LABORATÓRIO DE ESTATÍSTICA

Matemática Aplicada à Economia e à Gestão
1º Ano / 1º Semestre
2026/2027

CONTACTO

Professora: Elisabete Fernandes
E-mail: efernandes@iseg.ulisboa.pt

https://doity.com.br/estatistica-aplicada-a-nutricao

https://basiccode.com.br/produto/informatica-basica/

PROGRAMA

1. Conceitos Fundamentais de Estatística

2. Análise Exploratória de Dados

3. Redução e Sintetização de Dados

4. Associação e Relações entre Variáveis

5. Números Índices

6. Sucessões Cronológicas

AULA PRÁTICA 3: EXERCÍCIO 2

"The exercises planned for the class 3 are 2.1 and 2.4 " ✅

EXERCÍCIO 2.1: ANÁLISE EXPLORATÓRIA DE DADOS

2.1 Para efeitos de controle de qualidade um produto é vendido em caixas de 6 unidades, das quais se selecionam aleatoriamente e sem reposição 40 destas caixas, observando-se o número de produtos defeituosos incluídos em cada caixa. A amostra observada consta do quadro que se segue:

2, 0, 1, 3, 0, 3, 6, 4, 5, 3, 0, 1, 0, 3, 2, 0, 1, 0, 2, 3, 1, 0, 3, 0, 1, 0, 0, 1, 5, 6, 0, 2, 5, 1, 0, 3, 0, 2, 3, 0.

a) Defina a unidade estatística e a variável em análise, classificando-a.

b) Apresente o quadro de frequências.

c) Represente graficamente estes dados.

d) Qual o domínio de variação e o intervalo inter-quartis?

e) Como avalia a simetria ou assimetria desta distribuição?

EXERCÍCIO 2.1 A): SOLUÇÃO

✅ Resposta:

a) Statistical Unit and Variable

- **Statistical unit:** One box of six units.
- **Variable:** Number of defective units per box.
- **Classification:** Discrete quantitative variable (count data).

EXERCÍCIO 2.1 B): SOLUÇÃO

✅ Resposta:

b) Frequency Table

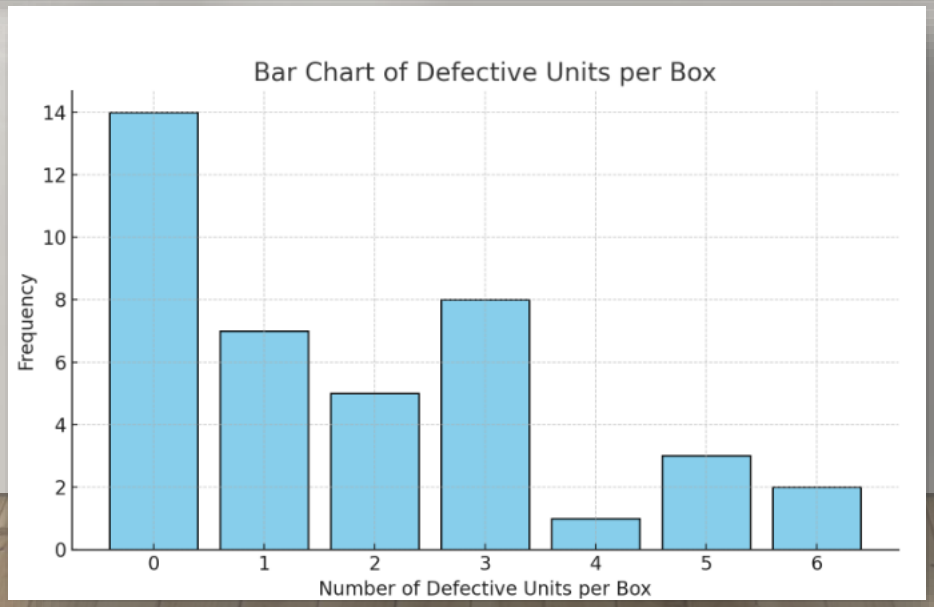
Defective Units (x)	Absolute Frequency (f)	Relative Frequency (%)
0	14	35.0%
1	7	17.5%
2	5	12.5%
3	8	20.0%
4	1	2.5%
5	3	7.5%
6	2	5.0%
Total	**40**	**100%**

EXERCÍCIO 2.1 C): SOLUÇÃO

✅ Resposta:

c) Graphical Representation

- **Bar chart** with x-axis = number of defectives (0–6) and y-axis = frequency.
- Modal value: 0 (most frequent).



