STATISTICAL METHODS

Master in Industrial Management,

Operations and Sustainability (MIMOS)

2nd 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. Descriptive Data Analysis
3. Introduction to Inferential Analysis
4. Parametric Hypothesis Testing
5. Non-Parametric Hypothesis Testing
6. Linear Regression Analysis

LECTURE 11 HOMEWORK: QUESTIONS AND SOLUTIONS

EXERCISE 9.12

9.12 A company that receives shipments of batteries tests a random sample of nine of them before agreeing to take a shipment. The company is concerned that the true mean lifetime for all batteries in the shipment should be at least 50 hours. From past experience it is safe to conclude that the population distribution of lifetimes is normal with a standard deviation of 3 hours. For one particular shipment the mean lifetime for a sample of nine batteries was 48.2 hours. Test at the 10% level the null hypothesis that the population mean lifetime is at least 50 hours.

Newbold et al (2013)

EXERCISE 9.12: SOLUTION

✅ Answer:

One-Sample z-Test for the Mean (known Variance)

Step 1 — Hypotheses

$$H_0: \mu \geq 50 \qquad \text{vs} \qquad H_1: \mu < 50$$

Left-tailed Test

(Left-tailed test.)

Step 2 — Data

$$\sigma = 3, \quad n = 9, \quad \bar{x} = 48.2, \quad \alpha = 0.10$$

Since σ is known and the population is normal, use the Z-test.

EXERCISE 9.12: SOLUTION

✅ Answer:

Step 3 — Test statistic

Standard error:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{9}} = 1$$

Test statistic:

$$Z = \frac{\bar{x} - \mu_0}{SE} = \frac{48.2 - 50}{1} = -1.8$$

Step 4 — Critical value / p-value

Critical value for left-tailed test at $\alpha = 0.10$:

RR =] -∞; -1.28]

$$Z_{critical} \approx -1.28$$

Comparison: $Z = -1.8 < -1.28$ → falls in the rejection region.

p-value: $P(Z \leq -1.8) \approx 0.0359$.

P-value = 1- 0.9641 = 0.0359
Approximate value (from the Standard normal distribution table)

EXERCISE 9.12: SOLUTION

✅ Answer:

Step 5 — Decision and conclusion

- Because $Z = -1.8$ is less than -1.28 (and $p\text{-value} \approx 0.036 < 0.10$), we **reject** H_0 at the 10% significance level.
- **Conclusion:** There is sufficient evidence at the 10% level to conclude the population mean lifetime is **less than 50 hours**. The shipment does not meet the company's requirement.

EXERCISE 9.26

9.26 An IT consultancy in Singapore that offers telephony solutions to small businesses claims that its new call-handling software will enable clients to increase successful inbound calls by an average of 75 calls per week. For a random sample of 25 small-business users of this software, the average increase in successful inbound calls was 70.2 and the sample standard deviation was 8.4 calls. Test, at the 5% level, the null hypothesis that the population mean increase is at least 75 calls. Assume a normal distribution.

Newbold et al (2013)

EXERCISE 9.26: SOLUTION

Answer:

Step 1 – Problem Setup

One-Sample t-Test for the Mean (unknown Variance)

- Sample size: $n = 25$
- Sample mean: $\bar{x} = 70.2$
- Sample standard deviation: $s = 8.4$
- Null hypothesis: $H_0 : \mu \geq 75$
- Alternative hypothesis: $H_1 : \mu < 75$ (left-tailed test)
- Significance level: $\alpha = 0.05$
- Test: **one-sample t-test** (population standard deviation unknown)

Left-tailed Test

Step 2 – Test Statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{70.2 - 75}{8.4/\sqrt{25}} = \frac{-4.8}{8.4/5} = \frac{-4.8}{1.68} \approx -2.857$$

Degrees of freedom: $df = n - 1 = 24$

EXERCISE 9.26: SOLUTION

✅ Answer:

Step 3 – Critical Value

- Left-tailed test, $\alpha = 0.05$, df = 24 → $t_{0.05,24} \approx -1.711$

Decision rule: Reject H_0 if $t < -1.711$

RR =] -∞; -1.711]

Step 4 – p-value

- Using t-distribution with 24 df:

$$p\text{-value} = P(T < -2.857) \approx 0.004$$

P-value ~ 0.005
Approximate value (from the Student's t-distribution table)

Step 5 – Conclusion

- Calculated t = -2.857 < -1.711 → **reject H_0**
- p-value ≈ 0.004 < 0.05

Interpretation: There is **sufficient evidence** at the 5% significance level to conclude that the population mean increase in successful inbound calls is **less than 75 calls per week**.

