

# 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



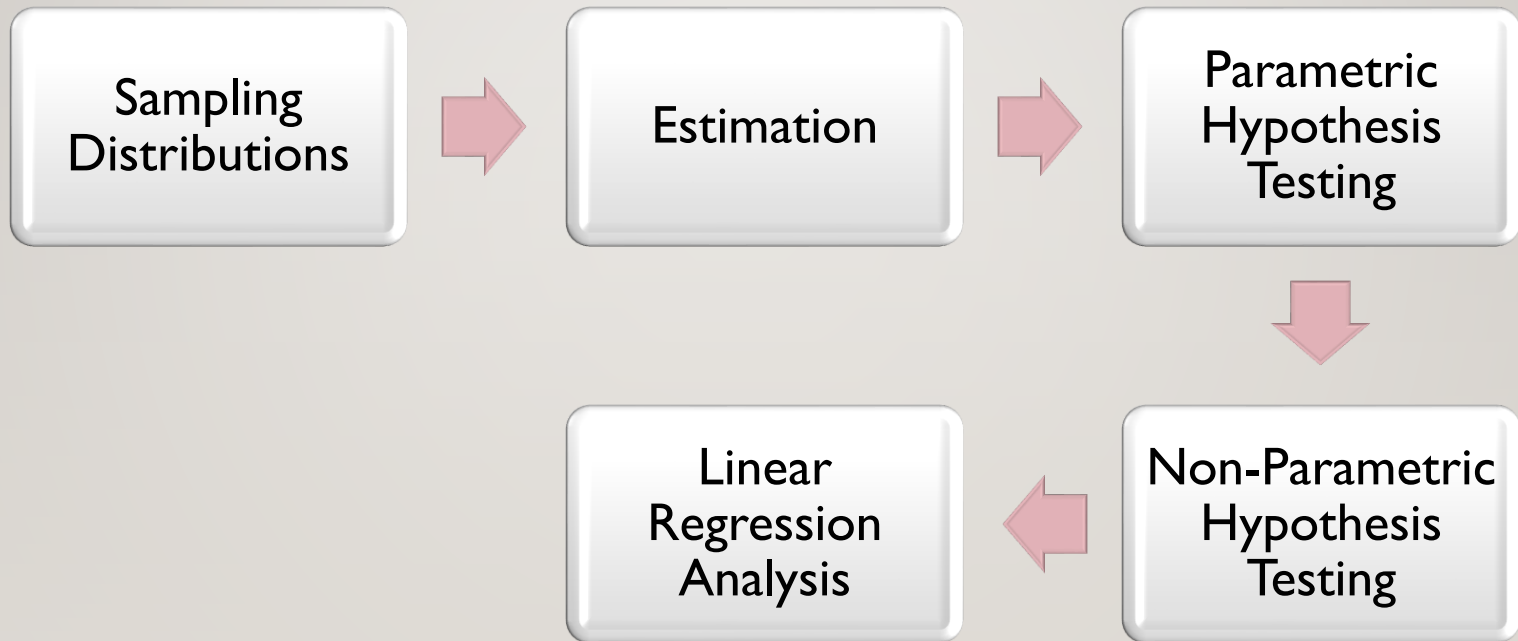
<https://doity.com.br/estatistica-aplicada-a-nutricao>



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

# PROGRAM

---



# PRATICAL CLASS 5

**Exercises 43 and 46/**

**Exercises 1, 2 and 3**

# EXERCISE 43

---

The fuel consumption (in liters per 100 km) of cars from brands **A** and **B** are independent random variables:

$$X_1 \sim N(\mu_1 = 1.8, \sigma_1 = 2.1), \quad X_2 \sim N(\mu_2 = 7.7, \sigma_2 = 5.1)$$

A potential buyer observes **two independent random samples** of  $n_1 = n_2 = 25$  cars from each brand, and chooses the car with the **lower sample mean**.

**Question:** What is the probability that the buyer chooses a car from brand A?

Murteira et al (2015), Chapter 6



# EXERCISE 43: SOLUTION



Answer:

Step 1: Define the difference of sample means

Let:

$$D = \bar{X}_2 - \bar{X}_1$$

- If  $D > 0$ , then  $\bar{X}_1 < \bar{X}_2 \rightarrow$  brand A has lower average consumption  $\rightarrow$  buyer chooses A

Since  $\bar{X}_1$  and  $\bar{X}_2$  are independent and normally distributed:

$$D \sim N(\mu_D, \sigma_D^2)$$

where

$$\mu_D = \mu_2 - \mu_1 = 7.7 - 1.8 = 5.9$$

$$\sigma_D = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{2.1^2}{25} + \frac{5.1^2}{25}}$$