EXERCÍCIO 2.1 D): SOLUÇÃO

Amplitude and Amplitude Inter-quartil (IQR)

✅ Answer: D = [0, 6] ; IQR = [0, 3]

Alternative Resolution

Exercise 2.1 (d) – Quartiles and IQR

Formula (Silvestre, 2015):

$$Q_p = (1-\alpha)\, x_{(r)} + \alpha\, x_{(r+1)}$$

with

$$\text{Position} = p(n+1)$$

$$r = \lfloor \text{Position} \rfloor \quad \text{(integer part of the position)}$$

$$\alpha = \text{Position} - r \quad \text{(decimal part of the position)}$$

Calculations (n = 40):

- Q1 (p = 0.25):

  $\text{Position} = 0.25 \times 41 = 10.25 \;\Rightarrow\; r = 10, \alpha = 0.25$

  $$Q1 = 0.75 \cdot x_{(10)} + 0.25 \cdot x_{(11)} = 0$$

- Q3 (p = 0.75):

  $\text{Position} = 0.75 \times 41 = 30.75 \;\Rightarrow\; r = 30, \alpha = 0.75$

  $$Q3 = 0.25 \cdot x_{(30)} + 0.75 \cdot x_{(31)} = 3$$

- IQR:

  $$IQR = Q3 - Q1 = 3 - 0 = 3$$

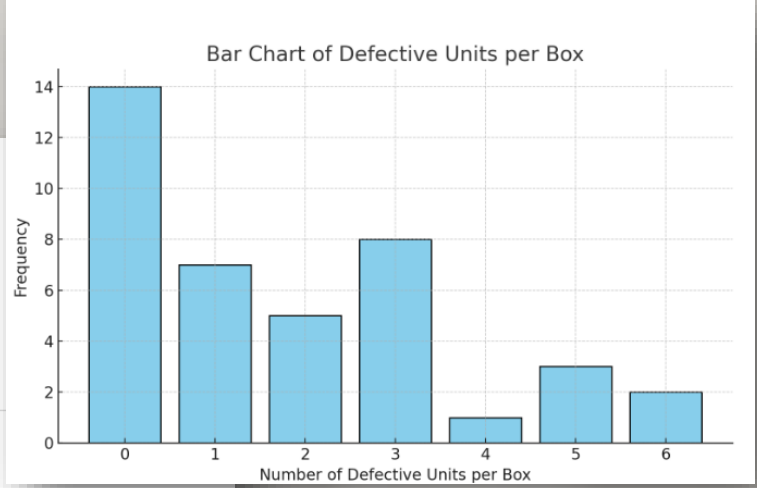
"The interquartile range (IQR) is **technically defined** as the difference between the third and first quartiles, that is, $IQR = Q3 - Q1$. This is the standard definition found in most statistics textbooks and software (Excel, R, Python).

However, some sources or people may informally refer to the 'interquartile interval' as the interval [Q1, Q3], describing the range that contains the middle 50% of the data. This is not the standard technical definition of the IQR, but rather a way of describing the central spread of the data."

EXERCÍCIO 2.1 D): SOLUÇÃO

✅ Resposta: Moda = 0

- Plot a **bar chart** of the data frequencies.
- Observation from the bar chart:
  - Many low values (0, 1) → high bars on the left
  - Fewer high values (4, 5, 6) → smaller bars on the right
- The **tail extends to the right** → distribution is **positively skewed (right-skewed)**



Conclusion:

Based on the **shape of the bar chart**, the distribution shows **positive skewness**, with a concentration of low values and a tail toward higher values.