EXERCISE 14.2

14.2 A 2008 survey investigated favorite water sports in Australia, and it found out that 45% of the interviewees voted for surfing, 40% voted for scuba diving, and the rest voted for other water sports. In 2011, a similar survey was conducted; out of a sample of 200 respondents, 102 declared they prefer surfing, 82 chose scuba diving, and the remaining 16 selected other water sports. Is it possible to conclude at the 5% level that in 2011 these preferences remained the same?

Newbold et al (2013)

EXERCISE 14.2: SOLUTION

✅ Answer:

Chi-Square Goodness-of-Fit Test

Step 1: Hypotheses + Observed vs. Expected Frequencies

$$H_0 : p_{\text{surfing}} = 0.45, \; p_{\text{scuba}} = 0.40, \; p_{\text{other}} = 0.15$$

$$H_a : \text{Preferences in 2011 differ from 2008}$$

Sample size: $n = 200$

Water Sport	Observed O_i	Expected $E_i = n \cdot p_i$
Surfing	102	200 × 0.45 = 90
Scuba	82	200 × 0.40 = 80
Other	16	200 × 0.15 = 30

EXERCISE 14.2: SOLUTION

✅ Answer:

Step 2: Test Statistic

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(102-90)^2}{90} + \frac{(82-80)^2}{80} + \frac{(16-30)^2}{30}$$

Compute each term:

$$\frac{(102-90)^2}{90} = \frac{12^2}{90} = \frac{144}{90} \approx 1.60$$

$$\frac{(82-80)^2}{80} = \frac{2^2}{80} = \frac{4}{80} = 0.05$$

$$\frac{(16-30)^2}{30} = \frac{(-14)^2}{30} = \frac{196}{30} \approx 6.53$$

$$\chi^2 \approx 1.60 + 0.05 + 6.53 = 8.18$$

EXERCISE 14.2: SOLUTION

✅ Answer:

Step 3: Rejection Region

- Degrees of freedom: $df = k - 1 = 3 - 1 = 2$
- Significance level: $\alpha = 0.05$
- Critical value: $\chi^2_{0.05,2} \approx 5.991$
- **Reject** H_0 **if** $\chi^2 > 5.991$

RR = [5.991; +∞[

Step 4: P-value

$$p = P(\chi^2_2 > 8.18) \approx 0.017$$

Step 5: Conclusion

- $\chi^2 = 8.18 > 5.991 \Rightarrow$ in the rejection region
- P-value $p = 0.017 < 0.05$

P-value ~ 0.01
Approximate value (from the Chi-square distribution table)

Decision: Reject H_0

Interpretation (slide-ready): There is **statistically significant evidence** that water sports preferences in 2011 **differ from those in 2008**.

LECTURE 12 EXAM SOLUTIONS: DESCRIPTIVE STATISTICS AND INFERENCE

EXERCISE 1

1. A dataset has a leptokurtic distribution. Which of the following statements is correct?[1.0]

a) The distribution has a flatter peak than a normal distribution.

b) The distribution has heavier tails than a normal distribution.

c) The distribution always has kurtosis equal to 3.

d) The skewness of the distribution is always zero.

e) None of the above.

EXERCISE I: SOLUTION

✅ Answer:

a) **The distribution has a flatter peak than a normal distribution.**

❌ False — that describes *platykurtic*, not leptokurtic.

b) **The distribution has heavier tails than a normal distribution.**

✅ True — leptokurtic distributions have heavier tails.

c) **The distribution always has kurtosis equal to 3.**

❌ False — kurtosis = 3 is the normal distribution (mesokurtic).

d) **The skewness of the distribution is always zero.**

❌ False — kurtosis has nothing to do with skewness; a leptokurtic distribution can be skewed.

e) **None of the above.**

❌ False — because (b) is correct.

