STATISTICAL LABORATORY

Applied Mathematics for Economics and Management
1st Year/1st Semester
2025/2026

CONTACT

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

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

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

PROGRAM

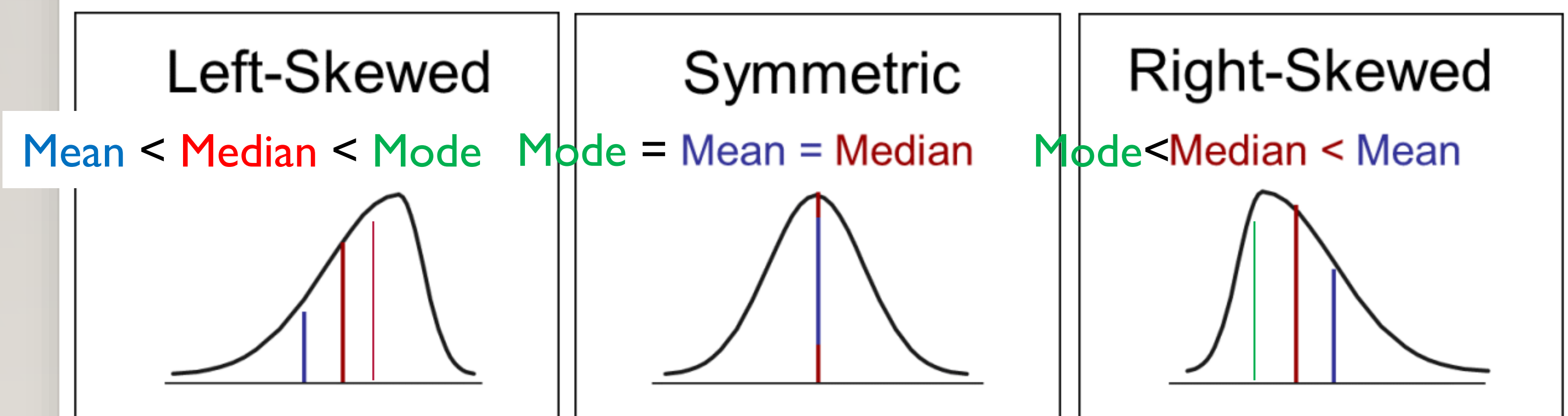
1. Fundamental Concepts of Statistics
2. Exploratory Data Analysis
3. Organizing and Summarizing Data
4. Association and Relationships Between Variables
5. Index Numbers
6. Time Series Analysis

LECTURE 5: MEASURES OF SKEWNESS

IDENTIFYING SKEWNESS USING MEAN, MEDIAN, AND MODE

- Skewness can be identified by the relationship between Mode, Median, and Mean.

- Mean > Median > Mode → positively skewed (right-skewed)
- Mean < Median < Mode → negatively skewed (left-skewed)
- Mean ≈ Median ≈ Mode → approximately symmetric



PEARSON'S COEFFICIENT OF SKEWNESS

According to **Silvestre (2007)**, the coefficient of skewness is defined as:

$$\boxed{g = \frac{\bar{x} - Mo}{s}}$$

where:

- $\bar{x}$ = sample mean
- Mo = mode
- s = standard deviation

Interpretation:

- $g = 0$: symmetric distribution
- $g > 0$: positively skewed (right tail)
- $g < 0$: negatively skewed (left tail)

BOWLEY'S COEFFICIENT OF SKEWNESS

According to **Silvestre (2007)**, Bowley's coefficient of skewness is defined as:

$$\boxed{g' = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}}$$

where:

- Q_1 = first quartile
- Q_2 = median (second quartile)
- Q_3 = third quartile

Interpretation:

- $g' = 0$: symmetric distribution
- $g' > 0$: positively skewed (right tail)
- $g' < 0$: negatively skewed (left tail)

Note:
This measure is **based on quartiles**, and therefore **less affected by extreme values** compared to **Pearson's coefficient**.

MOMENT COEFFICIENT OF SKEWNESS

According to **Silvestre (2007)**, the moment coefficient of skewness is defined as:

$$b_1 = \frac{m_3^2}{m_2^3}$$

or equivalently,

$$\boxed{g_1 = \sqrt{b_1} = \frac{m_3}{m_2^{3/2}}}$$

where:

- m_2 = **second central moment about the mean (variance)**
- m_3 = **third central moment about the mean**, given by:

$$\boxed{m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3} \text{(for ungrouped data)}$$

$$\boxed{m_3 = \frac{1}{n}\sum_{j=1}^{m} n_j(x_j' - \bar{x})^3} \quad \text{(for grouped data)}$$

where:

- m = number of classes
- x_j' = class midpoint
- n_j = class frequency
- $\bar{x}$ = grouped mean

Interpretation:

