STATISTICS II

Bachelor's degrees in Economics, Finance and Management

2nd year/2nd Semester
2025/2026

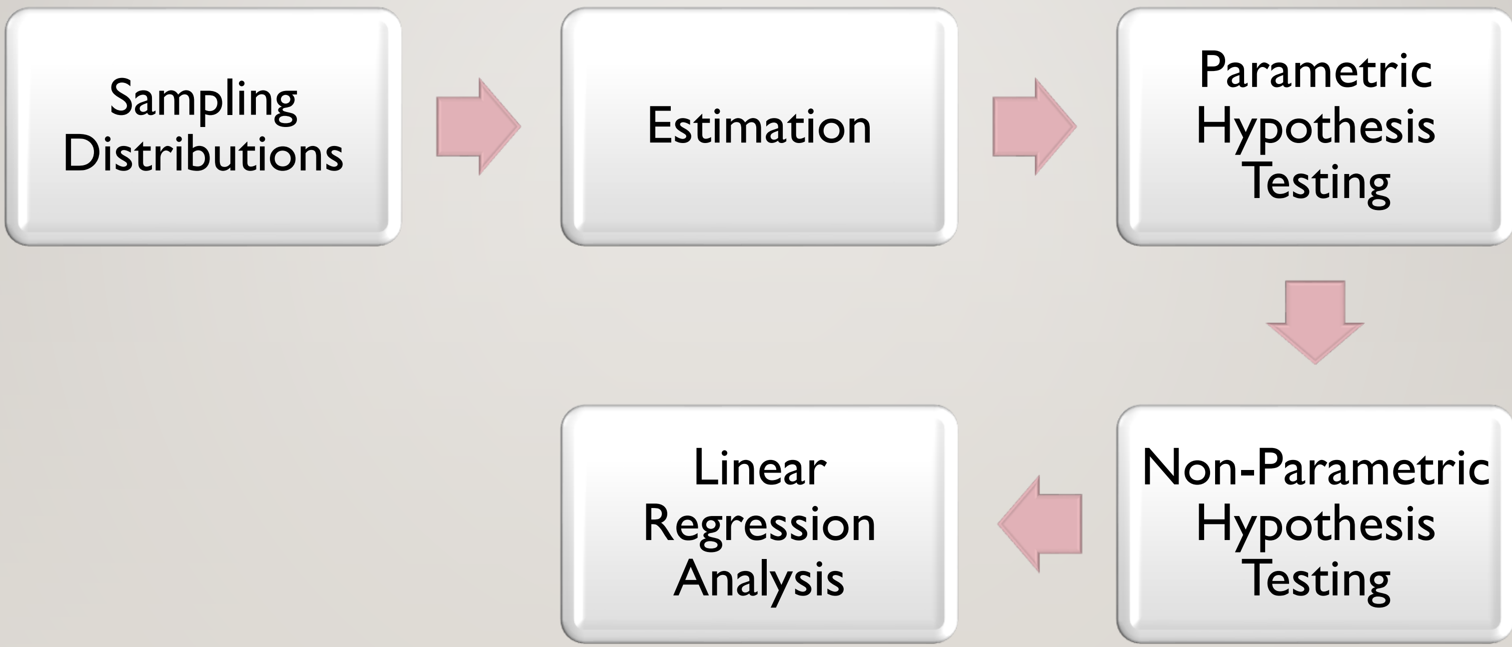
CONTACT

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

LECTURE 7 HOMEWORK: QUESTIONS AND SOLUTIONS

EXERCISE 7.70

7.70 The student government association at a university wants to estimate the percentage of the student body that supports a change being considered in the academic calendar of the university for the next academic year. How many students should be surveyed if a 90% confidence interval is desired and the margin of error is to be only 3%?

Newbold et al (2013)

EXERCISE 7.70: SOLUTION

Answer:

To determine the required sample size for estimating a **population proportion**, we use the formula

$$n = \left(\frac{z_{1-\alpha/2}}{\text{ME}}\right)^2 p(1-p)$$

Given information

- Confidence level: **90%** $\Rightarrow \alpha = 0.10$
- Margin of error: **ME = 0.03**
- Population proportion p: **unknown**

When p is unknown, we use the **conservative value**

$$p = 0.5$$

because it maximizes $p(1-p)$ and ensures a sufficiently large sample size.

For a 90% confidence level:

$$z_{1-\alpha/2} = z_{0.95} \approx 1.645$$

Sample size calculation

$$n = \left(\frac{1.645}{0.03}\right)^2 (0.5)(0.5)$$

$$n = (54.83)^2 \times 0.25 \approx 751.7$$

Final answer

$$\boxed{n = 752}$$

(The sample size is rounded **up** to ensure the desired margin of error.)