Step 2: Compute standard deviation of the difference

$$\frac{\sigma_1^2}{n_1} = \frac{4.41}{25} = 0.1764, \quad \frac{\sigma_2^2}{n_2} = \frac{26.01}{25} \approx 1.0404$$

$$\sigma_D = \sqrt{0.1764 + 1.0404} = \sqrt{1.2168} \approx 1.103$$

# EXERCISE 43: SOLUTION

---



Answer:

Step 3: Compute probability

The buyer chooses brand A if  $\bar{X}_1 < \bar{X}_2$ , i.e.,

$$P(\text{buy brand A}) = P(D > 0)$$

Standardize to Z:

$$Z = \frac{0 - \mu_D}{\sigma_D} = \frac{0 - 5.9}{1.103} \approx -5.35$$

$$P(D > 0) = P(Z > -5.35) \approx 1$$

Answer

$$P(\text{buy brand A}) \approx 1 \text{ (or 100\%)}$$

Interpretation:

Even considering sample variability, the probability is essentially 100% that the buyer will choose brand A because its average consumption is much lower.

# EXERCISE 46 A)

---

Consider two populations with independent random variables  $X$  and  $Y$ , both **normally distributed**:

$$E(X) = 20, \quad E(Y) = 22, \quad \text{Var}(X) = \text{Var}(Y) = 16$$

Two independent random samples are taken:

- Sample from  $X$ :  $n_1 = 9$
- Sample from  $Y$ :  $n_2 = 16$

**Question:** What is the probability that the **sample mean of Y exceeds the sample mean of X by more than 3 units?**

Murteira et al (2015), Chapter 6



# EXERCISE 46 A): SOLUTION



Answer:

Step 1: Define the difference of sample means

Let

$$D = \bar{Y} - \bar{X}$$

- We are asked for  $P(D > 3)$ .

Since  $\bar{X}$  and  $\bar{Y}$  are independent and normally distributed:

$$D \sim N(\mu_D, \sigma_D^2)$$

where

$$\mu_D = E(\bar{Y}) - E(\bar{X}) = 22 - 20 = 2$$

$$\sigma_D = \sqrt{\frac{\text{Var}(X)}{n_1} + \frac{\text{Var}(Y)}{n_2}} = \sqrt{\frac{16}{9} + \frac{16}{16}}$$

# EXERCISE 46 A): SOLUTION



Answer:

Step 2: Compute the standard deviation

$$\frac{16}{9} \approx 1.7778, \quad \frac{16}{16} = 1$$
$$\sigma_D = \sqrt{1.7778 + 1} = \sqrt{2.7778} \approx 1.6667$$

Step 3: Standardize to Z

$$Z = \frac{3 - \mu_D}{\sigma_D} = \frac{3 - 2}{1.6667} \approx 0.60$$

Step 4: Compute probability

$$P(D > 3) = P(Z > 0.60)$$

From the standard normal table:

$$P(Z > 0.60) \approx 0.274$$

Answer

$$P(\bar{Y} - \bar{X} > 3) \approx 0.274 \text{ (or 27.4\%)}$$

Interpretation:

There is roughly a 27% chance that the sample mean of the second population exceeds the first by more than 3 units, given the sample sizes and population variances.

# EXERCISE I: UNKNOWN AND EQUAL VARIANCES

---

A nutritionist wants to compare the **average daily calorie intake** of two groups of students following different diets.

- **Group A:** 15 students, sample standard deviation  $S_A = 120$  kcal
- **Group B:** 18 students, sample standard deviation  $S_B = 110$  kcal

Assume that caloric intake in both groups follows a normal distribution and that the population variances are **unknown but equal**.

During the study, the observed difference in sample means is

$$\bar{X}_B - \bar{X}_A = 150 \text{ kcal.}$$

**Question:**

What is the probability of observing a difference in sample means **greater than 150 kcal**, assuming that the true difference in population means is zero?