- $g_1 = 0$: symmetric distribution
- $g_1 > 0$: positively skewed (right tail)
- $g_1 < 0$: negatively skewed (left tail)

COMPARISON OF SKEWNESS COEFFICIENTS

Coefficient	Formula	Based on	Advantages ✅	Disadvantages ❌
Pearson's g	$g = \frac{\bar{x} - Mo}{s}$	Mean, mode, standard deviation	Easy to compute; intuitive	Sensitive to outliers; requires mode estimation
Bowley's g'	$g' = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}$	Quartiles (Q_1, Q_2, Q_3)	Robust to extreme values; useful for ordinal data	Less precise for small datasets; ignores all data outside quartiles
Moment g_1	$g_1 = \frac{m_3}{m_2^{3/2}}$	Central moments about the mean	Theoretical; widely used in inferential statistics	Sensitive to outliers; requires all data points

Interpretation (for all):

- $g = 0$: symmetric distribution
- $g > 0$: positively skewed (right tail)
- $g < 0$: negatively skewed (left tail)

SUMMARY: IDENTIFYING SKEWNESS

Graphical Methods:

- Histogram → continuous variables
- **Bar chart → discrete or ordinal variables (use ordinal only if 5 or more classes)**
- Boxplot → observe the **position of the median** relative to Q1 and Q3

Relation of Measures:

- Another way to identify skewness is through the relationship between **mean, median, and mode**:
  - Symmetric: mean ≈ median ≈ mode
  - Positively skewed: mean > median > mode
  - Negatively skewed: mean < median < mode

Numerical Methods:

- **Pearson's coefficient (g):** mean and mode (requires numeric coding)
- **Bowley's coefficient (g'):** quartiles (robust, suitable for ordinal data)
- **Moment coefficient (g_1):** central moments (requires numeric coding)

Key Points:

- Graphs give a **visual impression** of skewness
- Measures and coefficients give a **quantitative assessment**
- For ordinal variables with 5+ classes, **boxplots and Bowley's coefficient** are particularly useful

EXAMPLE 1: STUDY OF SKEWNESS FOR UNGROUPED DATA

Data (number of complaints in 10 firms):

$$X = [2, 3, 3, 4, 4, 5, 5, 6, 7, 9]$$

Step 1: Basic statistics

- Mean: $\bar{x} = 4.8$
- Median: $Me = 4.5$
- Mode: $Mo = 4$

Step 2: Skewness by relation of measures

- Mean > Median > Mode → **positively skewed**

Step 3: Pearson's coefficient

- Standard deviation: $s \approx 1.87$

$$g = \frac{\bar{x} - Mo}{s} = \frac{4.8 - 4}{1.87} \approx 0.43$$

EXAMPLE 1: STUDY OF SKEWNESS FOR UNGROUPED DATA

Data (number of complaints in 10 firms):

$$X = [2, 3, 3, 4, 4, 5, 5, 6, 7, 9]$$

Step 4: Bowley's coefficient (quartiles)

- Q1 = 3.25, Q2 = 4.5, Q3 = 6

$$g' = \frac{(Q3 - Q2) - (Q2 - Q1)}{(Q3 - Q2) + (Q2 - Q1)} = \frac{(6 - 4.5) - (4.5 - 3.25)}{(6 - 4.5) + (4.5 - 3.25)} = \frac{1.5 - 1.25}{1.5 + 1.25} = \frac{0.25}{2.75} \approx 0.09$$

Step 5: Moment coefficient

- Second central moment: $m_2 = s^2 = 3.5$
- Third central moment: $m_3 \approx 0.53$

$$g_1 = \frac{m_3}{m_2^{3/2}} \approx \frac{0.53}{3.5^{1.5}} \approx 0.09$$

Conclusion:

- All three methods indicate a **slight positive skewness**, consistent with **mean > median > mode.**

EXAMPLE 2: STUDY OF SKEWNESS FOR GROUPED DATA

Example: Skewness Analysis for Grouped Data (3 Classes)

Data

Class interval	Midpoint x'_j	Frequency n_j	Relative freq. f_j	Cumulative rel. freq. F^*_j
[0–10]	5	6	0.200	0.200
(10–20]	15	12	0.400	0.600
(20–30]	25	12	0.400	1.000

Total $n = 30$, $h = 10$.