Interpretation

At least **752 students** should be surveyed to estimate the proportion of students who support the change with **90% confidence** and a **margin of error of 3%**.

EXERCISE 41

A security card issuing center has **two personalization machines**, operating independently. The processing time (in seconds) for each machine is normally distributed with the **same mean**, with a standard deviation of 10 seconds for the first machine and 15 seconds for the second.

Samples of 16 cards are taken from each machine.

Questions:

a) Calculate the probability that the **absolute difference between the sample means** of the two machines exceeds 5 seconds.

b) What is the probability that the **sample standard deviation** of the first machine is greater than that of the second machine?

Murteira et al (2015), Chapter 6

Note: Only complete part (a) of Exercise 41, as part (b) involves topics not yet covered.

EXERCISE 41 A): SOLUTION

Answer:

Given:

- Machine 1: $\sigma_1 = 10$, $n_1 = 16$
- Machine 2: $\sigma_2 = 15$, $n_2 = 16$
- Both populations are normal with the **same mean** μ.

a) Probability that $|\bar{X}_1 - \bar{X}_2| > 5$

1. **Distribution of the difference of means**

$$D = \bar{X}_1 - \bar{X}_2 \sim N\left(0, \sigma_D^2\right)$$

$$\sigma_D = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{10^2}{16} + \frac{15^2}{16}} = \sqrt{\frac{100}{16} + \frac{225}{16}} = \sqrt{\frac{325}{16}}$$

$$\sigma_D = \sqrt{20.3125} \approx 4.51$$

2. **Standardize for Z**

$$P(|\bar{X}_1 - \bar{X}_2| > 5) = P(D > 5) + P(D < -5) = 2 \cdot P(D > 5)$$

$$Z = \frac{5 - 0}{4.51} \approx 1.11$$

$$P(D > 5) = P(Z > 1.11) \approx 0.1335$$

$$\boxed{P(|\bar{X}_1 - \bar{X}_2| > 5) = 2 \cdot 0.1335 \approx 0.267}$$

✅ Answer (a): $\mathbf{P} \approx \mathbf{0.267}$ (~26.7%)

EXERCISE 41 A): SOLUTION

Alternative Solution:

Answer:

Given:

- Machine 1: $\sigma_1 = 10$, $n_1 = 16$
- Machine 2: $\sigma_2 = 15$, $n_2 = 16$
- Both populations are normal with the **same mean** μ.

$X_1 \sim N(\mu_1, 10^2) \rightarrow$ Amostra casual : $m = 16$

$X_2 \sim N(\mu_2, 15^2) \rightarrow$ Amostra casual : $n = 16$, onde $\mu_1 = \mu_2$

(a)

Quer-se : $P(|\bar{X}_1 - \bar{X}_2| > 5) = 1 - P(|\bar{X}_1 - \bar{X}_2| \leq 5) = 1 - P(-5 \leq \bar{X}_1 - \bar{X}_2 \leq 5)$

Sabe-se que : $Z = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}} \sim N(0,1)$. Logo,

EXERCISE 41 A): SOLUTION

Answer:

Alternative Solution:

$$1 - P(-5 \le \bar{X}_1 - \bar{X}_2 \le 5) = 1 - P\left(\frac{-5-0}{\sqrt{\frac{10^2}{16} + \frac{15^2}{16}}} \le \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \le \frac{5-0}{\sqrt{\frac{10^2}{16} + \frac{15^2}{16}}} \right) =$$

$$= 1 - P\left(\underbrace{-\frac{5}{\sqrt{20.3125}}}_{(-1.11)} \le Z \le \underbrace{\frac{5}{\sqrt{20.3125}}}_{(1.11)} \right) = 1 - \left[\Phi(1.11) - \Phi(-1.11) \right] =$$

$$= 1 - \left[2\Phi(1.11) - 1 \right] = 2 - 2\Phi(1.11) = 2 - 2 \times 0.8665 = \boxed{0.267}$$

LECTURE 8: SAMPLING DISTRIBUTION OF THE DIFFERENCE BETWEEN TWO SAMPLE PROPORTIONS

DIFFERENCE OF TWO SAMPLE PROPORTIONS

Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ be two independent random samples of sizes n_1 and n_2 drawn from **two Bernoulli populations** with parameters p_1 and p_2, respectively, where X_{ij} takes the value 1 for a success and 0 for a failure. Define the **sample proportions**:

$$\hat{p}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}, \quad \hat{p}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i}.$$

Standard Deviation

If the sample sizes are large, then by the **De Moivre–Laplace theorem**:

Note: For large samples ($n_1 > 25$ and $n_2 > 25$), the sample proportions $\hat{p}_1$ and $\hat{p}_2$ have an approximately **normal distribution**, by the Central Limit Theorem (CLT).