EXERCÍCIO 2.1 D): SOLUÇÃO

✅ Resposta:

Alternative Resolution

Step 1: Basic statistics

- Number of observations: $n = 40$
- Minimum = 0, Maximum = 6
- Mode = 0 (most frequent value)
- **Median**:

$$\text{Position of median} = \frac{n+1}{2} = \frac{40+1}{2} = 20.5$$

Median = average of 20th and 21st observations → Median = 1

- Mean:

$$\text{Mean} = \frac{\text{Sum of all observations}}{n} = 1{,}80$$

Step 2: Skewness assessment

- Mode = 0, Median = 1, Mean = 1,80
- Mean > Median > Mode → **positively skewed (right-skewed) distribution**
- Many low values (0, 1) and a few higher values (4, 5, 6) confirm the **right tail**

Conclusion:

The distribution is **right-skewed (positively skewed)**, indicating a concentration of low values and a tail extending toward higher values.

EXERCÍCIO 2.4: ANÁLISE EXPLORATÓRIA DE DADOS

2.4 Considere as notas de 50 alunos desta uc obtidas no ano letivo anterior:

93, 77, 67, 72, 52, 83, 66, 84, 59, 63,
75, 97, 84, 73, 81, 42, 61, 51, 91, 87,
34, 54, 71, 47, 79, 70, 65, 57, 90, 83,
58, 69, 82, 76, 71, 60, 38, 81, 74, 69,
68, 76, 85, 58, 45, 73, 75, 42, 93, 65.

a) Represente os dados num diagrama de caule e folhas ordenado.

b) Calcule os cinco números sumário a partir do diagrama anterior.

c) Construa uma caixa de bigodes e analise a simetria ou assimetria dos dados.

d) Agrupe os dados em 5 classes de amplitude constante, com o limite inferior da primeira classe igual a 25 e construa um histograma.

EXERCÍCIO 2.4 A): SOLUÇÃO

✅ Resposta:

a) Stem-and-Leaf Plot

Steps:

1. Separate the tens and units of each grade. Tens → stem, units → leaf.
2. Order the leaves within each stem.

Example of ordered stem-and-leaf plot:

		n_i
3	4 8	2
4	2 2 5 7	4
5	1 2 4 7 8 8 9	7
6	0 1 3 5 5 6 7 8 9 9	10
7	0 1 1 2 3 3 4 5 5 6 6 7 9	13
8	1 1 2 3 3 4 4 5 7	9
9	0 1 3 3 7	5

EXERCÍCIO 2.4 B): SOLUÇÃO

✅ Answer:

1️⃣ Ordered data with positions

$$
\begin{array}{lllll}
x(1)=34, & x(2)=38, & x(3)=42, & x(4)=42, & x(5)=45, \\
x(6)=47, & x(7)=51, & x(8)=52, & x(9)=54, & x(10)=57, \\
x(11)=58, & x(12)=58, & x(13)=59, & x(14)=60, & x(15)=61, \\
x(16)=63, & x(17)=65, & x(18)=65, & x(19)=66, & x(20)=67, \\
x(21)=68, & x(22)=69, & x(23)=69, & x(24)=70, & x(25)=71, \\
x(26)=71, & x(27)=72, & x(28)=73, & x(29)=73, & x(30)=74, \\
x(31)=75, & x(32)=75, & x(33)=76, & x(34)=76, & x(35)=77, \\
x(36)=79, & x(37)=81, & x(38)=81, & x(39)=82, & x(40)=83, \\
x(41)=83, & x(42)=84, & x(43)=84, & x(44)=85, & x(45)=87, \\
x(46)=90, & x(47)=91, & x(48)=93, & x(49)=93, & x(50)=97
\end{array}
$$

EXERCÍCIO 2.4 B): SOLUÇÃO

✅ Resposta:

2️⃣ **Silvestre's formula for quartiles**

$$\text{position} = \alpha \cdot (n + 1)$$

- n = 50 (number of observations)
- α = 0.25 for Q1, 0.5 for Q2 (median), 0.75 for Q3
- If the position is not an integer, interpolate:

$$Q = (1 - \text{frac}) \cdot x(r) + (\text{frac}) \cdot x(r + 1)$$

where r is the integer part of the position, and frac is the decimal part.