EXAMPLE 2: STUDY OF SKEWNESS FOR GROUPED DATA

1) Mean

$$\bar{X} = \sum f_j x'_j = 0.2 \cdot 5 + 0.4 \cdot 15 + 0.4 \cdot 25 = 17.0$$

2) Median

Median class: first class where $F_j^* \geq 0.5$ → **(10–20]**

$$\text{Med} = l + \frac{0.5 - F^*(l)}{F^*(L) - F^*(l)}\, h = 10 + \frac{0.5 - 0.2}{0.6 - 0.2} \cdot 10 = 10 + 7.5 = 17.5$$

3) Mode

Modal class: class with largest $f(L)$ → **(10–20]**

$$f^* = f(L) - f(l) = 0.4 - 0.2 = 0.2, \quad f^{**} = f(L) - f(L+1) = 0.4 - 0.4 = 0$$

$$\text{Mo} = l + \frac{f^*}{f^* + f^{**}}\, h = 10 + \frac{0.2}{0.2 + 0} \cdot 10 = 20.0$$

EXAMPLE 2: STUDY OF SKEWNESS FOR GROUPED DATA

4) Skewness Coefficients

- Pearson: $g = \frac{\bar{X} - \mathrm{Mo}}{s}$
  - Variance $m_2 = \sum f_j (x'_j - \bar{X})^2 = 44.0$
  - SD $s = \sqrt{m_2} \approx 6.633$
  - $g \approx \frac{17.0 - 20.0}{6.633} \approx -0.452$ (negative skewness)
- Bowley (quartile-based):
  - $Q_1 \approx 11.667,\ Q_2 = \mathrm{Med} = 17.5,\ Q_3 \approx 23.333$
  - $g' = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)} = \frac{(23.333 - 17.5) - (17.5 - 11.667)}{(23.333 - 17.5) + (17.5 - 11.667)} = 0$ (symmetric by quartiles)
- Moment coefficient:

$$g_1 = \frac{m_3}{m_2^{3/2}}, \quad m_3 = \sum f_j (x'_j - \bar{X})^3 = 0$$

$\Rightarrow g_1 = 0$ (moment indicates symmetry)

Conclusion:

- Mean < Median < Mode → suggests **slight negative skew**
- Bowley and moment coefficients = 0 → **quartiles and central moments indicate approximate symmetry**
- Highlights how **different skewness measures can give slightly different indications** for small sample/grouped data.

EXERCISE 1: UNGROUPED DATA

Exercise 1: Ungrouped Data

The number of complaints in 12 firms is recorded as:

$$X = [1, 2, 2, 3, 4, 4, 5, 5, 6, 7, 8, 9]$$

Tasks:

1. Compute the **mean, median, and mode**.
2. Determine the skewness direction using the **relationship mean–median–mode**.
3. Calculate the **Pearson's coefficient of skewness**.
4. Calculate the **Bowley's coefficient** (use quartiles).
5. Calculate the **moment coefficient of skewness**.
6. Compare the results and interpret the overall skewness.

EXERCISE 1: SOLUTION

Answer:

Example 1 — Ungrouped data

Data: $X = [1, 2, 2, 3, 4, 4, 5, 5, 6, 7, 8, 9]$ (n = 12)

Basic stats

- $\bar{x} = 4.667$
- Median $= 4.5$
- Mode $= 4$ (tie: 2,4,5 — we choose 4)
- Second central moment (variance, population) $m_2 = 5.722$ → $s = \sqrt{m_2} = 2.392$
- Third central moment $m_3 = 3.426$

Relation mean–median–mode

$\bar{x} > \text{median} > \text{mode}$ → **positively skewed (right skew)**

EXERCISE 1: SOLUTION

Answer:

Pearson's coefficient

$$\boxed{g = \frac{\bar{x} - Mo}{s} = \frac{4.667 - 4}{2.392} \approx 0.28}$$

Bowley's coefficient (quartile-based)

Quartiles (by standard method): $Q_1 = 2.5,\ Q_2 = 4.5,\ Q_3 = 6.5$

$$\boxed{g' = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)} = \frac{(6.5 - 4.5) - (4.5 - 2.5)}{(6.5 - 4.5) + (4.5 - 2.5)} = 0}$$

(Here Bowley = 0 — quartiles are symmetric even though mean/median/mode and moments show slight positive skew. Bowley can be zero for small samples / particular quartile positions.)

Moment coefficient

$$\boxed{g_1 = \frac{m_3}{m_2^{3/2}} \approx \frac{3.426}{(5.722)^{1.5}} \approx 0.25}$$

Conclusion (Example 1)

- Pearson ≈ 0.28, Moment $g_1 \approx 0.25$ → **slight positive skew**.
- Bowley = 0 (quartile symmetry) — not uncommon with small samples. Overall: **slightly positively skewed** (mean–median–mode and moment/Pearson support this).

EXERCISE 2: GROUPED DATA

Exercise 2: Grouped Data

The number of customer complaints in 40 companies is grouped as follows:

Class (xj')	Frequency (nj)
0–10	5
10–20	8
20–30	12
30–40	10
40–50	5

Tasks:

1. Calculate the **grouped mean**.
2. Estimate the **median and mode**.
3. Determine skewness direction using the **mean–median–mode relationship**.
4. Calculate the **Pearson's coefficient**, **Bowley's coefficient**, and **moment coefficient**.
5. Compare the results and discuss whether the distribution is symmetric, positively skewed, or negatively skewed.

