
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

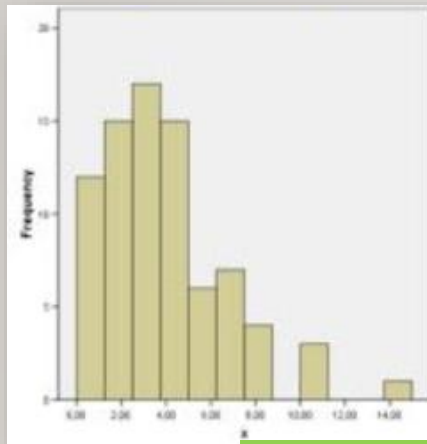


6. Sucessões
Cronológicas

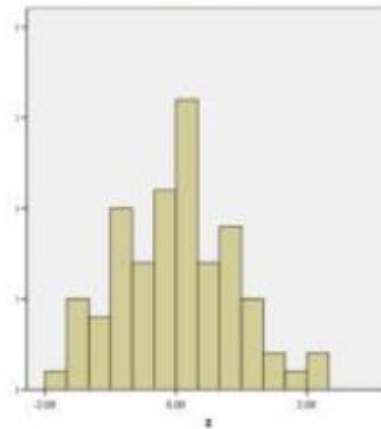
AULA 3: ANÁLISE EXPLORATÓRIA DE DADOS (CONTINUAÇÃO)

ANÁLISE DA SIMETRIA VIA HISTOGRAMA

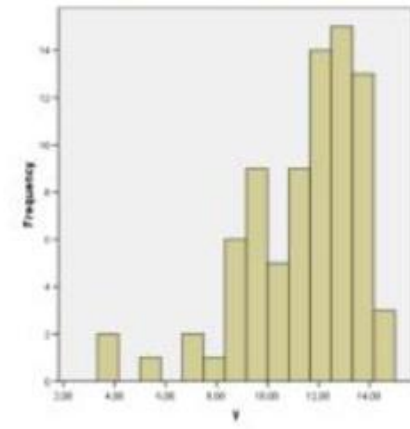
Right-Skewed



Symmetric



Left-Skewed



Distribuição Assimétrica à Direita (assimetria positiva):

- A cauda estende-se para a direita (valores mais elevados).
- A maioria dos dados está concentrada à esquerda.
- Assimetria > 0

Distribuição Assimétrica à Esquerda (assimetria negativa):

- A cauda estende-se para a esquerda (valores mais baixos).
- A maioria dos dados está concentrada à direita.
- Assimetria < 0

Distribuição Simétrica:

- As caudas são aproximadamente iguais em ambos os lados.
- Os dados estão distribuídos de forma uniforme em torno do centro.
- Assimetria ≈ 0

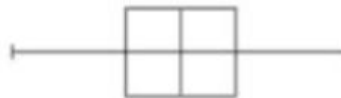
ANÁLISE DA SIMETRIA VIA CAIXA DE BIGODES OU BOXPLOT

O boxplot será abordado num slide posterior.