EXERCÍCIO 2.4 B): SOLUÇÃO

✅ Resposta:

3️⃣ Calculate Q1, Q2, Q3

Q1 (α = 0.25):

$$\text{position} = 0.25 \cdot (50 + 1) = 12.75$$

- r = 12 → x(12) = 58
- frac = 0.75 → x(13) = 59

$$Q1 = (1 - 0.75) \cdot x(12) + 0.75 \cdot x(13) = 0.25 \cdot 58 + 0.75 \cdot 59 = 58.75$$

Median (Q2, α = 0.5):

$$\text{position} = 0.5 \cdot 51 = 25.5$$

- r = 25 → x(25) = 71
- frac = 0.5 → x(26) = 71

$$Q2 = 0.5 \cdot x(25) + 0.5 \cdot x(26) = 71$$

EXERCÍCIO 2.4 B): SOLUÇÃO

✅ Resposta:

Q3 (α = 0.75):

$$\text{position} = 0.75 \cdot 51 = 38.25$$

- r = 38 → x(38) = 81
- frac = 0.25 → x(39) = 82

$$Q3 = 0.75 \cdot x(38) + 0.25 \cdot x(39) = 81.25$$

4 Minimum and Maximum

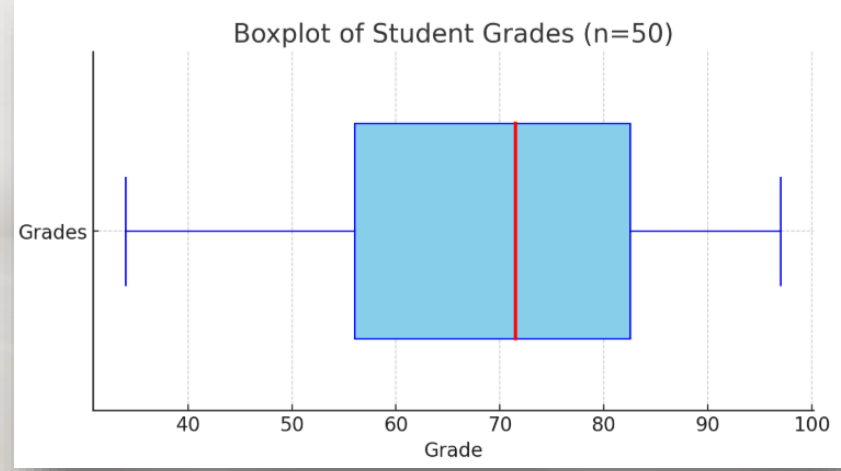
$$\text{Min} = x(1) = 34, \quad \text{Max} = x(50) = 97$$

5 Five-number summary

$$\boxed{(34, 58.75, 71, 81.25, 97)}$$

EXERCÍCIO 2.4 C): SOLUÇÃO

✅ Resposta:



c) Boxplot and Symmetry Assessment

Steps to construct boxplot:

- Draw a number line from minimum to maximum.
- Plot a box from Q1 to Q3 with a line at the median.
- Draw "whiskers" from min to Q1 and from Q3 to max.

Symmetry assessment:

- Since the median is slightly closer to Q1 than Q3 and the upper whisker is shorter than the lower whisker, the distribution is **slightly left-skewed**.

EXERCÍCIO 2.4 D): SOLUÇÃO

✅ Resposta:

1. Dados

- Mínimo real dos dados: **34**
- Máximo real dos dados: **97**
- Início da 1.ª classe: **25**
- Número de classes: **5**

2. Amplitude total

$$R' = \text{máximo desejado} - \text{mínimo desejado} = 97 - 25 = 72$$

3. Amplitude das classes

$$A = \frac{R'}{k} = \frac{72}{5} = 14,4$$

Arredondamos normalmente para **15** (para ficar em números inteiros e cobrir todo o interval

Width of each class = 15

Classes:

- $I_1 : [25, 40]$
- $I_2 :]40, 55]$
- $I_3 :]55, 70]$
- $I_4 :]70, 85]$
- $I_5 :]85, 100]$

Steps to create the histogram:

1. **Count the number of observations** in each class I_j.
2. **Calculate relative frequency**:

$$\text{Relative Frequency} = \frac{\text{Number of observations in class}}{n}$$

where n = total number of observations.