EXERCISE 2: SOLUTION

Answer:

Example 2: Grouped Data – Class Table (corrected)

Class interval	x'_j (Midpoint)	n_j	$f_j = n_j/n$	N_j (cum abs)	F^*_j (cum rel)
[0–10]	5	5	0.125	5	0.125
(10–20]	15	8	0.200	13	0.325
(20–30]	25	12	0.300	25	0.625
(30–40]	35	10	0.250	35	0.875
(40–50]	45	5	0.125	40	1.000

EXERCISE 2: SOLUTION

Answer:

- **Mean (grouped)**

$$\boxed{\bar{X} = \sum f_j x'_j = 0.125(5) + 0.200(15) + 0.300(25) + 0.250(35) + 0.125(45) = 25.50}$$

- **Median** (use median class interpolation with cumulative relative freqs)

  median class is the first with $F^*_j \geq 0.5$ → **[20–30)**.

  Apply

$$\boxed{\mathrm{Med} = L + \frac{0.5 - F^*_{L-1}}{F^*_L - F^*_{L-1}}\, h}$$

  where $L = 20,\ F^*_{L-1} = 0.325,\ F^*_L = 0.625,\ h = 10$. Thus

$$\mathrm{Med} = 20 + \frac{0.5 - 0.325}{0.625 - 0.325} \times 10 = 20 + \frac{0.175}{0.300} \times 10 = 20 + 5.8333 = 25.8333$$

- **Mode** (grouped, using relative frequencies)

  modal class = [20–30) (largest $f_L = 0.30$). Use

$$f^* = f_L - f_{L-1}, \qquad f^{**} = f_L - f_{L+1}$$

$$\mathrm{Mo} = L + \frac{f^*}{f^* + f^{**}}\, h$$

  Here $f_L = 0.30,\ f_{L-1} = 0.20,\ f_{L+1} = 0.25$ so $f^* = 0.10,\ f^{**} = 0.05$. Then

$$\boxed{\mathrm{Mo} = 20 + \frac{0.10}{0.10 + 0.05} \times 10 = 20 + \frac{0.10}{0.15} \times 10 = 20 + 6.6667 = 26.6667}$$

EXERCISE 2: SOLUTION

✅ Answer:

Relation mean — median — mode

$$\bar{X} = 25.50, \quad \mathrm{Med} \approx 25.833, \quad \mathrm{Mo} \approx 26.667$$

So $\bar{X} < \mathrm{Med} < \mathrm{Mo}$ → this indicates a **slight negative (left) skew**.

Central moments (grouped, about grouped mean)

Using midpoints and f_j (population central moments):

- $m_2 = \sum f_j (x'_j - \bar{X})^2 = 144.75$
  $\Rightarrow s = \sqrt{m_2} = 12.0312$
- $m_3 = \sum f_j (x'_j - \bar{X})^3 = -167.25$

(rounded values shown above; calculations use midpoints and f_j.)

EXERCISE 2: SOLUTION

Answer:

Skewness coefficients

- **Pearson (using mean & mode):**

$$\boxed{g = \frac{\bar{X} - \mathrm{Mo}}{s} = \frac{25.50 - 26.6667}{12.0312} \approx -0.097}$$

- **Bowley (quartile-based)** — compute quartiles by interpolation with F^*:
  - $Q_1 \approx 17.5,\ Q_2 = \mathrm{Med} \approx 25.8333,\ Q_3 \approx 33.0$

$$\boxed{g' = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)} \approx \frac{(33.0 - 25.8333) - (25.8333 - 17.5)}{(33.0 - 25.8333) + (25.8333 - 17.5)} \approx -0.022}$$

- **Moment coefficient (Fisher / g_1):**

$$\boxed{g_1 = \frac{m_3}{m_2^{3/2}} = \frac{-167.25}{(144.75)^{1.5}} \approx -0.096}$$

EXERCISE 2: SOLUTION

Answer:

Final interpretation (concise)

- All three numerical measures (Pearson ≈ −0.097, Bowley ≈ −0.022, Moment g_1 ≈ −0.096) are **small in magnitude** and **slightly negative**, and the ordering mean < median < mode also indicates **slight negative skewness**.
- Conclusion: the grouped distribution is **approximately symmetric but with a very small left skew**.

EXERCISE 3: GRAPHICAL ANALYSIS

Exercise 3: Graphical Analysis

For both Exercise 1 and Exercise 2:

1. Draw a **histogram** (or bar chart for Exercise 2) and a **boxplot**.
2. Use the graphs to **visualize skewness**.
3. Compare the graphical impression with the calculated coefficients.

THANKS!

Questions?