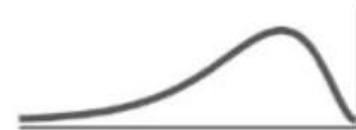
Right-Skewed



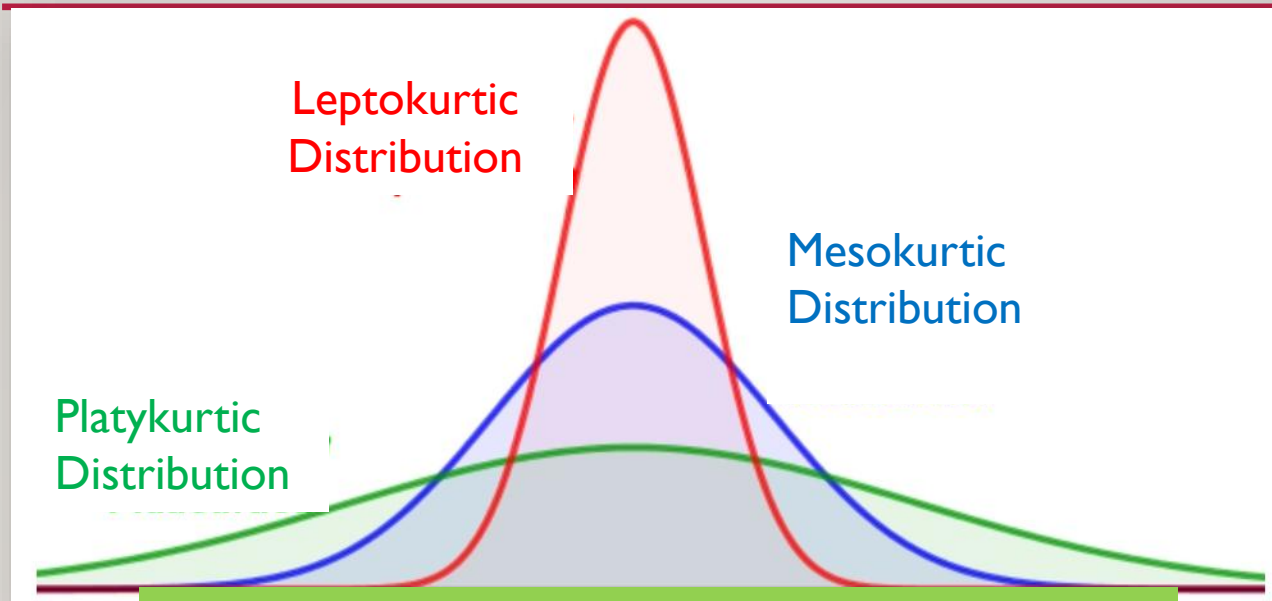
Symmetric



Left-Skewed



ANÁLISE DO ACHATAMENTO



Distribuição Leptocúrtica (pico elevado, caudas pesadas):

- Maior concentração dos dados em torno da média.
- Maior probabilidade de ocorrência de valores extremos (outliers).

Distribuição Platicúrtica (pico baixo, caudas leves):

- Os dados estão mais uniformemente dispersos.
- Menor ocorrência de valores extremos.

Distribuição Mesocúrtica (pico moderado, caudas normais):

- Concentração moderada dos dados em torno da média.

ANÁLISE DA DISPERSÃO

Low Variation (narrow, tall curve)

High Variation (wide, flat curve)

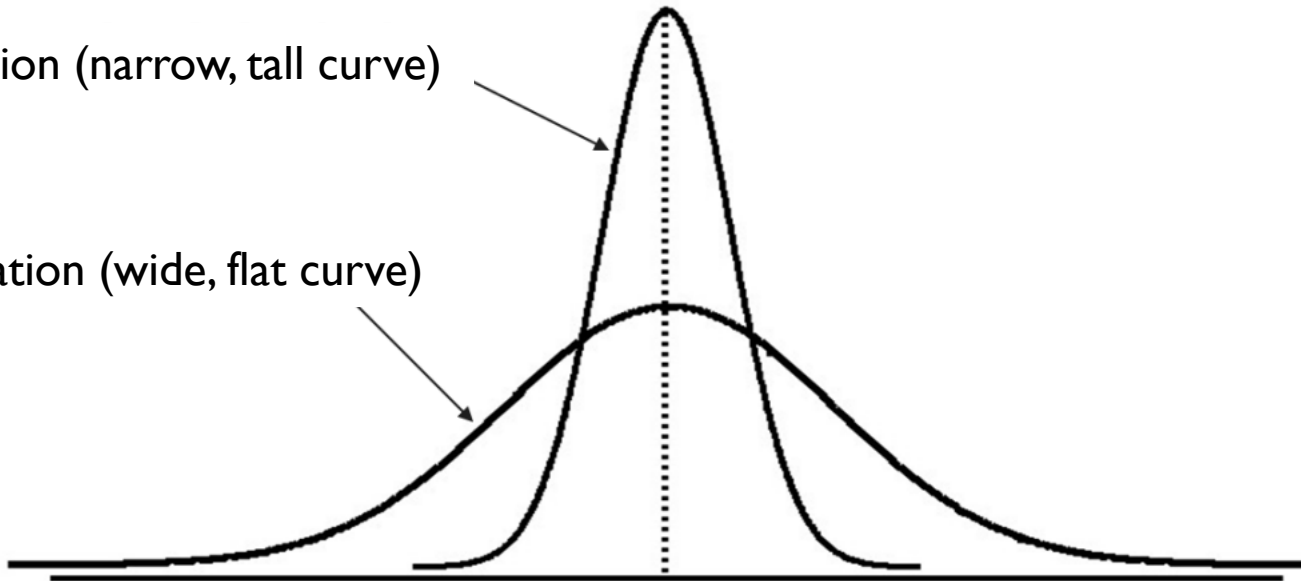


DIAGRAMA DE CAULE-E-FOLHAS: EXEMPLO

Data in ordered array:

21, 24, 24, 26, 27, 27, 30, 32, 38, 41

- Completed stem-and-leaf diagram:

Stem	Leaves
2	1 4 4 6 7 7
3	0 2 8
4	1

Uma forma simples de visualizar os detalhes da distribuição de um conjunto de dados.

Método: Separar a série de dados ordenada em **dígitos iniciais (caule)** e **dígitos finais (folhas)**.