3. Draw the histogram:

- X-axis → classes I_j
- Y-axis → relative frequency
- Height of each bar = relative frequency of that class

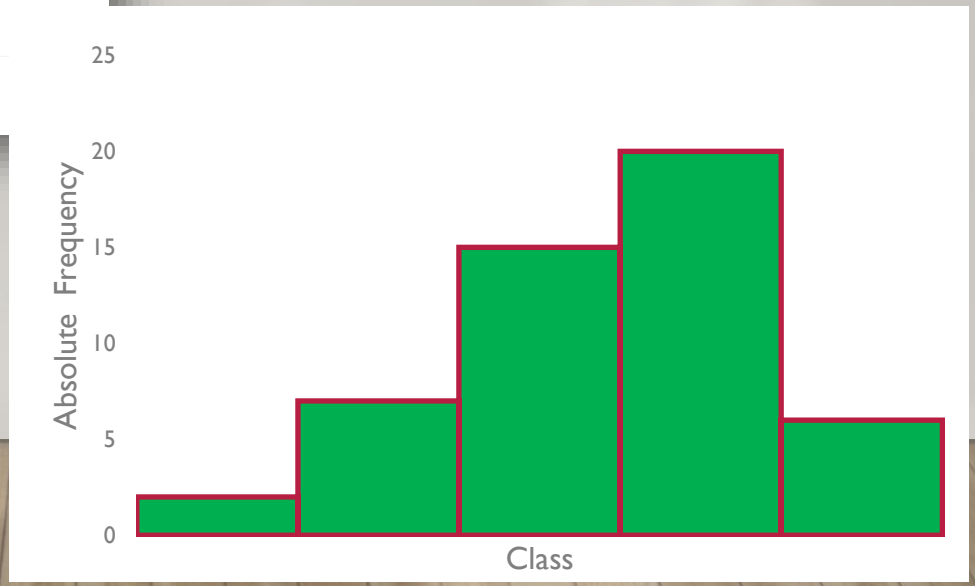
EXERCÍCIO 2.4 D): SOLUÇÃO

✅ Resposta:

Frequência relativa $f_r = f/n$

Class Interval	Midpoint x_j	Absolute Frequency f	Relative Frequency f_r
[25, 40]	32.5	2	0.04
(40, 55]	47.5	7	0.14
(55, 70]	62.5	15	0.30
(70, 85]	77.5	20	0,4
(85, 100]	92.5	6	0,12

Classe Modal =]70, 85]



EXERCÍCIO 2.4 D): SOLUÇÃO

✅ Resposta:

Frequency Table

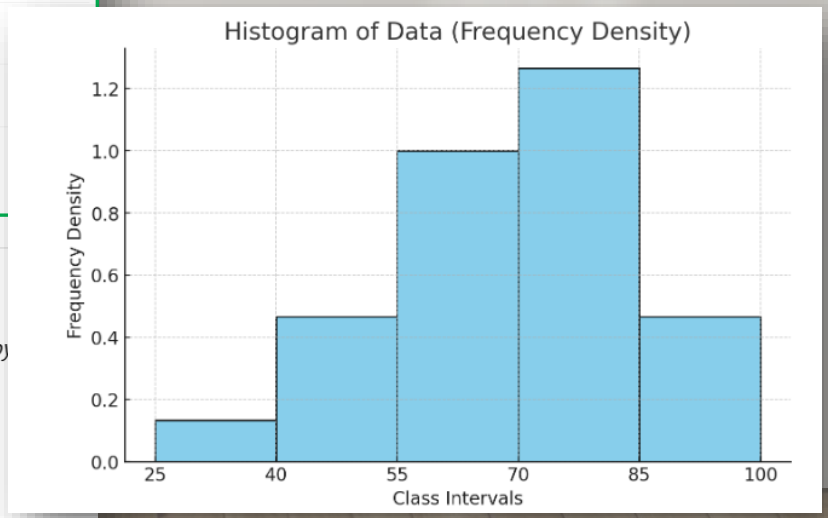
Class Interval	Midpoint x_j	Absolute Frequency f	Relative Frequency f_r	Frequency Density f/h
[25, 40]	32.5	2	0.04	0.1333
(40, 55]	47.5	7	0.14	0.4667
(55, 70]	62.5	15	0.30	1.0000
(70, 85]	77.5	20	0,4	1,333
(85, 100]	92.5	6	0,12	0,4

Note (in English):

The "Frequency Density" column represents the height of each class in a histogram. It is calculated by dividing the absolute frequency by the class width:

$$\text{Frequency Density} = \frac{f_j}{h_j}$$

This is especially useful when class widths are different, so the histogram accurately reflects the distribution of data. In this case, all class widths are equal (h = 15), so the density is proportional to the frequency.



