STATISTICS II

Bachelor's degrees in Economics, Finance and Management

2nd year/2nd Semester
2025/2026

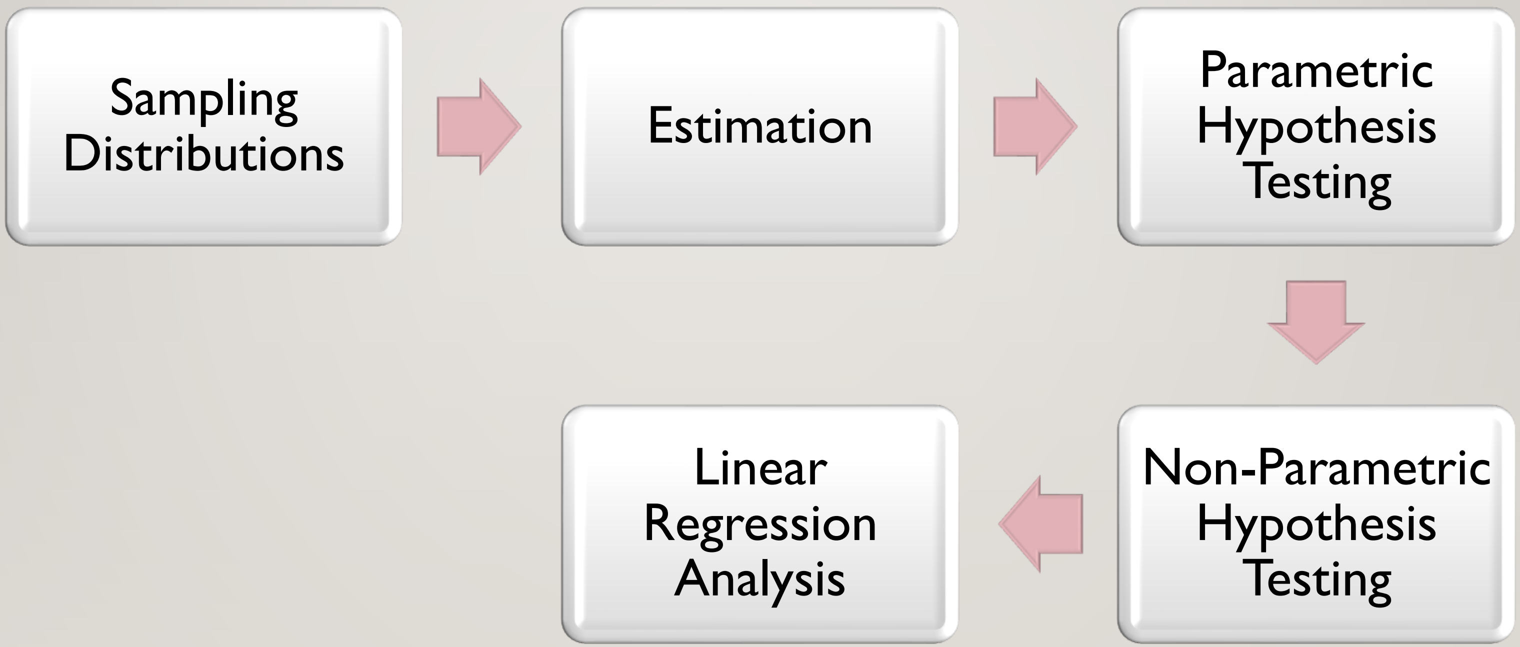
CONTACT

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


Sampling Distributions
Estimation
Parametric Hypothesis Testing
Non-Parametric Hypothesis Testing
Linear Regression Analysis

PRATICAL CLASS 12

Exercises 11.12, 11.25 a), 11.28, 11.32 b), 11.34, 11.35

EXERCISE 11.21

11.21 A corporation administers an aptitude test to all new sales representatives. Management is interested in the extent to which this test is able to predict sales representatives' eventual success. The accompanying table records average weekly sales (in thousands of dollars) and aptitude test scores for a random sample of eight representatives.

Weekly sales	10	12	28	24	18	16	15	12
Test score	55	60	85	75	80	85	65	60

a. Estimate the linear regression of weekly sales on aptitude test scores.
b. Interpret the estimated slope of the regression line.

Newbold et al (2013)

EXERCISE 11.21 A): SOLUTION

Answer:

a) Estimation of the linear regression

Step 1: Compute sample means

$$\bar{x} = \frac{55+60+85+75+80+85+65+60}{8} = \frac{565}{8} = 70.625$$

$$\bar{y} = \frac{10+12+28+24+18+16+15+12}{8} = \frac{135}{8} = 16.875$$

Step 2: Compute S_{xx} and S_{xy}

$$S_{xx} = \sum (x_i - \bar{x})^2 = 906.88$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = 301.88$$

EXERCISE 11.21 A): SOLUTION

✅ Answer:

Step 3: Estimate the regression coefficients

Slope:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{301.88}{906.88} \approx 0.333$$

Intercept:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$\hat{\beta}_0 = 16.875 - 0.333(70.625) \approx -6.66$$

Estimated regression equation

$$\boxed{\hat{Y} = -6.66 + 0.333X}$$

where:

- Y = weekly sales (thousands of dollars)
- X = aptitude test score

EXERCISE 11.21 B): SOLUTION

✅ Answer:

b) Interpretation of the slope

The estimated slope $\hat{\beta}_1 = 0.333$ indicates that:

> **For each additional one-point increase in the aptitude test score, average weekly sales are expected to increase by approximately 0.33 thousand dollars (about $330), on average.**

This suggests a positive relationship between aptitude test performance and sales success.

EXERCISE 11.25 A)

11.25 Compute SSR, SSE, s_e^2, and the coefficient of determination, given the following statistics computed from a random sample of pairs of X and Y observations.

a. $\sum_{i=1}^{n}(y_i - \overline{y})^2 = 100{,}000$, $R^2 = 0.50$, $n = 52$

Newbold et al (2013)

EXERCISE 11.25 A): SOLUTION

✅ Answer:

Given

- Total sum of squares:

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = 100{,}000$$

- Coefficient of determination:

$$R^2 = 0.50$$

- Sample size:

$$n = 52$$

1) Regression sum of squares (SSR)

By definition,

$$R^2 = \frac{SSR}{SST}$$

So,

$$SSR = R^2 \times SST = 0.50 \times 100{,}000 = 50{,}000$$

EXERCISE 11.25 A): SOLUTION

✅ Answer:

2) Error sum of squares (SSE)

$$SSE = SST - SSR = 100{,}000 - 50{,}000 = 50{,}000$$

3) Estimator of the error variance s_e^2

In simple linear regression:

$$s_e^2 = \frac{SSE}{n-2}$$

$$s_e^2 = \frac{50{,}000}{52-2} = \frac{50{,}000}{50} = 1{,}000$$

4) Coefficient of determination

This was already given, but for completeness:

$$R^2 = 0.50$$

EXERCISE 11.28

11.28 Find and interpret the coefficient of determination for the regression of DVD system sales on price, using the following data.

Sales	420	380	350	400	440	380	450	420
Price	98	194	244	207	89	261	149	198

Newbold et al (2013)

EXERCISE 11.28: SOLUTION

✅ Answer:

Step 1: Compute sample means

$$\bar{x} = \frac{98 + 194 + 244 + 207 + 89 + 261 + 149 + 198}{8} = 180$$

$$\bar{y} = \frac{420 + 380 + 350 + 400 + 440 + 380 + 450 + 420}{8} = 405$$

Step 2: Compute sums of squares

$$S_{xx} = \sum (x_i - \bar{x})^2 = 22{,}088$$

$$S_{yy} = \sum (y_i - \bar{y})^2 = 8{,}200$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = -8{,}180$$

Step 3: Compute the correlation coefficient

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{-8{,}180}{\sqrt{22{,}088 \times 8{,}200}} \approx -0.608$$

EXERCISE 11.28: SOLUTION

✅ Answer:

Step 4: Coefficient of determination

$$R^2 = r^2 = (-0.608)^2 \approx 0.37$$

Interpretation

The coefficient of determination $R^2 \approx 0.37$ means that:

> **Approximately 37% of the variation in DVD system sales can be explained by changes in price through the linear regression model.**

The remaining 63% of the variation in sales is due to other factors not included in the model, such as advertising, product features, competition, or random variation.

EXERCISE 11.32 B)

11.32 Given the simple regression model

$$Y = \beta_0 + \beta_1 X$$

and the regression results that follow, test the null hypothesis that the slope coefficient is 0 versus the alternative hypothesis of greater than zero using probability of Type I error equal to 0.05, and determine the two-sided 95% and 99% confidence intervals.

b. A random sample of size $n = 46$ with $b_1 = 5.2 \quad s_{b_1} = 2.1$

Newbold et al (2013)

EXERCISE 11.32 B): SOLUTION

Answer:

Given

- Sample size: $n = 46$
- Estimated slope: $\hat{b}_1 = 5.2$
- Standard error of the slope: $s_{b_1} = 2.1$
- Significance level: $\alpha = 0.05$

Degrees of freedom:

$$df = n - 2 = 44$$

1) Hypothesis test for the slope

Hypotheses

$$H_0 : \beta_1 \leq 0$$

$$H_1 : \beta_1 > 0$$

Test statistic

$$t = \frac{\hat{b}_1}{s_{b_1}} = \frac{5.2}{2.1} \approx 2.48$$

EXERCISE 11.32 B): SOLUTION

✅ Answer:

Critical value

For a **one-sided test** with $\alpha = 0.05$ and $df = 44$:

$$t_{0.95,44} \approx 1.68$$

Decision

$$2.48 > 1.68$$

👉 **Reject H_0**

Conclusion

At the 5% significance level, there is sufficient evidence to conclude that the slope coefficient is **greater than zero**.

Thus, X has a statistically significant positive effect on Y.

EXERCISE 11.32 B): SOLUTION

Answer:

2) Two-sided 95% confidence interval for β_1

Critical value:

$$t_{0.975,44} \approx 2.02$$

$$\hat{b}_1 \pm t \cdot s_{b_1} = 5.2 \pm 2.02(2.1) = 5.2 \pm 4.24$$

$$\boxed{(0.96,\ 9.44)}$$

---

3) Two-sided 99% confidence interval for β_1

Critical value:

$$t_{0.995,44} \approx 2.69$$

$$5.2 \pm 2.69(2.1) = 5.2 \pm 5.65$$

$$\boxed{(-0.45,\ 10.85)}$$

EXERCISE 11.32 B): SOLUTION

✅ Answer:

✅ **Final summary**

- The slope is **significantly greater than zero** at the 5% level.
- **95% CI:** $(0.96,\ 9.44)$ → excludes 0
- **99% CI:** $(-0.45,\ 10.85)$ → includes 0

This is consistent: significance at 5%, but not at 1%.

EXERCISE 11.34

11.34 Mumbai Electronics is planning to extend its marketing region from the western United States to include the midwestern states. In order to predict its sales in this new region, the company has asked you to develop a linear regression of DVD system sales on price, using the following data supplied by the marketing department:

Sales	418	384	343	407	432	386	444	427
Price	98	194	231	207	89	255	149	195

a. Use an unbiased estimation procedure to find an estimate of the variance of the error terms in the population regression.

b. Use an unbiased estimation procedure to find an estimate of the variance of the least squares estimator of the slope of the population regression line.

c. Find a 90% confidence interval for the slope of the population regression line.

Newbold et al (2013)

EXERCISE 11.34: SOLUTION

✅ Answer:

Step 1: Compute sample means

$$\bar{x} = \frac{98 + 194 + 231 + 207 + 89 + 255 + 149 + 195}{8} = 177.25$$

$$\bar{y} = \frac{418 + 384 + 343 + 407 + 432 + 386 + 444 + 427}{8} = 405.13$$

Step 2: Compute sums of squares

$$S_{xx} = \sum (x_i - \bar{x})^2 = 21{,}584$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = -7{,}503$$

$$S_{yy} = \sum (y_i - \bar{y})^2 = 8{,}468$$

Step 3: Estimate the regression coefficients

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-7{,}503}{21{,}584} \approx -0.347$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 405.13 + 0.347(177.25) \approx 466.7$$

Regression line:

$$\hat{Y} = 466.7 - 0.347X$$

EXERCISE 11.34 A): SOLUTION

Answer:

a) Unbiased estimate of the variance of the error terms

First compute:

$$SSR = \frac{S_{xy}^2}{S_{xx}} = \frac{(-7{,}503)^2}{21{,}584} \approx 2{,}609$$

$$SSE = S_{yy} - SSR = 8{,}468 - 2{,}609 = 5{,}859$$

The unbiased estimator of the error variance is:

$$\boxed{\hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{5{,}859}{6} \approx 976.5}$$

EXERCISE 11.34 B): SOLUTION

✅ Answer:

b) Unbiased estimate of the variance of the slope estimator

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{xx}}$$

$$\boxed{\mathrm{Var}(\hat{\beta}_1) = \frac{976.5}{21{,}584} \approx 0.0453}$$

Standard error of the slope:

$$s_{\hat{\beta}_1} = \sqrt{0.0453} \approx 0.213$$

EXERCISE 11.34 C): SOLUTION

✅ Answer:

c) 90% confidence interval for the slope

Degrees of freedom:

$$df = n - 2 = 6$$

Critical value:

$$t_{0.95,6} \approx 1.943$$

Confidence interval:

$$\hat{\beta}_1 \pm t \cdot s_{\hat{\beta}_1} = -0.347 \pm 1.943(0.213)$$

$$= -0.347 \pm 0.414$$

$$\boxed{(-0.761,\ 0.067)}$$

EXERCISE 11.35

11.35 A fast-food chain decided to carry out an experiment to assess the influence of advertising expenditure on sales. Different relative changes in advertising expenditure, compared to the previous year, were made in eight regions of the country, and resulting changes in sales levels were observed. The accompanying table shows the results.

Increase in advertising expenditure (%)	0	4	14	10	9	8	6	1
Increase in sales (%)	2.4	7.2	10.3	9.1	10.2	4.1	7.6	3.5

a. Estimate by least squares the linear regression of increase in sales on increase in advertising expenditure.

Newbold et al (2013)

EXERCISE 11.35 A): SOLUTION

✅ Answer:

(a) Least squares regression of sales on advertising

Step 1: Compute sample means

$$\bar{X} = \frac{0+4+14+10+9+8+6+1}{8} = \frac{52}{8} = 6.5$$

$$\bar{Y} = \frac{2.4+7.2+10.3+9.1+10.2+4.1+7.6+3.5}{8} = \frac{54.4}{8} = 6.8$$

Step 2: Compute sums of squares

$$S_{xx} = \sum (X_i - \bar{X})^2 = 146$$

$$S_{xy} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) = 80.6$$

Step 3: Estimate slope and intercept

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{80.6}{146} \approx 0.552$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 6.8 - 0.552(6.5) \approx 3.21$$

✅ Estimated regression line

$$\boxed{\hat{Y} = 3.21 + 0.552X}$$

Interpretation:

An increase of 1% in advertising expenditure is associated with an average increase of about **0.55% in sales**.

EXERCISE 11.35 B): SOLUTION

✅ Answer:

(b) 90% confidence interval for the slope

Step 1: Compute residual variance

$$S_{yy} = \sum (Y_i - \bar{Y})^2 = 64.64$$

$$\text{SSE} = S_{yy} - \hat{\beta}_1 S_{xy} = 64.64 - (0.552)(80.6) \approx 20.15$$

$$\text{MSE} = \frac{\text{SSE}}{n-2} = \frac{20.15}{6} \approx 3.36$$

Step 2: Standard error of the slope

$$\text{SE}(\hat{\beta}_1) = \sqrt{\frac{\text{MSE}}{S_{xx}}} = \sqrt{\frac{3.36}{146}} \approx 0.152$$

Step 3: Critical value

Degrees of freedom:

$$df = n - 2 = 6$$

For a **90% confidence interval**:

$$t_{0.05,6} \approx 1.943$$

EXERCISE 11.35 B): SOLUTION

✅ Answer:

Step 4: Confidence interval

$$\hat{\beta}_1 \pm t \cdot \text{SE} = 0.552 \pm 1.943(0.152)$$

$$= 0.552 \pm 0.295$$

✅ **90% confidence interval for the slope**

$$\boxed{(0.26,\ 0.85)}$$

THANKS!

Questions?