✅ **Correct answer: b) The distribution has heavier tails than a normal distribution.**

EXERCISE 2

2. A simple linear regression analysis was performed. It can be assumed that: [1.0]

a) All variables are qualitative.
b) A histogram was constructed.
c) The coefficient of determination was calculated.
d) The coefficient of variation was calculated.
e) A contingency table was constructed.

EXERCISE 2: SOLUTION

✅ Answer:

a) **All variables are qualitative.**

❌ False — regression requires **quantitative** variables.

b) **A histogram was constructed.**

❌ False — a histogram is not required for regression analysis.

c) **The coefficient of determination was calculated.**

✅ True — R^2 **is always computed** in simple linear regression.

d) **The coefficient of variation was calculated.**

❌ False — regression does not require or guarantee a CV calculation.

e) **A contingency table was constructed.**

❌ False — contingency tables are used for qualitative variables, not regression.

✅ **Correct answer: c) The coefficient of determination was calculated.**

EXERCISE 3

3. In a financial company, the profit generated by each investment over the past year has a population mean of €8,200 and a population standard deviation of €3,600. Consider a random sample of 100 investments.

3.1. The sampling distribution of the mean of the 100 investments is approximately: [2.5]

a) $t_{(99)}$.
b) $N(0, 1)$.
c) $N(\mu = 8200, \sigma = 3600)$.
d) $N(\mu = 8200, \sigma = 360)$.
e) None of the above.

3.2. The probability that the sample mean differs from the population mean by more than €806.76 is: [2.5]

a) 0.0125.
b) 0.025.
c) 0.9575.
d) 0.9899.
e) None of the above.

EXERCISE 3.1: SOLUTION

✅ Answer:

3.1. Sampling distribution of the sample mean

- Population mean $\mu = 8200$.
- Population standard deviation $\sigma = 3600$.
- Sample size $n = 100$.
- By the Central Limit Theorem (and because σ is known), the sampling distribution of the sample mean $\bar{X}$ is approximately normal with

$$\bar{X} \sim N\left(\mu = 8200,\ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3600}{\sqrt{100}} = 360\right).$$

✔ **Correct answer: d)** $N(\mu = 8200,\ \sigma = 360)$.

EXERCISE 3.2: SOLUTION

Answer:

3.2. Probability that $|X - \mu| > 806.76$

We compute the Z–score for the deviation:

$$Z = \frac{806.76}{\sigma_{\bar{X}}} = \frac{806.76}{360} = 2.241.$$

This is a two–sided probability:

$$\boxed{P\left(|\bar{X} - \mu| > 806.76\right) = 2\left(1 - \Phi(2.241)\right).}$$

Using the standard normal CDF,

$$\Phi(2.241) \approx 0.98749 \quad \Rightarrow \quad 2(1 - \Phi(2.241)) \approx 0.02503 \approx 0.025.$$

✔ **Correct answer: b) 0.025.**

EXERCISE 4

4. A pharmaceutical company produces tablets in which the variability of the active ingredient from one tablet to another must be very small. The population standard deviation is supposedly 1 milligram. Inspectors from the Ministry of Health selected a random sample of 16 tablets. Assuming the population is normally distributed, the probability that the corrected sample variance exceeds 0.736 mg² is: [2.5]

a) 0.25.
b) 0.30.
c) 0.75.
d) 0.78.
e) None of the above.