EXERCISE 2.2: EXPLORATORY DATA ANAYSIS

2.2 Na seguinte tabela apresenta-se o tempo de estudo semanal fora das aulas (em horas) observado numa amostra de estudantes de MAEG:

6.7	8.2	8.5	5.7	9	8	18.2	9.5	5.6	11.2	9.8	3.1	7.6	8.8	6.3

a) Defina a unidade estatística e variável em estudo.

b) Calcule os cincos números sumários para este conjunto de dados.

c) Analise a simetria dos dados a partir de um diagrama caixa de bigodes

EXERCISE 2.2 A): SOLUTION

✅ Answer:

a) Statistical Unit and Variable

- **Statistical unit:** each MAEG student
- **Variable:** number of weekly study hours outside classes (quantitative continuous)

EXERCISE 2.2 B): SOLUTION

✅ Answer:

b) Five-Number Summary (Silvestre, 2013)

- **Ordered data:** 3.1, 5.6, 5.7, 6.3, 6.7, 7.6, 8, 8.2, 8.5, 8.8, 9, 9.5, 9.8, 11.2, 18.2
- **Quantile formula:**

$$\text{Position} = \alpha \cdot (n + 1)$$

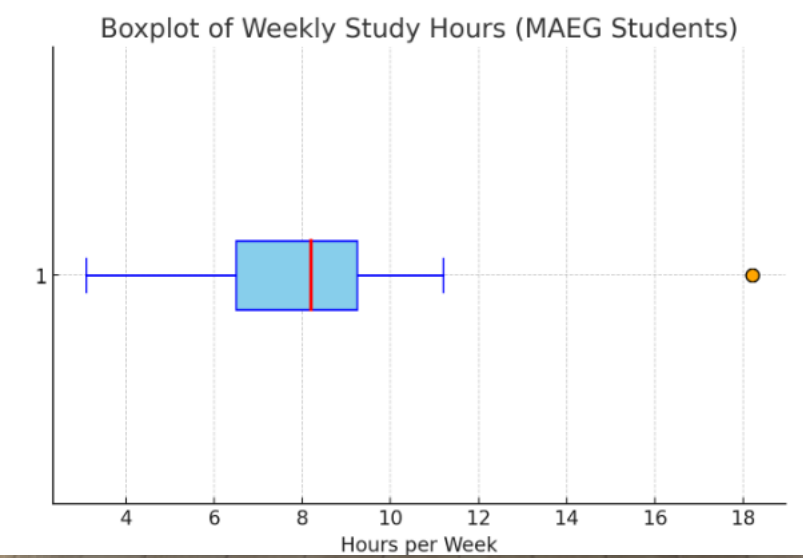
- **Q1 (α = 0.25):** Position = 0.25 × 16 = 4 → 4th value = 6.3
- **Median (Q2, α = 0.5):** Position = 0.5 × 16 = 8 → 8th value = 8.2
- **Q3 (α = 0.75):** Position = 0.75 × 16 = 12 → 12th value = 9.5
- **Five-number summary:**
  - Minimum = 3.1
  - Q1 = 6.3
  - Median = 8.2
  - Q3 = 9.5
  - Maximum = 18.2

EXERCISE 2.2 C): SOLUTION

✅ Answer:

c) Symmetry Analysis (Boxplot)

- **Box:** Q1 (6.3) to Q3 (9.5)
- **Median:** 8.2 (closer to Q1) → longer tail to the right
- **Maximum (18.2):** far from Q3 → potential outlier
- **Conclusion:** Data is **positively skewed** (right-skewed)



EXERCISE 2.3: EXPLORATORY DATA ANAYSIS

2.3 A informação estatística relativa aos salários mensais auferidos por 100 trabalhadores de uma empresa têxtil consta do quadro que se segue:

Salário (euros)	Trabalhadores (%)
[600 ; 900]	9
]900 ; 1000]	18
]1000 ; 1100]	35
]1100 ; 1200]	22
]1200 ; 1300]	16

a) Defina a variável em estudo e classifique-a.

b) Represente graficamente os dados.

c) Avalie e simetria dos dados a partir da representação gráfica realizada.