EXERCÍCIO 1.32

1.32 Consider the following data:

17	62	15	65
28	51	24	65
39	41	35	15
39	32	36	37
40	21	44	37
59	13	44	56
12	54	64	59

- a. Construct a frequency distribution.
- b. Construct a histogram.
- c. Construct an ogive.
- d. Construct a stem-and-leaf display

Newbold et al (2013)



EXERCÍCIO 1.32 D): SOLUÇÃO



Resposta:

Step 1 – Data

 Copiar código

17, 62, 15, 65, 28, 51, 24, 65, 39, 41, 35, 15, 39, 32, 36, 37, 40, 21, 44, 37, 59, 13, 44, 56, 12

Step 2 – Determine stems

- Tens digit → **stem**
- Units digit → **leaf**

Step 3 – Assign leaves to stems (unsorted)

- Stem 1 → Leaves: 7, 5, 5, 3, 2 (frequency = 6)
- Stem 2 → Leaves: 8, 4, 1 (frequency = 3)
- Stem 3 → Leaves: 9, 5, 2, 6, 7, 7, 9 (frequency = 7)
- Stem 4 → Leaves: 1, 0, 4, 4 (frequency = 4)
- Stem 5 → Leaves: 1, 4, 6, 9, 9 (frequency = 5)
- Stem 6 → Leaves: 2, 5, 5, 4 (frequency = 4)

EXERCÍCIO 1.32 D): SOLUÇÃO



Resposta:

Step 4 – Order leaves within each stem

- Stem 1 → 2, 3, 5, 5, 5, 7
- Stem 2 → 1, 4, 8
- Stem 3 → 2, 5, 6, 7, 7, 9, 9
- Stem 4 → 0, 1, 4, 4
- Stem 5 → 1, 4, 6, 9, 9
- Stem 6 → 2, 4, 5, 5

Step 5 – Final stem-and-leaf plot with frequencies

markdown

Stem | Leaf | Frequency

1	2 3 5 5 5 7	6
2	1 4 8	3
3	2 5 6 7 7 9 9	7
4	0 1 4 4	4
5	1 4 6 9 9	5
6	2 4 5 5	4

Step 6 – Interpretation

- Most data are concentrated in the 30s.
- Frequencies show which stems contain more values.
- Leaves are ordered to make it easier to read the distribution.

EXERCÍCIO 1.32 D): SOLUÇÃO



Resposta:

SPSS Output: Steam and Leaf Plot

d) Stem-and-Leaf Display

Stem	Leaf	Frequency
1	2 3 5 5 5 7	6
2	1 4 8	3
3	2 5 6 7 7 9 9	7
4	0 1 4 4	4
5	1 4 6 9 9	5
6	2 4 5 5	4

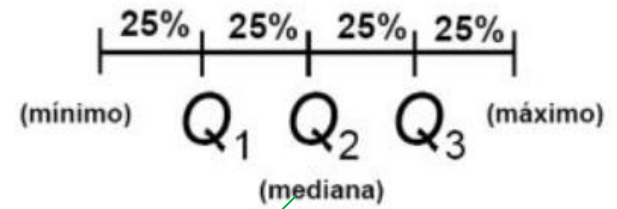
👉 Here, the **stem** is the tens digit, and the **leaf** is the ones digit.
For example, 1 | 2 3 5 5 7 means 12, 13, 15, 15, 17.

Data Gráfico de Ramos e Folhas

Frequency	Stem & Leaf
5,00	1 . 23557
3,00	2 . 148
7,00	3 . 2567799
4,00	4 . 0144
5,00	5 . 14699
4,00	6 . 2455

Largura do ramo: 10
Cada folha: 1 caso(s)

ESTATÍSTICAS DE ORDEM E QUANTIS: DEFINIÇÃO



1 What are Quantiles?

- Quantiles are values that divide a dataset into equal parts.
- Special cases:
 - Quartiles $\rightarrow Q_1, Q_2, Q_3, Q_4$ (divide data into 4 equal parts)
 - Median = Q_2
 - Deciles $\rightarrow D_1, D_2, \dots, D_{10}$ (divide data into 10 equal parts)
 - Median = D_5
 - Percentiles $\rightarrow P_1, P_2, \dots, P_{100}$ (divide data into 100 equal parts)
 - Median = P_{50}

Order- p quantile: q_p with $0 < p < 1$ – the observed value that divides the sample data into two subsets: values below q_p and values above q_p .

Seja x uma variável discreta relativamente à qual se recolheram n observações: $(x_1, x_2, \dots, x_n)$.
Seja $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ com $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, a amostra ordenada correspondente.

PASSOS PARA CALCULAR O QUANTIL DE ORDEM P (Q_p)

Step 1 – Order the sample

- Arrange the data in ascending order:

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}$$

Step 2 – Determine the quantile position

- Compute $(n + 1)p$, where p is the **order of the quantile** ($0 < p < 1$).
- Denote $(n + 1)p = r$ if it is an integer, or $(n + 1)p = r + a$ if not, where r is the integer part and a is the fractional part.