# EXERCISE I: SOLUTION



Answer:

Given:

- Group A:  $n_A = 15$ ,  $S_A = 120$  kcal
- Group B:  $n_B = 18$ ,  $S_B = 110$  kcal
- Observed difference in sample means:  $\bar{Y} - \bar{X} = 150$  kcal
- Population variances unknown but equal, assume normality

Step 1: Compute pooled standard deviation

$$P(\bar{X}_B - \bar{X}_A > 150 \mid \mu_B - \mu_A = 0)$$

$$S_p^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}$$

$$S_p^2 = \frac{(15 - 1)120^2 + (18 - 1)110^2}{15 + 18 - 2} = \frac{14 \cdot 14400 + 17 \cdot 12100}{31}$$

$$14 \cdot 14400 = 201,600, \quad 17 \cdot 12100 = 205,700$$

$$S_p^2 = \frac{201,600 + 205,700}{31} = \frac{407,300}{31} \approx 13,145.16$$

$$S_p \approx \sqrt{13,145.16} \approx 114.7 \text{ kcal}$$

# EXERCISE I: SOLUTION



Answer:

Given:

- Group A:  $n_A = 15$ ,  $S_A = 120$  kcal
- Group B:  $n_B = 18$ ,  $S_B = 110$  kcal
- Observed difference in sample means:  $\bar{Y} - \bar{X} = 150$  kcal
- Population variances unknown but equal, assume normality

Step 2: Compute standard error

$$SE = S_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} = 114.7 \sqrt{\frac{1}{15} + \frac{1}{18}}$$

$$\frac{1}{15} + \frac{1}{18} = 0.0667 + 0.0556 = 0.1223$$

$$SE \approx 114.7 \cdot \sqrt{0.1223} = 114.7 \cdot 0.3497 \approx 40.1$$

Step 3: Compute t-value

$$t_{\text{obs}} = \frac{150}{40.1} \approx 3.74$$

$$t = \frac{\text{observed difference}}{SE}$$

$$\nu = n_A + n_B - 2 = 31$$

# EXERCISE I: SOLUTION

---



Answer:

Given:

- Group A:  $n_A = 15$ ,  $S_A = 120$  kcal
- Group B:  $n_B = 18$ ,  $S_B = 110$  kcal
- Observed difference in sample means:  $\bar{Y} - \bar{X} = 150$  kcal
- Population variances unknown but equal, assume normality

$$P(\bar{X}_B - \bar{X}_A > 150) = P(T_{31} > 3.74)$$

Consultando a t de Student (ou usando software):

$$P(T_{31} > 3.74) \approx 0.0004$$

# EXERCISE 2: UNKNOWN AND UNEQUAL VARIANCES

---

A quality control engineer wants to compare the **average processing time** (minutes) for two different machines.

- **Machine X:** 12 samples, sample standard deviation  $S_X = 4.5$  min
- **Machine Y:** 10 samples, sample standard deviation  $S_Y = 6.2$  min

The observed **difference in sample means** is 3 minutes (Machine Y – Machine X).

Assume normality but that the **variances are unknown and may be different**.

**Question:** What is the probability of observing a sample mean difference greater than 3, assuming that the true difference in population means is zero?



# EXERCISE 2: SOLUTION



Answer:

Given:

- Machine X:  $n_X = 12$ ,  $S_X = 4.5$  min
- Machine Y:  $n_Y = 10$ ,  $S_Y = 6.2$  min
- Observed difference in sample means:  $\bar{Y} - \bar{X} = 3$  min
- Population variances unknown and possibly unequal

Step 1: Compute standard error (Welch's formula)

$$P(\bar{Y} - \bar{X} > 3 \mid \mu_B - \mu_A = 0)$$

$$SE = \sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}} = \sqrt{\frac{4.5^2}{12} + \frac{6.2^2}{10}}$$

$$\frac{4.5^2}{12} = \frac{20.25}{12} \approx 1.6875, \quad \frac{6.2^2}{10} = \frac{38.44}{10} \approx 3.844$$

$$SE = \sqrt{1.6875 + 3.844} = \sqrt{5.5315} \approx 2.35 \text{ min}$$

Step 2: Compute t-value

$$t = \frac{\text{observed difference} - 0}{SE} = \frac{3}{2.35} \approx 1.277$$

# EXERCISE 2: SOLUTION



Answer:

Given:

- Machine X:  $n_X = 12$ ,  $S_X = 4.5$  min
- Machine Y:  $n_Y = 10$ ,  $S_Y = 6.2$  min
- Observed difference in sample means:  $\bar{Y} - \bar{X} = 3$  min
- Population variances unknown and possibly unequal

Step 3: Compute approximate degrees of freedom (Welch-Satterthwaite)

$$df = \frac{(S_X^2/n_X + S_Y^2/n_Y)^2}{\frac{(S_X^2/n_X)^2}{n_X-1} + \frac{(S_Y^2/n_Y)^2}{n_Y-1}}$$

$$S_X^2/n_X = 1.6875, \quad S_Y^2/n_Y = 3.844$$

$$\text{numerator} = (1.6875 + 3.844)^2 = 5.5315^2 \approx 30.61$$

$$\text{denominator} = \frac{1.6875^2}{11} + \frac{3.844^2}{9} = \frac{2.849}{11} + \frac{14.77}{9} \approx 0.259 + 1.641 \approx 1.90$$

$$df \approx \frac{30.61}{1.90} \approx 16.1 \approx 16 \text{ (rounded down)}$$

# EXERCISE 2: SOLUTION

---



Answer:

Given:

- Machine X:  $n_X = 12$ ,  $S_X = 4.5$  min
- Machine Y:  $n_Y = 10$ ,  $S_Y = 6.2$  min
- Observed difference in sample means:  $\bar{Y} - \bar{X} = 3$  min
- Population variances unknown and possibly unequal

$$P(\bar{Y} - \bar{X} > 3 \mid \mu_B - \mu_A = 0)$$



$$P(T_{16} > 1.277) \approx 0.11$$

# EXERCISE 3: KNOWN AND UNKNOWN VARIANCES

---

An exhaustive study was carried out in two poultry farms, **Nitrofrango** and **Frangosao**, and it was found that the average development time of a chicken was 38 days in the first farm and 42 days in the second farm. It is reasonable to assume that the distributions are normal.

A random sample of 25 chickens was collected from the Nitrofrango farm and 20 chickens from the Frangosao farm.

a) What is the probability that the *sample mean development time* is higher in the Nitrofrango farm than in the Frangosao farm, given that:

- i. The population standard deviations are known and equal to 6 and 5 days, respectively?
- ii. The observed sample standard deviations were 6 and 5 days, respectively? (Assume unequal population variances.)

b) Comment on the differences obtained in part (a).

ProbabilidadesEstatistica\_2019, Ex. 11 page 143



# EXERCISE 3 A) I: SOLUTION



Answer:

Let:

- Nitrofrango (1):  
 $\mu_1 = 38, n_1 = 25$
- Frangosao (2):  
 $\mu_2 = 42, n_2 = 20$

We want to compute:

$$P(\bar{X}_1 > \bar{X}_2) = P(\bar{X}_1 - \bar{X}_2 > 0)$$

The expected difference is:

$$\mu_1 - \mu_2 = 38 - 42 = -4$$

a.i) Population standard deviations known

Given:

- $\sigma_1 = 6$
- $\sigma_2 = 5$

The standard deviation of the difference of sample means is:

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{36}{25} + \frac{25}{20}} = \sqrt{1.44 + 1.25} = \sqrt{2.69} \approx 1.64$$

Standardizing:

$$Z = \frac{0 - (-4)}{1.64} \approx 2.44$$

Thus,

$$P(\bar{X}_1 > \bar{X}_2) = P(Z > 2.44) \approx 0.0073$$

✓ Probability  $\approx 0.73\%$

# EXERCISE 3 A) II: SOLUTION



Answer:

Let:

- Nitrofrango (1):  
 $\mu_1 = 38, n_1 = 25$
- Frangosao (2):  
 $\mu_2 = 42, n_2 = 20$

We want to compute:

$$P(\bar{X}_1 > \bar{X}_2) = P(\bar{X}_1 - \bar{X}_2 > 0)$$

The expected difference is:

$$\mu_1 - \mu_2 = 38 - 42 = -4$$

a.ii) Population standard deviations unknown (unequal variances)

Given sample standard deviations:

- $s_1 = 6$
- $s_2 = 5$

The estimated standard error is the same:

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{2.69} \approx 1.64$$

We now use a **t distribution** with Welch–Satterthwaite degrees of freedom:

$$\nu \approx \frac{(2.69)^2}{\frac{(1.44)^2}{24} + \frac{(1.25)^2}{19}} \approx 43$$

Test statistic:

$$t = \frac{4}{1.64} \approx 2.44$$

Thus,

$$P(\bar{X}_1 > \bar{X}_2) = P(T_{43} > 2.44) \approx 0.009$$

✓ Probability  $\approx 0.9\%$

# EXERCISE 3 B): SOLUTION

---



Answer:

## b) Comment on the results

In both cases, the probability that the sample mean development time is higher in **Nitrofrango** than in **Frangosao** is very small (below 1%).

- Knowing the population standard deviations leads to a slightly **smaller probability**, due to reduced uncertainty.
- When standard deviations are estimated from the samples, the probability increases marginally because of the additional variability introduced by estimation.

## ✦ Conclusion:

Regardless of whether the variances are known or estimated, it is **highly unlikely** that the average development time in Nitrofrango exceeds that of Frangosao, which is consistent with the fact that the true mean in Nitrofrango is 4 days lower.

# THANKS!

**Questions?**