EXERCISE 4: SOLUTION

✅ Answer:

We use the fact that for a normal population the corrected sample variance S^2 satisfies

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Given $\sigma = 1$, $n = 16$ and the threshold 0.736, compute the chi-square value:

$$\chi^2_{\text{obs}} = \frac{(n-1)\,0.736}{1^2} = 15 \times 0.736 = 11.04.$$

We want

$$P\big(S^2 > 0.736\big) = P\big(\chi^2_{15} > 11.04\big).$$

Using the chi-square distribution (df = 15),

$$P(\chi^2_{15} > 11.04) \approx 0.7498 \approx 0.75.$$

✔ **Correct answer: c) 0.75.**

EXERCISE 5

5. The daily price (€) per liter of crude oil on the stock market is a random variable, X, with distribution $N(\mu, \sigma)$. To better understand this price, a random sample of 30 days was selected, and the corresponding daily prices per liter of crude oil were recorded. From this data, the sample mean and corrected sample variance were obtained: $\bar{x}$ = €1.16 and s^2 = 0.1567 €^{2}. A 95% confidence interval for the true mean daily price per liter of crude oil on the stock market is given by: [3.5]

a) (1.0183; 1,3017).
b) (1.0122; 1.3078).
c) (1.0148; 1.3052).
d) (1.0411; 1.2789).
e) None of the above.

EXERCISE 5: SOLUTION

Answer:

Given: $\bar{x} = 1.16, \; s^2 = 0.1567, \; n = 30$, confidence level 95%.

Since σ is unknown and $n = 30$, use the t-distribution with $df = n - 1 = 29$.

1. Compute the sample standard deviation s:

$$s = \sqrt{0.1567} = 0.3958535083 \text{ (approx.)}$$

2. Standard error of the mean:

$$SE = \frac{s}{\sqrt{n}} = \frac{0.3958535083}{\sqrt{30}} = 0.07227263198 \text{ (approx.)}$$

3. t-critical value for a two-sided 95% CI with 29 degrees of freedom:

$t_{0.975;\,29} \approx 2.045.$

EXERCISE 5: SOLUTION

✅ Answer:

4. Margin of error:

$$\text{ME} = t_{0.975;\,29} \times SE = 2.045 \times 0.07227263198 \approx 0.1477975324$$

5. 95% confidence interval:

$$\bar{x} \pm \text{ME} = 1.16 \pm 0.1477975324$$

$$\Rightarrow (1.0122024676,\ 1.3077975324)$$

Rounded to the 4 decimals shown in the choices, this is approximately **(1.0122; 1.3078)**.

✔ **Correct answer: b) (1.0122; 1.3078).**

EXERCISE 6

6. In a confidence interval for the population proportion, p, with 95% confidence, if the sample size decreases, the margin of error: [1.0]

a) Increases.

b) Decreases.

c) Does not change.

d) Becomes zero.

e) Becomes negative.

EXERCISE 6: SOLUTION

Answer:

The **margin of error (ME)** for a proportion is given by:

$$ME = z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Where:

- $z_{1-\alpha/2}$ the critical value from the standard normal distribution (for 95% confidence, $z_{0.975} \approx 1.96$),
- $\hat{p}$ is the sample proportion,
- n is the sample size.

Step 1: Look at the formula

$$ME \propto \frac{1}{\sqrt{n}}$$

- This means as n **increases, the margin of error decreases.**
- Conversely, **if n decreases, the margin of error increases.**

EXERCISE 6: SOLUTION

✅ Answer:

Step 2: Analyze the options

a) **Increases.** ✅ This matches the formula.

b) **Decreases.** ❌ This is the opposite of what happens.

c) **Does not change.** ❌ The margin of error depends on n, so it changes.

d) **Becomes zero.** ❌ It cannot become zero unless n is infinite, which is not the case.

e) **Becomes negative.** ❌ Margin of error is always positive.

✅ Step 3: Correct answer

a) **Increases.**

EXERCISE 7

7. In a hypothesis test H_0: μ ≥ 25 versus H_1: μ < 25, the p-value obtained was 0.057. For a significance level of α = 0.05, which of the following conclusions is appropriate? [1.0]

a) There is statistical evidence to conclude that the population mean is 25.

b) There is no statistical evidence to conclude that the population mean is less than 25.

c) There is statistical evidence to conclude that the population mean is less than 25.

d) There is no statistical evidence to conclude that the population mean is 25.

e) There is no statistical evidence to conclude that the population mean is greater than or equal to 25.