Step 3 – Calculate the quantile

- If r is an integer:

$$q_p = x_{(r)}$$

- If r is not an integer:

$$q_p = (1 - a) x_{(r)} + a x_{(r+1)}$$

PASSOS PARA CALCULAR O QUANTIL DE ORDEM P (Q_p): EXEMPLO

Step 1 – Order the sample

- Data: 12, 15, 7, 20, 18
- Arrange in ascending order:

$$x_{(1)} = 7, x_{(2)} = 12, x_{(3)} = 15, x_{(4)} = 18, x_{(5)} = 20$$

Step 2 – Determine the quantile position

- Suppose we want the **0.4 quantile** ($p = 0.4$, i.e., the 40th percentile).
- Compute $(n + 1)p = (5 + 1) \cdot 0.4 = 2.4$
- Here, $r = 2$ (integer part), $a = 0.4$ (fractional part)

Step 3 – Calculate the quantile

- Since r is not an integer, interpolate:

$$q_{0.4} = (1 - 0.4)x_{(2)} + 0.4x_{(3)} = 0.6 \cdot 12 + 0.4 \cdot 15 = 13.2$$

Note

- Other formulas exist (e.g., SPSS, Excel) → results may **differ slightly**.
- Key: choose **one method** and apply consistently.

PERCENTIS E QUARTIS

Percentiles and Quartiles

To find percentiles and quartiles, data must first be arranged in order from the smallest to the largest values.

The ***P*th percentile** is a value such that approximately $P\%$ of the observations are at or below that number. **Percentiles** separate large ordered data sets into 100ths. The 50th percentile is the median.

The P th percentile is found as follows:

$$P\text{th percentile} = \text{value located in the } (P/100)(n + 1)\text{th ordered position} \quad (2.6)$$

Quartiles are descriptive measures that separate large data sets into four quarters. The **first quartile**, Q_1 , (or 25th *percentile*) separates approximately the smallest 25% of the data from the remainder of the data. The **second quartile**, Q_2 , (or 50th *percentile*) is the median (see Equation 2.3).

Newbold et al (2013)

PERCENTIS, DECIS E QUARTIS

- Percentiles and Quartiles indicate the position of a value relative to the entire set of data
- Generally used to describe large data sets
- Example: An IQ score at the 90th percentile means that 10% of the population has a higher IQ score and 90% have a lower IQ score.

P^{th} percentile = value located in the $\left(\frac{P}{100}\right)(n+1)^{\text{th}}$ ordered position

Newbold et al (2013)

QUARTIS

- Quartiles split the ranked data into 4 segments with an equal number of values per segment (note that the widths of the segments may be different)



- The first quartile, Q_1 , is the value for which 25% of the observations are smaller and 75% are larger
- Q_2 is the same as the median (50% are smaller, 50% are larger)
- Only 25% of the observations are greater than the third quartile

QUARTIL: FÓRMULAS

$$(n + 1)p$$

Find a quartile by determining the value in the appropriate position in the ranked data, where

First quartile position: $Q_1 = 0.25(n + 1)$

Second quartile position:
(the median position) $Q_2 = 0.50(n + 1)$

Third quartile position: $Q_3 = 0.75(n + 1)$

where n is the number of observed values

QUARTIL: EMPLO

- Example: Find the first quartile

Sample Ranked Data: 11 12 13 16 16 17 18 21 22

$(n = 9)$



Q_1 is in the $0.25(9+1) = 2.5$ position of the ranked data
so use the value $Q_1 = (1 - 0.5) \cdot 12 + 0.5 \cdot 13 = 0.5 \cdot 12 + 0.5 \cdot 13 = 12.5$,

so $Q_1 = 12.5$

AMPLITUDE INTER-QUARTIL

- Can eliminate some outlier problems by using the interquartile range
- Eliminate high-and low-valued observations and calculate the range of the middle 50% of the data
- Interquartile range = 3rd quartile – 1st quartile

$$IQR = Q_3 - Q_1$$

Newbold et al (2013)

"The interquartile range (IQR) is **technically defined** as the difference between the third and first quartiles, that is, $IQR = Q_3 - Q_1$. 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 $[Q_1, Q_3]$, 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."

AMPLITUDE INTER-QUARTIL

- The interquartile range (IQR) measures the spread in the middle 50% of the data
- Defined as the difference between the observation at the third quartile and the observation at the first quartile

$$\text{IQR} = Q_3 - Q_1$$

Newbold et al (2013)

CONJUNTO DE CINCO- NÚMEROS SÍNTESE

The **five-number summary** refers to five descriptive measures:

minimum

first quartile

median

third quartile

maximum

$$\text{minimum} < Q_1 < \text{median} < Q_3 < \text{maximum}$$

Newbold et al (2013)

EXERCÍCIO 2.6

2.6 During the last 3 years Consolidated Oil Company expanded its gasoline stations into convenience food stores (CFSs) in an attempt to increase total sales revenue. The daily sales (in hundreds of dollars) from a random sample of 10 weekdays from one of its stores are:

6 8 10 12 14 9 11 7 13 11

- a. Find the mean, median and mode for this store.
- b. Find the five-number summary.