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

and by the **additivity property of the normal distribution**:

Note: The dot (or the letter *"a"*) above the distribution symbol denotes that the distribution is **approximately normal**.

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

Note: The normal distribution can be denoted as $N(\mu, \sigma^2)$ or $N(\mu, \sigma)$,depending on whether the second parameter represents the **variance** or the **standard deviation**.

that is,

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \dot{\sim} N(0,1).$$

EXERCISE 1: DIFFERENCE OF TWO SAMPLE PROPORTIONS

The proportion of customers who chose the Noko brand at the **TeleMN** store was 0.35, while at the **Optcel** store it was 0.29.

Question: If a sample of 200 customers is taken from TeleMN and a sample of 150 customers from Optcel, what is the probability that the **sample proportion of Noko customers at TeleMN exceeds the sample proportion at Optcel**?

ProbabilidadesEstatistica2019.pdf

EXERCISE 1: SOLUTION

✅ Answer:

Step 1: Define the sample proportions

Let

$$\hat{p}_1 = \text{sample proportion of Noko customers at TeleMN}, \quad n_1 = 200$$

$$\hat{p}_2 = \text{sample proportion of Noko customers at Optcel}, \quad n_2 = 150$$

We are asked to find

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

EXERCISE 1: SOLUTION

Answer:

Step 2: Compute the mean and standard deviation of the difference

The difference of sample proportions:

$$D = \hat{p}_1 - \hat{p}_2$$

$$E(D) = p_1 - p_2 = 0.35 - 0.29 = 0.06$$

$$\mathrm{Var}(D) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

$$\frac{p_1(1-p_1)}{n_1} = \frac{0.35 \cdot 0.65}{200} = \frac{0.2275}{200} \approx 0.0011375$$

$$\frac{p_2(1-p_2)}{n_2} = \frac{0.29 \cdot 0.71}{150} = \frac{0.2059}{150} \approx 0.0013727$$

$$\mathrm{Var}(D) = 0.0011375 + 0.0013727 \approx 0.0025102$$

$$\sigma_D = \sqrt{0.0025102} \approx 0.0501$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \mathrel{\dot\sim} N(0,1).$$

Standard Deviation of the Difference of Two Sample Proportions

EXERCISE 1: SOLUTION

Answer:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \overset{\cdot}{\sim} N(0,1).$$

Step 3: Standardize to Z

$$Z = \frac{0 - E(D)}{\sigma_D} = \frac{0 - 0.06}{0.0501} \approx -1.1976$$

Step 4: Compute probability

$$P(\hat{p}_1 - \hat{p}_2 > 0) = P(Z > -1.198)$$

From standard normal tables:

P(Z > -1.198) = 1- P(Z < -1.198) = 1- $\Phi(-1.198) = 1 - [1 - \Phi(1.198)]$
$= \Phi(1.198) \sim 0.883$

Standard Normal Distribution Table

Answer

$$P(\hat{p}_1 > \hat{p}_2) \approx 0.883 \ (or\ 88.3\%)$$

Interpretation:

There is approximately an **88% chance** that the sample proportion of Noko customers at TeleMN will be higher than at Optcel, given the sample sizes and population proportions.

EXERCISE 1: SOLUTION

Answer:

Alternative Solution:

$p_1 = 0{,}35$ e $p_2 = 0{,}29$.

$$\left.\begin{array}{l} X_1 \text{ e } X_2 \text{ dist. Bernoulli} \\ n_1\ (= 200) \text{ e } n_2\ (= 150) \text{ grandes} \end{array}\right| \Rightarrow Z = \frac{(\overline{P}_1 - \overline{P}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \overset{\circ}{\sim} N(0;1).$$

$$P(\overline{P}_1 > \overline{P}_2) = P(\overline{P}_1 - \overline{P}_2 > 0) = 1 - P(\overline{P}_1 - \overline{P}_2 \leq 0) = 1 - P\left(Z \leq \frac{0 - (0{,}35 - 0{,}29)}{\sqrt{\frac{0{,}35(1-0{,}35)}{200} + \frac{0{,}29(1-0{,}29)}{150}}}\right)$$

$$= 1 - \Phi(-1{,}2) = 1 - \big(1 - \Phi(1{,}2)\big) = \Phi(1{,}2) = 0{,}8849.$$

ProbabilidadesEstatistica2019.pdf

LECTURE 8: SAMPLING DISTRIBUTION OF THE RATIO OF TWO SAMPLE VARIANCES

RATIO OF TWO SAMPLE VARIANCES

Let

$$X_{11}, X_{12}, \ldots, X_{1n_1} \quad \text{and} \quad X_{21}, X_{22}, \ldots, X_{2n_2}$$

be **two independent random samples** of sizes n_1 and n_2, drawn from **two normal populations**, with

$$X_1 \sim N(\mu_1, \sigma_1^2) \quad \text{and} \quad X_2 \sim N(\mu_2, \sigma_2^2),$$

respectively.