EXERCISE 7: SOLUTION

✅ Answer:

We are given:

- Null hypothesis: $H_0 : \mu \geq 25$
- Alternative hypothesis: $H_1 : \mu < 25$ (this is a **left-tailed test**)
- p-value = 0.057
- Significance level: $\alpha = 0.05$

Step 1: Compare p-value with α

- **Rule:** If $p\text{-value} \leq \alpha$, reject H_0.
- Here: $p = 0.057 > 0.05$ → **do not reject** H_0.

Step 2: Interpret the conclusion

- "Do not reject H_0" means **there is not enough statistical evidence to support** H_1.
- H_1 is $\mu < 25$, so we **cannot conclude the mean is less than 25.**

EXERCISE 7: SOLUTION

✅ Answer:

Step 3: Check the options

a) **There is statistical evidence to conclude that the population mean is 25.** ❌

- We never "conclude" a specific value; we only fail to reject H_0.

b) **There is no statistical evidence to conclude that the population mean is less than 25.** ✅

- This matches our reasoning.

c) **There is statistical evidence to conclude that the population mean is less than 25.** ❌

- p-value > 0.05 → no evidence for H_1.

d) **There is no statistical evidence to conclude that the population mean is 25.** ❌

- The test does not prove equality; failing to reject H_0 does **not prove** $\mu = 25$.

e) **There is no statistical evidence to conclude that the population mean is greater than or equal to 25.** ❌

- Actually, we **do not reject** H_0, so there is no evidence against $\mu \geq 25$.

✅ Step 4: Correct answer

b) There is no statistical evidence to conclude that the population mean is less than 25.

EXERCISE 8 TO 10

8 to 10. The store "Grand Sales" has observed in recent years that 15% of its customers pay for their purchases by check, 38% by credit card, 32% by debit card, and 15% in cash. A sample of 160 sales made during the week before Christmas revealed the following results:

Payment Type	Check	Credit Card	Debit Card	Cash
Number of Sales	27	65	48	20

Is the type of payment used by "Grand Sales" customers during the Christmas period consistent with the store's historical information (α = 10%)? To answer this question, a statistician performed a statistical analysis and obtained the following outputs from the SPSS software package (Asymp. Sig. = p-value):

	Observed N	Expected N	Residual
Cheque	27	24,0	3,0
Cartão de crédito	65	60,8	4,2
Cartão de débito	48	51,2	-3,2
Dinheiro	20	24,0	-4,0
Total	160		

	Tipo de pagamento
Chi-Square	1,532[a]
df	3
Asymp. Sig.	,675

EXERCISE 8 TO 10

8. The null hypothesis (H_o) of the hypothesis test is (where p_1, p_2, p_3, and p_4 are the proportions of customers who paid by check, credit card, debit card, and cash, respectively): [1.0]

a) H_0: $p_1 = 0.25$, $p_2 = 0.25$, $p_3 = 0.25$, $p_4 = 0.25$.

b) H_0: $p_1 = 0.24$, $p_2 = 0.608$, $p_3 = 0.512$, $p_4 = 0.24$.

c) H_0: $p_1 = 0.15$, $p_2 = 0.38$, $p_3 = 0.32$, $p_4 = 0.15$.

d) H_0: $p_1 = 0.50$, $p_2 = 0.50$, $p_3 = 0.50$, $p_4 = 0.50$.

e) H_0: $p_1 = 0.25$, $p_2 = 0.50$, $p_3 = 0.50$, $p_4 = 0.25$.