Newbold et al (2013)



EXERCÍCIO 2.6 B): SOLUÇÃO



Resposta:

Step 1 – Order the sample

$x_{(1)} = 6, x_{(2)} = 7, x_{(3)} = 8, x_{(4)} = 9, x_{(5)} = 10, x_{(6)} = 11, x_{(7)} = 11, x_{(8)} = 12, x_{(9)} = 13, x_{(10)} = 14$

EXERCÍCIO 2.6 B): SOLUÇÃO



Resposta:

Step 2 – Determine quantile positions

- Minimum: $x_{(1)} = 6$
- Maximum: $x_{(10)} = 14$
- Q1 (25th percentile): $p = 0.25$

$$(n + 1)p = (10 + 1) \cdot 0.25 = 2.75$$

- Integer part $r = 2$, fraction $a = 0.75$

$$Q1 = (1 - 0.75)x_{(2)} + 0.75 x_{(3)} = 0.25 \cdot 7 + 0.75 \cdot 8 = 7.75$$

- Median (50th percentile): $p = 0.5$

$$(n + 1)p = 11 \cdot 0.5 = 5.5$$

- $r = 5, a = 0.5$

$$Median = (1 - 0.5)x_{(5)} + 0.5 x_{(6)} = 0.5 \cdot 10 + 0.5 \cdot 11 = 10.5$$

- Q3 (75th percentile): $p = 0.75$

$$(n + 1)p = 11 \cdot 0.75 = 8.25$$

- $r = 8, a = 0.25$

$$Q3 = (1 - 0.25)x_{(8)} + 0.25 x_{(9)} = 0.75 \cdot 12 + 0.25 \cdot 13 = 12.25$$

EXERCÍCIO 2.6 B): SOLUÇÃO



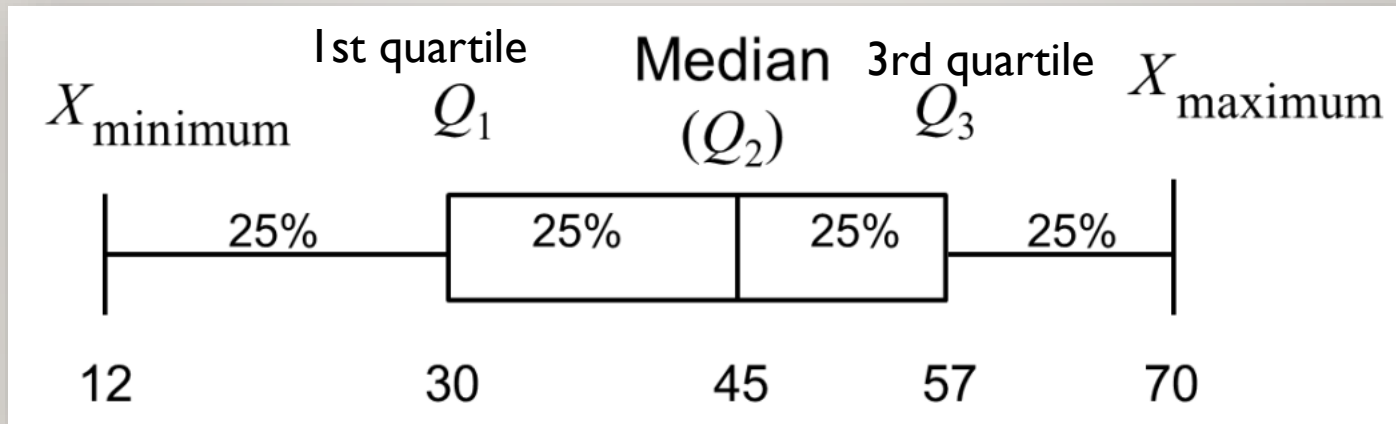
Resposta:

,

Step 3 – Five-number summary

Minimum = 6, $Q1 = 7.75$, Median = 10.5, $Q3 = 12.25$, Maximum = 14

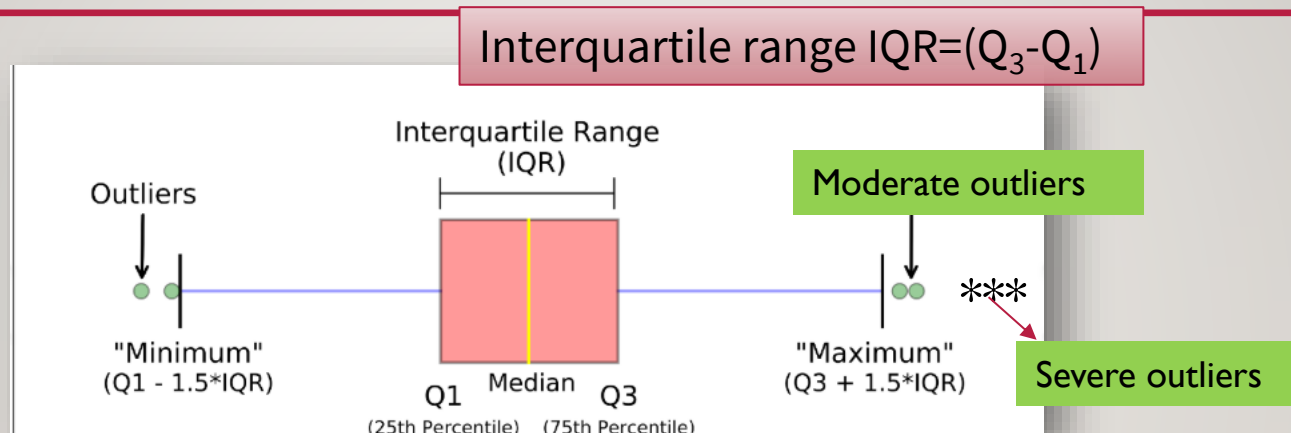
CAIXA COM BIGODES / BOXPLOT: EXEMPLO



Newbold et al (2013)

O gráfico pode ser orientado **horizontal** ou **verticalmente**.