Let the **corrected sample variances** be defined by

Note: The corrected sample variances are those computed by dividing by $n-1$, rather than by n.

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2, \qquad S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2.$$

Then, by a theorem concerning the **F distribution**, it follows that

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1,\, n_2-1}.$$

Note: The variable F has a Snedecor's F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

ProbabilidadesEstatistica2019.pdf

EXERCISE 1: RATIO OF TWO SAMPLE VARIANCES

A pharmaceutical company launched a new sleep medication that has been used in hospitals. It was observed that:

- Patients **not taking the medication** slept on average 7.5 hours with a standard deviation of 1.4 hours
- Patients **taking the medication** slept on average 8 hours with a standard deviation of 2 hours

At a particular hospital, **samples** were taken:

- $n_1 = 31$ patients not taking the medication
- $n_2 = 61$ patients taking the medication

Question: Assuming normality of the data, what is the probability that the **sample variance of the first group is smaller than that of the second group**?

ProbabilidadesEstatistica2019.pdf

EXERCISE 1: SOLUTION

Note: The F table provides **right-tail probabilities**, that is, values of the form $P(F > a)$.
Therefore, it is convenient to express the probability in this form.
In this case, the numerator degrees of freedom are 30 and the denominator degrees of freedom are 60.
The value in the table closest to 2.04 is 2.03, which corresponds to a right-tail probability of 0.01.
Hence, this value is used.

Answer:

Step 1: Define the random variable

Let S_1^2 and S_2^2 be the **sample variances** of the two groups.

For independent samples from normal populations, the ratio

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1,\, n_2-1}.$$

where F_{df_1,df_2} is the **F-distribution** with $df_1 = n_2 - 1$ and $df_2 = n_1 - 1$ degrees of freedom.

We are asked:

$$P(S_1^2 < S_2^2) = P\left(\frac{S_1^2}{S_2^2} < 1\right)$$

X_1 e X_2 dist. Normal
$n_1 = 31$ e $n_2 = 61$

$$\Rightarrow F = \frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1;\, n_2-1 = 30;60}$$

F Distribution Table

$$P(S_1^2 < S_2^2) = P\left(\frac{S_1^2}{S_2^2} < 1\right) = P\left(F < 1 \times \frac{2^2}{1.4^2}\right) = P(F < 2{,}04)$$ = 1 - P(F > 2.04) ~1-0.01 = 0.99

EXERCISE 1: SOLUTION

Answer:

Answer

$$\boxed{P(S_1^2 < S_2^2) \approx \text{ 0.99 (or 99\%)}}$$

Interpretation:

There is a **very high probability** that the sample variance of patients **not taking the medication** is smaller than that of patients **taking the medication**, which aligns with the population variances (1.96 vs 4).

LECTURE 8 HOMEWORK: QUESTIONS

EXERCISE 27

After an intensive advertising campaign, the market share of "Crispy Chips" increased from 8% to 10%. Suppose two independent surveys were conducted:

- **Before the campaign:** sample of size $n_1 = 100$
- **After the campaign:** sample of size $n_2 = 300$

Questions:

a) What is the probability that the surveys would conclude that the **market share gain exceeded 5 percentage points**?

b) What is the probability that the surveys would conclude a **loss in market share**?

Murteira et al (2015), Chapter 6

EXERCISE 41

A security card issuing center has **two independent personalization machines**. The processing time (in seconds) for each machine is **normally distributed** with the **same mean**, but different standard deviations:

- Machine 1: $\sigma_1 = 10$
- Machine 2: $\sigma_2 = 15$

Samples of **16 cards** are taken from each machine.

Questions:

a) What is the probability that the **absolute difference between the two sample means** exceeds 5 seconds?

b) What is the probability that the **sample standard deviation of the processing times for machine 1** is greater than that of machine 2?

Murteira et al (2015), Chapter 6

Note: Part (a) was solved earlier in a similar context and will not be repeated here.

THANKS!

Questions?