EXERCISE 8: SOLUTION

✅ Answer:

We are given:

- Historical proportions of payment types:
  - Check = 15% → $p_1 = 0.15$
  - Credit card = 38% → $p_2 = 0.38$
  - Debit card = 32% → $p_3 = 0.32$
  - Cash = 15% → $p_4 = 0.15$
- Sample size $n = 160$ and observed counts (from table)

Question 8: Null hypothesis H_0

The **null hypothesis** for a chi-square goodness-of-fit test is:

$$H_0 : \text{The observed proportions match the historical proportions.}$$

This means:

$$H_0 : p_1 = 0.15,\ p_2 = 0.38,\ p_3 = 0.32,\ p_4 = 0.15$$

✅ **Correct answer: c)**

EXERCISE 9

9. The rejection region (RR) of the hypothesis test mentioned above is (α = 10%): [2.5]

a) RR =]-∞; 6.251].

b) RR =]-∞; -6.251].

c) RR =]-∞; -6.251] U [6.251;+∞[.

d) RR = [-6.251;+∞[.

e) RR = [6.251;+∞[.

EXERCISE 9: SOLUTION

Answer:

Step 1: State the hypotheses

- **Null hypothesis** H_0: The observed proportions match historical proportions.

$$H_0 : p_1 = 0.15,\ p_2 = 0.38,\ p_3 = 0.32,\ p_4 = 0.15$$

- **Alternative hypothesis** H_1: At least one proportion differs from historical.

$$H_1 : \text{At least one } p_i \neq \text{historical } p_i$$

Step 2: Identify the test statistic

The **chi-square statistic** is:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where:

- O_i = observed frequency
- $E_i = n \cdot p_i$ = expected frequency

EXERCISE 9: SOLUTION

✅ Answer:

Step 3: Determine the rejection region (RR)

- For a chi-square goodness-of-fit test:
  - The test statistic χ^2 **cannot be negative**.
  - RR is in the **right tail** of the chi-square distribution.
  - Degrees of freedom: $df = k - 1 = 4 - 1 = 3$
- At $\alpha = 0.10$, the **critical value** $\chi^2_{0.10,3} \approx 6.251$ (from chi-square table).
- **Rejection region:**

$$RR = \{\chi^2 > 6.251\} \quad \text{or in interval notation: } [6.251, +\infty[$$

Step 4: Match with options

- e) RR = [6.251;+∞[✅ Correct.

Step 5: Notes

- Options a–d involve negative numbers, which are impossible for chi-square.
- Chi-square values are always ≥ 0, so only a right-tail region makes sense.

EXERCISE 10

10. Observing the SPSS outputs (above), it is known that the p-value = 0.675. Therefore, it can be stated that: [1.5]

a) At a 10% significance level, the type of payment used by customers during the Christmas period is consistent with the store's historical information.

b) At a 10% significance level, the type of payment used by customers during the Christmas period is not consistent with the store's historical information.

c) The type of payment used by customers during the Christmas period is consistent with the store's historical information.

d) The type of payment used by customers during the Christmas period is not consistent with the store's historical information.

e) No conclusion can be drawn with the data available.

EXERCISE 10: SOLUTION

✅ Answer:

We are given:

- **p-value = 0.675**
- **Significance level α = 0.10**
- Hypothesis test: chi-square goodness-of-fit (testing whether observed payment types match historical proportions)

Step 1: Compare p-value with α

- **Rule:**
  - If $p \leq \alpha$, reject H_0 → evidence that the distribution differs
  - If $p > \alpha$, **do not reject** H_0 → insufficient evidence to say the distribution differs

Here: $p = 0.675 > 0.10$ → **do not reject** H_0

EXERCISE 10: SOLUTION

✅ Answer:

Step 2: Interpret conclusion

- **Failing to reject** H_0 means:
  - The observed payment types are **consistent with historical proportions**
  - **We cannot say they are inconsistent**

EXERCISE 10: SOLUTION

✅ Answer:

Step 3: Check the options

a) **At a 10% significance level, the type of payment used by customers during the Christmas period is consistent with the store's historical information.** ✅ Correct

b) At a 10% significance level, the type of payment used … is **not consistent** ❌ Wrong (we did not reject H_0)

c) The type of payment used … is consistent … ❌ Slightly overgeneralized; significance level must be mentioned

d) … is not consistent ❌ Wrong

e) No conclusion … ❌ Wrong (we can conclude at the given significance level)

✅ **Step 4: Correct answer**

a) At a 10% significance level, the type of payment used by customers during the Christmas period is consistent with the store's historical information.

THANKS!

Questions?