BOXPLOT E OUTLIERS



Moderate outliers (marked with a circle)

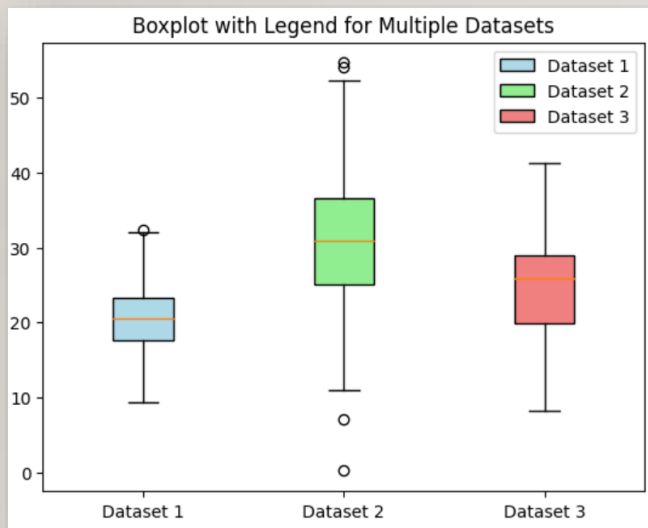
$(Q_1 - 1,5 \times IQR; Q_3 + 1,5 \times IQR)$ Inner fences

Severe outliers (marked with an asterisk)

$(Q_1 - 3 \times IQR; Q_3 + 3 \times IQR)$ Outer fences

COMPARAÇÃO DE CONJUNTOS DE DADOS

- Utilizar medidas de **tendência central, dispersão, assimetria e forma**.
- Visualizar os dados através de **histogramas, boxplots e diagramas de caule e folhas**.
- Identificar **padrões, semelhanças e diferenças** entre conjuntos de dados.



Adding Legend to Boxplot with Multiple Plots - GeeksforGeeks

smokers			non smokers	
		10	0	8
	1	11	1	5 8 9
	3 0	12	3	6 6 8 8 9 9
	4 2 0	13	0	1 7 7 8 8 9
	6 5 3 1	14	0	1 5 6 9 1
9	7 6 5 0	15	2	4 2
2	9 7 6 4 1	16	0	1 8
	9 8 7 4 3	17	4	6
	6 5 3 0	18	1	
	7 5 1	19		
	5 0	20		

www.analyzemath.com

DIAGRAMA DE QUANTIS (GRÁFICO Q-Q)

• Definição

- Um **Q-Q plot (gráfico Q-Q)** é uma ferramenta gráfica utilizada para comparar a distribuição de dois conjuntos de dados ou para verificar se um conjunto de dados segue uma determinada distribuição teórica.

Objetivo

- Se os dados e a distribuição de referência tiverem a mesma forma, os pontos formam uma **linha reta**.
- Os desvios em relação à linha indicam diferenças entre as distribuições.

Aplicações comuns

- **Verificação da normalidade:** comparar os quantis dos dados com os quantis de uma distribuição normal.
- **Comparação de distribuições:** comparar dois conjuntos de dados.
- **Deteção de outliers:** identificar valores que se desviam do padrão esperado.

DIAGRAMA DE QUANTIS (GRÁFICO Q-Q): CONSTRUÇÃO & INTERPRETAÇÃO

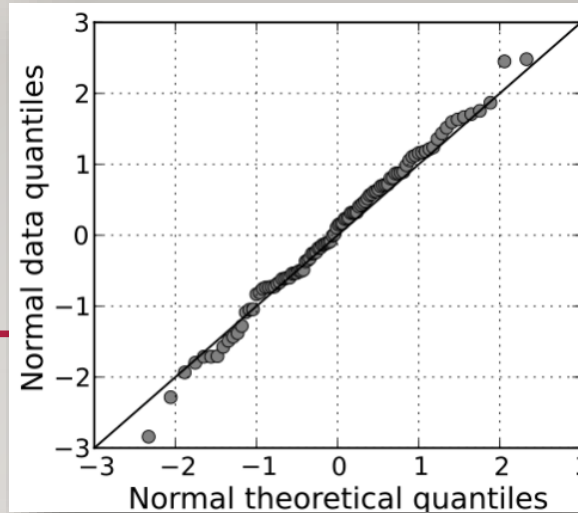
Construction Steps

1. Sort the observed data in ascending order.
2. Calculate the corresponding theoretical quantiles.
3. Plot observed quantiles (y-axis) against theoretical quantiles (x-axis).

