40 |
|]60, 80] | 10 |
|]80, 100] | 5 |

a) Obtenha a tabela de frequências completa, incluindo a frequência absoluta acumulada, relativa e relativa acumulada.

A partir dos dados informados na tabela, obtém-se os seguintes resultados:

| Escalão Etário | Freq. Abs | Freq. Rel (%) | Freq. Abs. Acum | Freq. Rel. Acum (%) |
|----------------|-----------|---------------|-----------------|---------------------|
| [0, 20] | 20 | 20 | 20 | 20 |
|]20, 40] | 25 | 25 | 45 | 45 |
|]40, 60] | 40 | 40 | 85 | 85 |
|]60, 80] | 10 | 10 | 95 | 95 |
|]80, 100] | 5 | 5 | 100 | 100 |

b) Calcule, para a idade:

i A média:

A média pode ser dada pelo valor médio nos intervalos de classe multiplicados pela frequência relativa.

$$\bar{x} = \sum_{i=1}^n x_i \cdot \frac{f_i}{100}$$

Onde x_i é a idade média em cada intervalo, f_i é a frequência relativa, e \bar{x} é a média.

$$\bar{x} = 10 \times 0,2 + 30 \times 0,25 + 50 \times 0,4 + 70 \times 0,1 + 90 \times 0,05 = 41 \text{ anos} \quad (1)$$

ii A mediana:

A mediana pode ser definida somando os dois números do meio e dividindo por dois em valores ordenados. Quando a quantidade de números N que estamos avaliando for par ou escolhendo o número do meio caso N seja ímpar.

Para o caso específico de intervalos de classe, com dados agrupados, podemos fazer uma regra de três com a diferença entre metade do tamanho da amostra $\frac{n}{2} = 50$ e o valor da frequência absoluta acumulada f_{ac} no intervalo anterior $]20, 40] = 45$, juntamente ao valor da frequência absoluta $f_{ab} = 40$ e a amplitude do intervalo $]40, 60] \rightarrow h = 60 - 40 = 20$. Após isso, soma-se o encontrado com o limite inferior da sua classe pertencente ($L_{min} = 40$). Matematicamente, pode-se representar como:

$$\text{mediana} = L_{min} + \frac{\left(\frac{n}{2} - f_{ac}\right)}{f_{ab}} \cdot h$$

$$\text{mediana} = 40 + \frac{(50 - 45) \cdot 20}{40} = 42,5 \quad (2)$$

iii A moda:

A moda pode ser definida como a média dos valores mais frequentes. Com essa definição, observa-se inicialmente que o intervalo de classe $[40, 60]$ é o mais frequente. Pela fórmula de Czuber, podemos obter a moda como:

$$\text{moda} = L_{min} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot h$$

Onde Δ_1 é a diferença entre a frequência da classe modal e a classe anterior e Δ_2 é a diferença entre a frequência da classe modal e a classe seguinte.

$$\Delta_1 = 40 - 25 = 15;$$

$$\Delta_2 = 40 - 10 = 30;$$

$$h = 60 - 40 = 20$$

$$\text{moda} = 40 + \frac{15}{15 + 30} \cdot 20 \approx 42,67 \quad (3)$$

c) Obtenha o boxplot. Interprete o boxplot.

Para montar o boxplot, é necessário primeiro obter os dados dos primeiro e terceiro quartis, além de uma medida de dispersão que pode ser dada como a diferença entre o terceiro e primeiro quartis.

Para o cálculo dos quartis, as equações são similares à mediana (que é o segundo quartil):

$$q_1 = L_{min} + \frac{\left(\frac{n}{4} - f_{ac}\right)}{f_{ab}} \cdot h = 20 + \frac{(25 - 20) \cdot 20}{25} = 24$$

$$q_3 = L_{min} + \frac{\left(3 \cdot \frac{n}{4} - f_{ac}\right)}{f_{ab}} \cdot h = 40 + \frac{(75 - 45) \cdot 20}{40} = 55$$

Com os quartis calculados, obtém-se a dispersão e os limites do boxplot:

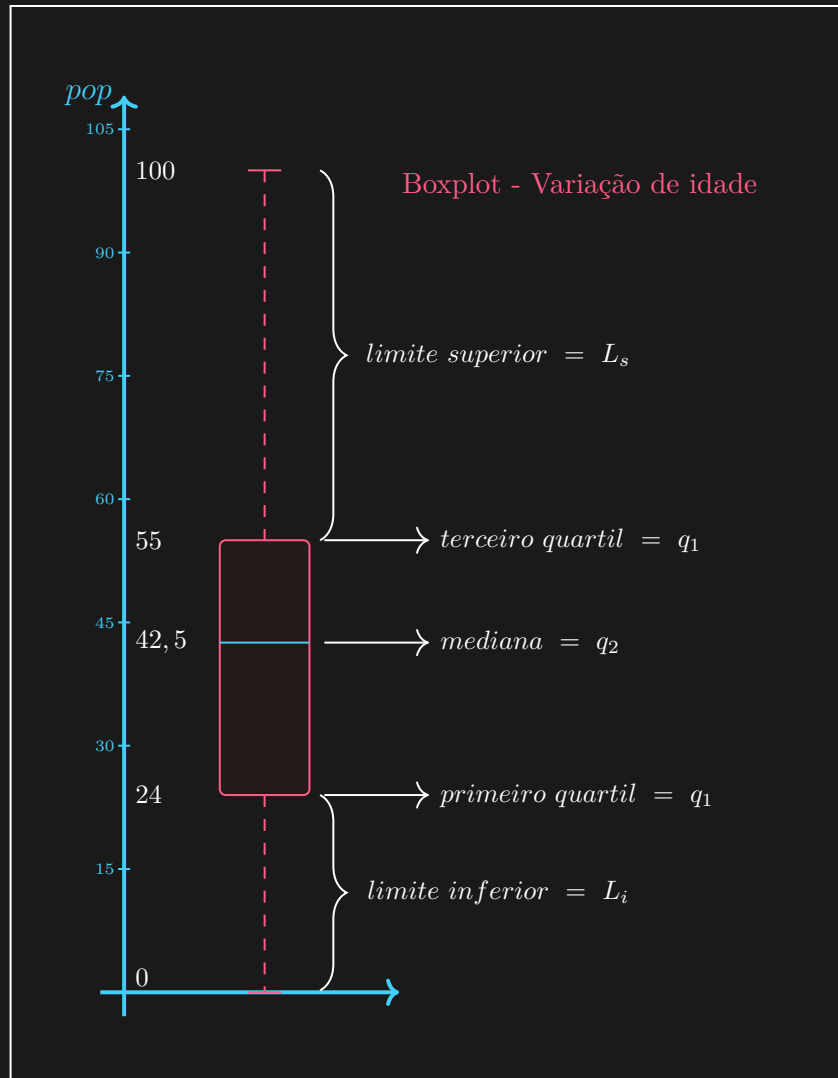
$$d_q = q_3 - q_1 = 55 - 24 = 31$$

$$L_s = q_3 + 1,5 \cdot d_q = 45 + 1,5 \cdot 31 = 101,5$$

$$L_i = q_1 - 1,5 \cdot d_q = 24 - 1,5 \cdot 31 = -22,5$$

Como o limite inferior de idade na amostra é 0, então $L_i = 0$. Assim como o limite superior é 100, então $L_s = 100$.

A partir desses valores, podemos montar o boxplot como:



No boxplot podemos observar uma assimetria que se manifesta na maior concentração de pessoas com idades entre 24 e 55 anos (primeiro e terceiro quartis), com maior prevalência para idades próximas à mediana. Nesse intervalo de dados, **não há nenhum valor** além do limite superior nem inferior. Quando isso ocorre, esses valores são tidos como outliers. Do ponto de vista estatístico eles podem ser considerados um erro de observação ou arredondamento. Contudo, não necessariamente se pode afirmar tal relação, pois a situação ainda é possível além do limite superior, mesmo sendo atípica.

d) Qual a percentagem de pessoas que têm idade inferior a 50 anos?

Tem-se a soma das frequências relativas dos dois primeiros intervalos de classe, juntamente com metade da frequência do terceiro. Isto é:

$$f_{rel} = f_{rel}^{[0,20]} + f_{rel}^{[20,40]} + \frac{f_{rel}^{[40,60]}}{2} = 20\% + 25\% + 20\% = 65\%$$

e) Quantas pessoas têm idades entre 40 e 80 anos?

Para essa questão, basta apenas somar os valores da frequência absoluta do terceiro e quarto intervalos:

$$f_{abs} = f_{abs}^{[40,60]} + f_{abs}^{[60,80]} = 40 + 10 = 50 \text{ pessoas}$$

Questão 02) Um inquérito a 100 pessoas em Salvador sobre as suas idades resultou no quadro seguinte: