# STATISTICS II

Lisbon School of Economics & Management
Universidade de Lisboa

**Bachelor's degrees in Economics, Finance and Management**

2nd year/2nd Semester
2025/2026

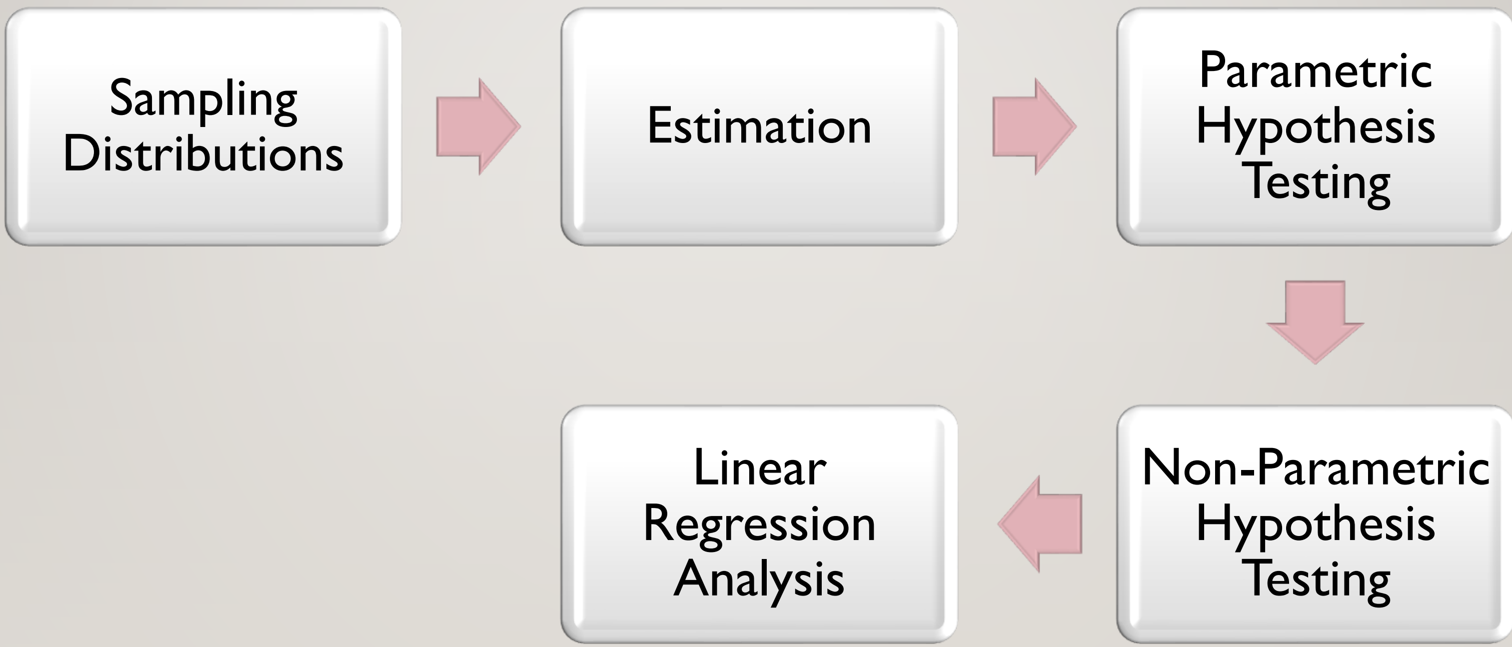
# CONTACT

**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 6

**Exercises 26 and 45**

**Exercises 8.12, 8.15, 8.16, 8.22**

# EXERCISE 26

At the beginning of a referendum campaign, the proportion of voters supporting "**YES**" in the population was 53%. At the end of the campaign, this proportion decreased to 49%.

Suppose that a random sample of size $n_1 = 100$ was taken at the beginning of the campaign and another **independent** random sample of size $n_2 = 120$ was taken at the end.

**Question:**

What is the probability that the **sample results indicate an increase in the support for "YES"**?

Murteira et al (2015), Chapter 6

# EXERCISE 26: SOLUTION

Answer:

**Step 1: Define the sample proportions**

Let

$\hat{p}_1$ = sample proportion of "YES" at the beginning, $\hat{p}_2$ = sample proportion of "YES" at the end.

Population parameters:

$$p_1 = 0.53, \quad p_2 = 0.49.$$

We are asked to compute:

$$P(\hat{p}_2 > \hat{p}_1) = P(\hat{p}_2 - \hat{p}_1 > 0).$$

# EXERCISE 26: SOLUTION

✅ Answer:

**Step 2: Sampling distribution of the difference**

For large samples, by the normal approximation:

$$\hat{p}_2 - \hat{p}_1 \mathrel{\dot{\sim}} N\left(p_2 - p_1,\ \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right).$$

Mean:

$$E(\hat{p}_2 - \hat{p}_1) = 0.49 - 0.53 = -0.04.$$

Variance:

$$\frac{0.53 \cdot 0.47}{100} = 0.002491, \quad \frac{0.49 \cdot 0.51}{120} = 0.002083.$$

$$\sigma = \sqrt{0.002491 + 0.002083} = \sqrt{0.004574} \approx 0.0676.$$

# EXERCISE 26: SOLUTION

✅ Answer:

**Step 3: Standardization**

$$Z = \frac{0 - (-0.04)}{0.0676} = \frac{0.04}{0.0676} \approx 0.592.$$

**Step 4: Compute the probability**

$$P(\hat{p}_2 > \hat{p}_1) = P(Z > 0.592).$$

From the standard normal table:

$$P(Z > 0.592) \approx 0.277.$$

**Final Answer**

$$\boxed{P(\hat{p}_2 > \hat{p}_1) \approx 0.28}$$

# EXERCISE 45

From **two independent normal populations with equal variances**, two random samples were taken with sizes $n_1 = 10$ and $n_2 = 5$, respectively.

**Question:** Determine the values between which, with 95% probability, the **ratio of the corrected (sample) variances** lies.

Murteira et al (2015), Chapter 6

# EXERCISE 45: SOLUTION

Answer:

Let $S_1^2$ and $S_2^2$ denote the **corrected sample variances** of the two samples.

Since the populations are normal and have **equal variances** $(\sigma_1^2 = \sigma_2^2)$, the ratio

$$\frac{S_1^2}{S_2^2}$$

follows an **F distribution**:

$$\frac{S_1^2}{S_2^2} \sim F_{n_1-1,\, n_2-1} = F_{9,4}.$$

**Step 1: Confidence interval for the ratio of variances**

We want numbers $a$ and $b$ such that

$$P\left(a \leq \frac{S_1^2}{S_2^2} \leq b\right) = 0.95.$$

This corresponds to the central 95% interval of the $F_{9,4}$ distribution:

$$P\left(F_{0.025;9,4} \leq \frac{S_1^2}{S_2^2} \leq F_{0.975;9,4}\right) = 0.95.$$

# EXERCISE 45: SOLUTION

✅ Answer:

**Step 2: Obtain critical values**

From F-distribution tables:

$$F_{0.025;9,4} \approx 0.104, \quad F_{0.975;9,4} \approx 9.12.$$

**Final Answer**

$$\boxed{0.104 \ \le \ \frac{S_1^2}{S_2^2} \ \le \ 9.12}$$

**Interpretation**

With **95% probability**, the ratio of the corrected sample variances lies between **0.104 and 9.12**. This wide interval reflects the **small sample sizes**, especially the second sample ($n_2 = 5$).

# EXERCISE 45: SOLUTION

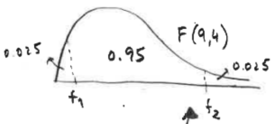
Answer:

**Alternative Solution:**

$X_1 \sim N(\mu_1, \sigma_1^2)$ → Amostra casual : $m = 10$

$X_2 \sim N(\mu_2, \sigma_2^2)$ → Amostra casual : $n = 5$

$\sigma_1^2 = \sigma_2^2$




Quer-se encontrar os valores $f_1$ e $f_2$, tais que: $P\left(f_1 \leq \frac{S_1'^2}{S_2'^2} \leq f_2\right) = 0.95$

Sabe-se que $F = \frac{S_1'^2}{S_2'^2} \cdot \underbrace{\frac{\sigma_2^2}{\sigma_1^2}}_{1} \sim F(m-1, n-1) = F(9,4)$ .. Logo,

$P\left(f_1 < \frac{S_1'^2}{S_2'^2} < f_2\right) = 0.95 \Rightarrow P\left(f_1 \times 1 < \frac{S_1'^2}{S_2'^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} < f_2 \times 1\right) = 0.95 \Rightarrow P(f_1 < F < f_2) = 0.95$

$\Rightarrow \begin{cases} P(F < f_1) = 0.025 \\ P(F > f_2) = 0.025 \end{cases} \Rightarrow \begin{cases} P\left(\frac{1}{F} > \frac{1}{f_1}\right) = 0.025 \text{, onde } \frac{1}{F} \sim F(4,9) \\ f_2 = 8.90 \end{cases} \Rightarrow$

$\Rightarrow \begin{cases} \frac{1}{f_1} = 4.72 \\ \text{"} \end{cases} \Rightarrow \begin{cases} f_1 = \frac{1}{4.72} \\ \text{"} \end{cases} \Rightarrow \begin{cases} f_1 = 0.212 \\ f_2 = 8.90 \end{cases}$

# EXERCISE 8.12

8.12 A manufacturer knows that the numbers of items produced per hour by machine A and by machine B are normally distributed with a standard deviation of 8.4 items for machine A and a standard deviation of 11.3 items for machine B. The mean hourly amount produced by machine A for a random sample of 40 hours was 130 units; the mean hourly amount produced by machine B for a random sample of 36 hours was 120 units. Find the 95% confidence interval for the difference in mean parts produced per hour by these two machines.

Newbold et al (2013)

# EXERCISE 8.12: SOLUTION

✅ Answer:

We are asked to find a **95% confidence interval** for the difference in mean hourly production:

$$\mu_A - \mu_B$$

The population standard deviations are **known**, and the populations are normal, so we use the **two-sample $z$-confidence interval**.

---

### Given data

- Machine A:
  $\sigma_A = 8.4, \quad n_A = 40, \quad \bar{x}_A = 130$
- Machine B:
  $\sigma_B = 11.3, \quad n_B = 36, \quad \bar{x}_B = 120$

# EXERCISE 8.12: SOLUTION

Answer:

**Step 1: Difference in sample means**

$$\bar{x}_A - \bar{x}_B = 130 - 120 = 10$$

**Step 2: Standard error**

$$SE = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{8.4^2}{40} + \frac{11.3^2}{36}}$$

$$SE = \sqrt{\frac{70.56}{40} + \frac{127.69}{36}} = \sqrt{1.764 + 3.547} = \sqrt{5.311} = 2.304$$

**Step 3: Critical value**

For a **95% confidence level**:

$$z_{0.975} = 1.96$$

# EXERCISE 8.12: SOLUTION

Answer:

### Step 4: Confidence interval

Margin of error:

$$ME = 1.96 \times 2.304 = 4.51$$

$$\text{CI} = 10 \pm 4.51$$

$$\boxed{(5.49,\ 14.51)}$$

---

### Final Answer

A **95% confidence interval** for the difference in mean parts produced per hour $(\mu_A - \mu_B)$ is:

$$\boxed{(5.49,\ 14.51)}$$

**Interpretation:**

We are 95% confident that machine A produces between **5.49 and 14.51 more items per hour** on average than machine B.

# EXERCISE 8.15

Explain your findings.

8.15 Recent business graduates currently employed in full-time positions were surveyed. Family backgrounds were self-classified as relatively high or low socioeconomic status. For a random sample of 16 high-socioeconomic-status recent business graduates, the mean total compensation was $34,500 and the sample standard deviation was $8,520. For an independent random sample of 9 low-socioeconomic-status recent business graduates, the mean total compensation was $31,499 and the sample standard deviation was $7,521. Find a 90% confidence interval for the difference between the two population means.

Newbold et al (2013)

# EXERCISE 8.15: SOLUTION

✅ Answer:

We are asked to find a **90% confidence interval** for the difference between the two population means:

$$\mu_H - \mu_L$$

where:

- **High socioeconomic status (H)**:
  $n_H = 16,\ \bar{x}_H = 34{,}500,\ s_H = 8{,}520$
- **Low socioeconomic status (L)**:
  $n_L = 9,\ \bar{x}_L = 31{,}499,\ s_L = 7{,}521$

The samples are **small**, the populations are assumed **normal**, and **equal variances are not stated**, so we use the **two-sample $t$-confidence interval with unequal variances (Welch's method)**.

# EXERCISE 8.15: SOLUTION

Answer:

**Step 1: Difference in sample means**

$$\bar{x}_H - \bar{x}_L = 34{,}500 - 31{,}499 = 3{,}001$$

**Step 2: Standard error**

$$SE = \sqrt{\frac{s_H^2}{n_H} + \frac{s_L^2}{n_L}} = \sqrt{\frac{8{,}520^2}{16} + \frac{7{,}521^2}{9}}$$

$$SE = \sqrt{4{,}534{,}800 + 6{,}286{,}269} = \sqrt{10{,}821{,}069} = 3{,}289$$

**Step 3: Degrees of freedom (Welch–Satterthwaite)**

$$df \approx \frac{\left(\frac{s_H^2}{n_H} + \frac{s_L^2}{n_L}\right)^2}{\frac{\left(\frac{s_H^2}{n_H}\right)^2}{n_H - 1} + \frac{\left(\frac{s_L^2}{n_L}\right)^2}{n_L - 1}} \approx 17$$

# EXERCISE 8.15: SOLUTION

Answer:

**Step 4: Critical value**

For a **90% confidence level** with $df \approx 17$:

$$t_{0.95,17} = 1.739$$

**Step 5: Confidence interval**

Margin of error:

$$ME = 1.739 \times 3{,}289 = 5{,}720$$

$$\text{CI} = 3{,}001 \pm 5{,}720$$

$$\boxed{(-2{,}719,\ 8{,}721)}$$

**Final Answer**

A **90% confidence interval** for the difference between the population mean total compensations $(\mu_{\text{high}} - \mu_{\text{low}})$ is:

$$\boxed{(-\$2{,}719,\ \$8{,}721)}$$

**Interpretation:**

At the 90% confidence level, the interval includes 0, so there is **no statistically significant difference** in mean total compensation between recent business graduates from high and low socioeconomic backgrounds.

# EXERCISE 8.16

8.16 Suppose that for a random sample of 200 firms that revalued their fixed assets, the mean ratio of debt to tangible assets was 0.517 and the sample standard deviation was 0.148. For an independent random sample of 400 firms that did not revalue their fixed assets, the mean ratio of debt to tangible assets was 0.489 and the sample standard deviation was 0.158. Find a 99% confidence interval for the difference between the two population means.

Newbold et al (2013)

# EXERCISE 8.16: SOLUTION

✅ Answer:

We are asked to find a **99% confidence interval** for the difference between two population means:

$$\mu_1 - \mu_2$$

where:

- Firms that **revalued** fixed assets:
  $n_1 = 200,\ \bar{x}_1 = 0.517,\ s_1 = 0.148$
- Firms that **did not revalue** fixed assets:
  $n_2 = 400,\ \bar{x}_2 = 0.489,\ s_2 = 0.158$

Since both sample sizes are large, we use the **large-sample two-sample $z$-confidence interval**, without assuming equal variances.

**Step 1: Difference in sample means**

$$\bar{x}_1 - \bar{x}_2 = 0.517 - 0.489 = 0.028$$

---

**Step 2: Standard error**

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{0.148^2}{200} + \frac{0.158^2}{400}}$$

$$SE = \sqrt{0.00010952 + 0.00006241} = \sqrt{0.00017193} = 0.01311$$

# EXERCISE 8.16: SOLUTION

Answer:

**Step 3: Critical value**

For a **99% confidence level**:

$$z_{0.995} = 2.576$$

**Step 4: Confidence interval**

Margin of error:

$$ME = 2.576 \times 0.01311 = 0.0338$$

$$\text{CI} = 0.028 \pm 0.0338$$

$$\boxed{(-0.0058,\ 0.0618)}$$

# EXERCISE 8.16: SOLUTION

✅ Answer:

**Final Answer**

A **99% confidence interval** for the difference between the population mean ratios ($\mu_{\text{revalued}} - \mu_{\text{not revalued}}$) is:

$$\boxed{(-0.0058,\ 0.0618)}$$

**Interpretation:**

At the 99% confidence level, the interval includes 0, so there is **no statistically significant difference** between the population mean debt-to-tangible-assets ratios of firms that revalued their fixed assets and those that did not.

# EXERCISE 8.22

8.22 Would you use the library more if the hours were extended? From a random sample of 138 freshmen, 80 indicated that they would use the school's library more if the hours were extended. In an independent random sample of 96 sophomores, 73 responded that they would use the library more if the hours were extended. Estimate the difference in proportion of first-year and second-year students responding affirmatively to this question. Use a 95% confidence level.

Newbold et al (2013)

# EXERCISE 8.22: SOLUTION

✅ Answer:

Let

$p_1$ = population proportion of **freshmen** who would use the library more

$p_2$ = population proportion of **sophomores** who would use the library more

From the samples:

$$\hat{p}_1 = \frac{80}{138} \approx 0.5797, \qquad \hat{p}_2 = \frac{73}{96} \approx 0.7604$$

The point estimate of the difference is:

$$\hat{p}_1 - \hat{p}_2 = 0.5797 - 0.7604 = -0.1807$$

Since both samples are sufficiently large, we use the **normal approximation**.

**Standard Error**

$$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{138} + \frac{\hat{p}_2(1-\hat{p}_2)}{96}}$$

$$SE \approx \sqrt{\frac{0.5797(0.4203)}{138} + \frac{0.7604(0.2396)}{96}} \approx \sqrt{0.00366} \approx 0.0605$$

# EXERCISE 8.22: SOLUTION

Answer:

**95% Confidence Interval**

For a 95% confidence level, $z_{0.975} = 1.96$.

Margin of error:

$$ME = 1.96 \times 0.0605 \approx 0.1186$$

Confidence interval:

$$(\hat{p}_1 - \hat{p}_2) \pm ME$$

$$(-0.1807 - 0.1186,\ -0.1807 + 0.1186)$$

$$\boxed{(-0.299,\ -0.062)}$$

**Interpretation**

We are 95% confident that the proportion of **freshmen** who would use the library more if hours were extended is **between 6.2% and 29.9% lower** than the proportion of **sophomores**.

# THANKS!

**Questions?**