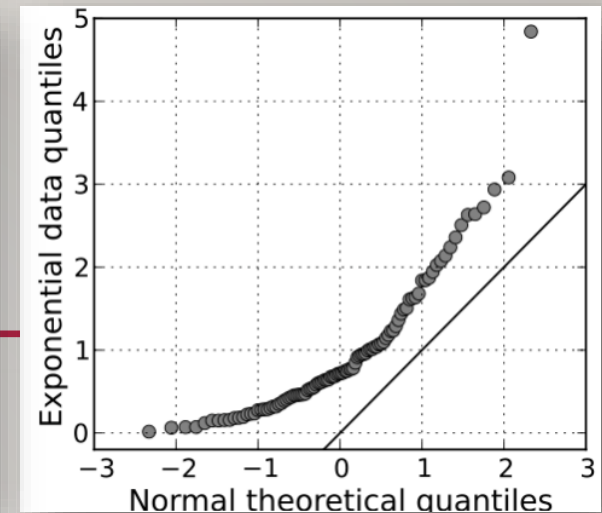
Interpretation

- Points close to a straight line → data follows the reference distribution.
- Points deviating from the line → data differs from the reference distribution.

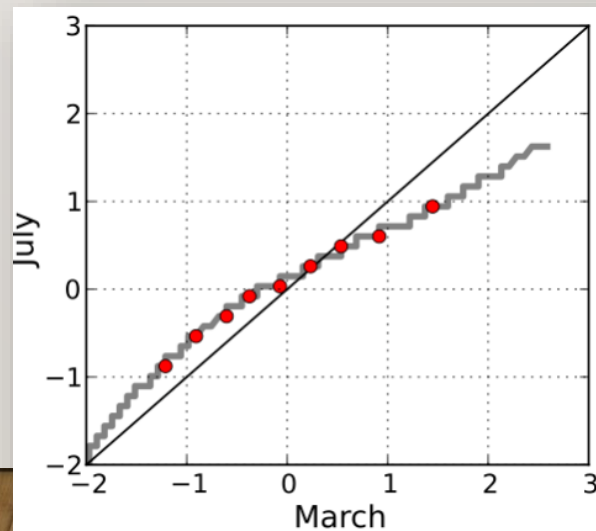
GRÁFICOS QQ: EXEMPLOS



A normal Q-Q plot comparing randomly generated, independent standard normal data on the vertical axis to a standard normal population on the horizontal axis. The linearity of the points suggests that the data are normally distributed.



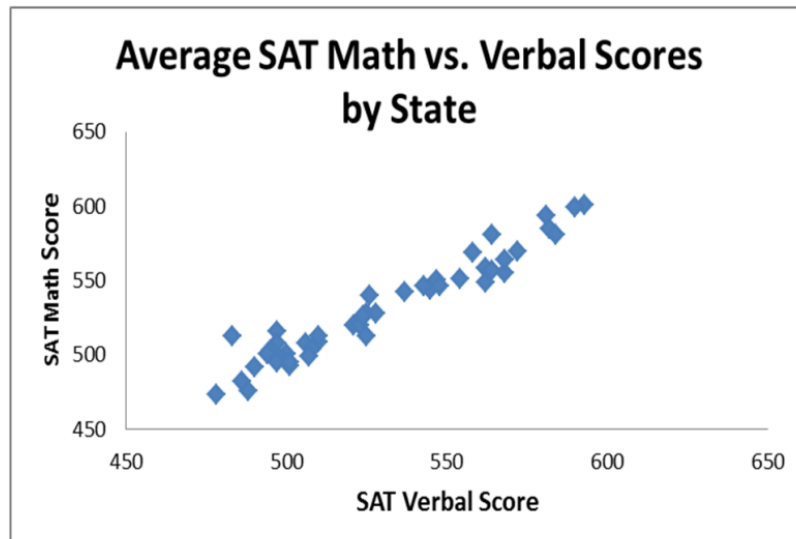
A normal Q-Q plot of randomly generated, independent standard exponential data, ($X \sim \text{Exp}(1)$). This Q-Q plot compares a sample of data on the vertical axis to a statistical population on the horizontal axis. The points follow a strongly nonlinear pattern, suggesting that the data are not distributed as a standard normal ($X \sim N(0,1)$). The offset between the line and the points suggests that the mean of the data is not 0. The median of the points can be determined to be near 0.7.



A Q-Q plot comparing the distributions of standardized daily maximum temperatures at 25 stations in the US state of Ohio in March and in July. The curved pattern suggests that the central quantiles are more closely spaced in July than in March, and that the July distribution is skewed to the left compared to the March distribution. The data cover the period 1893–2001.