EXERCISE 2.3 A): SOLUTION

✅ Answer:

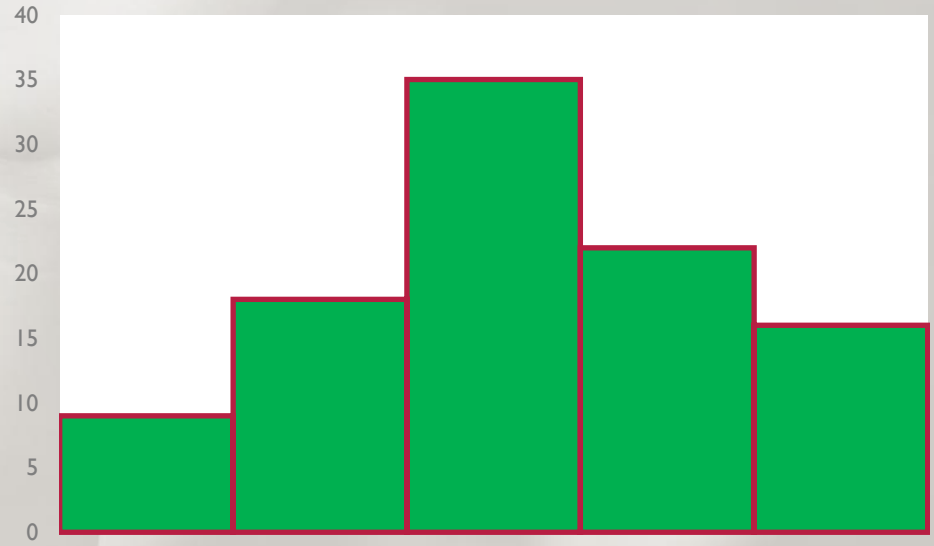
The variable "**monthly wages**" is an example of a **quantitative continuous variable**.

Explanation:

- **Quantitative** because it represents a **numerical measurement**.
- **Continuous** because, in theory, it can take **any value within a range** (e.g., $1200.50, $1200.75, $1200.756, etc.), not just integer values.

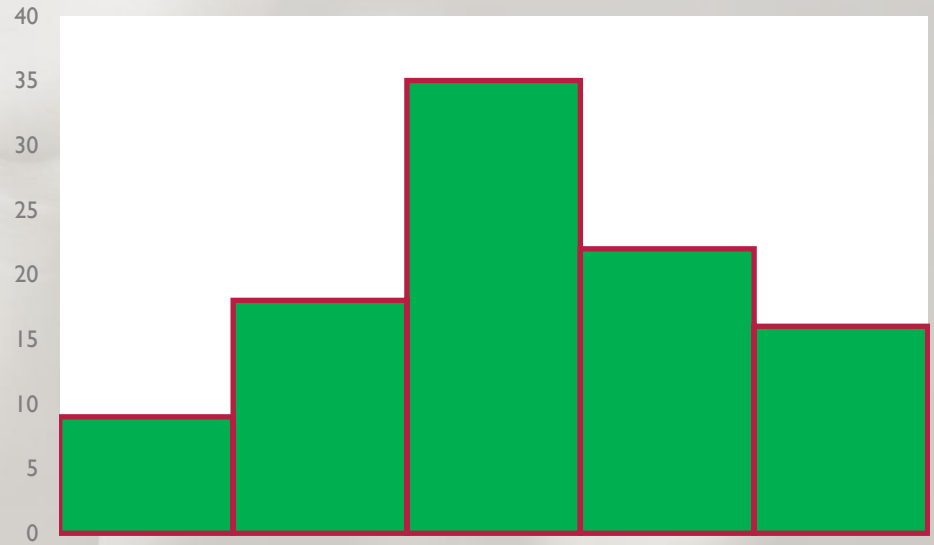
EXERCISE 2.3 B): SOLUTION

✅ Answer:



EXERCISE 2.3 C): SOLUTION

✅ Answer:



Negative Skewness

OBRIGADA!

Questões?