GRÁFICO DE DISPERSÃO /SCATTERPLOT: EXEMPLO

Average SAT scores by state: 1998		
	Verbal	Math
Alabama	562	558
Alaska	521	520
Arizona	525	528
Arkansas	568	555
California	497	516
Colorado	537	542
Connecticut	510	509
Delaware	501	493
D.C.	488	476
Florida	500	501
Georgia	486	482
Hawaii	483	513
...		
W.Va.	525	513
Wis.	581	594
Wyo.	548	546



Newbold et al (2013)

Os diagramas de dispersão (Scatter Diagrams) são utilizados para observações emparelhadas provenientes de duas variáveis numéricas.

Uma variável é representada no **eixo vertical** e a outra variável é representada no **eixo horizontal**.

CRONOGRAMA

Definição

- Um cronograma é uma representação gráfica que apresenta a **sequência temporal ou a ordem dos acontecimentos**. Em Estatística, é frequentemente utilizado para **visualizar a ocorrência ou o momento em que as observações acontecem**, ajudando a detetar tendências, padrões ou irregularidades ao longo do tempo.

Objetivo

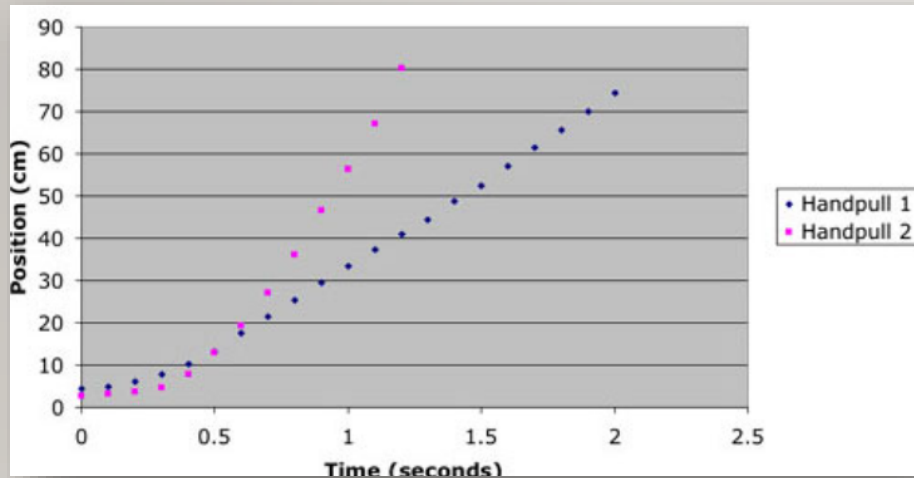
- **Apresentar os dados ou acontecimentos por ordem cronológica.**
- **Identificar padrões, tendências ou anomalias** no momento em que os acontecimentos ocorrem.
- **Apoiar a análise descritiva** de dados relacionados com o tempo.

Principais características

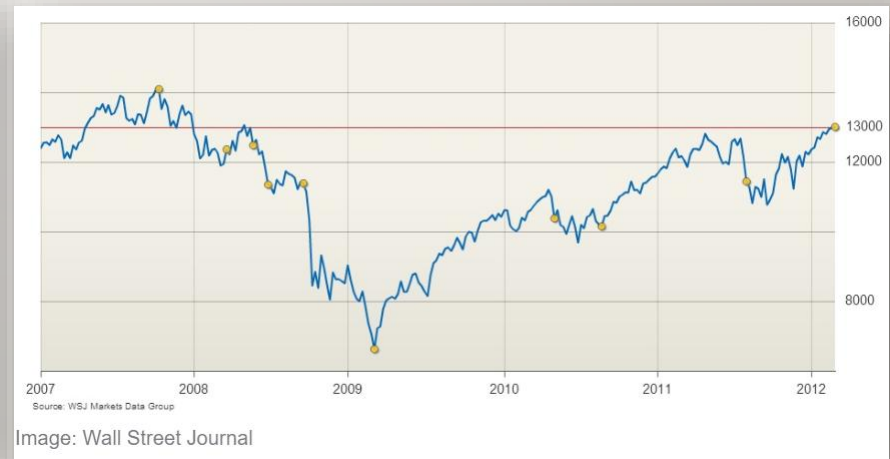
- O **tempo ou a sequência** é representado num dos eixos (geralmente no eixo horizontal — eixo x).
- As **observações ou acontecimentos** são representados no eixo vertical (eixo y).
- É útil no **controlo de qualidade, na análise de sobrevivência e em estudos de eventos**.

Silvestre (2007)

CRONOGRAMA: EXEMPLOS



O gráfico seguinte apresenta um **gráfico temporal** relacionado com a **Física**, representando a **posição em função do tempo** para duas fitas de papel de faíscas (*spark tapes*) que foram puxadas através de um **temporizador de faíscas**, a diferentes velocidades constantes.



Embora um **gráfico temporal** possa ser semelhante a um **diagrama de dispersão**, apresentando uma série de pontos, é frequente encontrar estes gráficos com os **pontos ligados por linhas**, sobretudo em publicações financeiras, como o *The Wall Street Journal*.

OBRIGADA!

Questões?

