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 7

Exercises 8.23 and 1

Exercises 3, 4, 7 and 8

EXERCISE 8.23

8.23 A random sample of 100 men contained 61 in favor of a state constitutional amendment to retard the rate of growth of property taxes. An independent random sample of 100 women contained 54 in favor of this amendment. A confidence interval extending from 0.04 to 0.10 was calculated for the difference between the population proportions. Determine the confidence level of this interval.

Newbold et al (2013)

EXERCISE 8.23: SOLUTION

Answer:

Let

p_1 = population proportion of **men** in favor

p_2 = population proportion of **women** in favor

From the samples:

$$\hat{p}_1 = \frac{61}{100} = 0.61, \qquad \hat{p}_2 = \frac{54}{100} = 0.54$$

The point estimate of the difference is:

$$\hat{p}_1 - \hat{p}_2 = 0.61 - 0.54 = 0.07$$

The given confidence interval is:

$$(0.04,\ 0.10)$$

Step 1: Margin of error

The margin of error is half the length of the interval:

$$ME = \frac{0.10 - 0.04}{2} = 0.03$$

EXERCISE 8.23: SOLUTION

Answer:

Step 2: Standard error

For two population proportions:

$$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$SE = \sqrt{\frac{0.61(0.39)}{100} + \frac{0.54(0.46)}{100}} = \sqrt{0.002379 + 0.002484} = \sqrt{0.004863} \approx 0.0697$$

Step 3: Critical value

$$z = \frac{ME}{SE} = \frac{0.03}{0.0697} \approx 0.43$$

EXERCISE 8.23: SOLUTION

✅ Answer:

Step 4: Confidence level

A critical value of $z \approx 0.43$ corresponds to:

$$P(|Z| \leq 0.43) \approx 0.33$$

Final Answer

$$\boxed{\text{The confidence level is approximately } 33\%.}$$

Interpretation

This is a **very low confidence level**, which explains why the confidence interval is relatively narrow despite the moderate sample sizes.

EXERCISE 1: RATIO OF TWO POPULATION VARIANCES

To replace an old machine, there are two alternatives: **Equipment A** or **Equipment B**.

Since this is a decision involving considerable costs, as the equipment is quite expensive, both machines were tested during an experimental period.

At the end of the experimental period, **31** and **61** pieces were selected from the production of Equipment A and B, respectively. The following values were recorded for a characteristic used to evaluate the quality of the machines:

- Equipment A: $\sum_{i=1}^{31} x_{iA} = 43.4, \quad \sum_{i=1}^{31} x_{iA}^2 = 123.76$
- Equipment B: $\sum_{i=1}^{61} x_{iB} = 91.5, \quad \sum_{i=1}^{61} x_{iB}^2 = 269.25$

Using a **95% confidence interval**, determine if there is reason to believe that **Machine A achieves less variability** in the characteristic than Machine B. Assume normality of the distributions.

ProbabilidadesEstatistica_2019 (uevora.pt)

EXERCISE 1: SOLUTION

✅ Answer:

Step 1: Compute sample variances

The **sample variance formula** is:

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}$$

Equipment A:

$$s_A^2 = \frac{123.76 - \frac{(43.4)^2}{31}}{31-1} = \frac{123.76 - \frac{1883.56}{31}}{30} = \frac{123.76 - 60.77}{30} = \frac{62.99}{30} \approx 2.10$$

Equipment B:

$$s_B^2 = \frac{269.25 - \frac{(91.5)^2}{61}}{61-1} = \frac{269.25 - \frac{8372.25}{61}}{60} = \frac{269.25 - 137.24}{60} = \frac{132.01}{60} \approx 2.20$$

EXERCISE 1: SOLUTION

✅ Answer:

Step 2: Set up the confidence interval

For normal populations, the **pivotal quantity** is:

$$F = \frac{s_A^2}{s_B^2} \cdot \frac{\sigma_B^2}{\sigma_A^2} \sim F_{n_A-1, n_B-1}$$

A **95% confidence interval** for the ratio of population variances $\frac{\sigma_A^2}{\sigma_B^2}$ is:

$$\left(\frac{s_A^2}{s_B^2} \cdot \frac{1}{F_{1-\alpha/2; n_A-1, n_B-1}}, \frac{s_A^2}{s_B^2} \cdot \frac{1}{F_{\alpha/2; n_A-1, n_B-1}} \right)$$

Where:

- $n_A = 31 \Rightarrow df_1 = 30$
- $n_B = 61 \Rightarrow df_2 = 60$
- $\alpha = 0.05$

EXERCISE 1: SOLUTION

✅ Answer:

Step 3: Compute the ratio of sample variances

$$\frac{s_A^2}{s_B^2} = \frac{2.10}{2.20} \approx 0.955$$

Step 4: Find F quantiles

From $F_{30,60}$ tables (or approximate):

$$F_{0.975;30,60} \approx 1.78, \quad F_{0.025;30,60} \approx \frac{1}{1.78} \approx 0.562$$

Step 5: Compute confidence interval

$$\text{Lower bound} = \frac{0.955}{1.78} \approx 0.537$$

$$\text{Upper bound} = \frac{0.955}{0.562} \approx 1.70$$

Step 6: Interpretation

The **95% confidence interval** for $\frac{\sigma_A^2}{\sigma_B^2}$ is approximately:

$$\boxed{(0.54,\ 1.70)}$$

- Since the interval **includes 1**, there is **no evidence** that Machine A has **significantly lower variability** than Machine B at the 5% significance level.

✅ Both machines have similar variability in the characteristic evaluated.

EXERCISE 3

Consider a random variable X whose distribution depends on the parameters α and θ, such that

$$E(X) = \alpha\theta \quad \text{and} \quad \mathrm{Var}(X) = \alpha\theta^2.$$

From a random sample of size $n = 320$, the following quantities were obtained:

$$\sum_{i=1}^{320} x_i = 22.2, \qquad \sum_{i=1}^{320} x_i^2 = 535.8.$$

Provide, with justification, estimates for the unknown parameters.

Murteira (2015), Chapter 7

EXERCISE 3: SOLUTION

Answer:

V. A. com $F_X(x \mid \alpha, \theta)$

$E(X) = \alpha\theta \qquad Var(X) = \alpha\theta^2$

Amostra casual: $(X_1, \dots, X_{320})$

$$\sum_{i=1}^{320} x_i = 22.2$$

$$\sum_{i=1}^{320} x_i^2 = 535.8$$

Note: We will use the method of moments to estimate the parameters, since the likelihood function is not known. Therefore, the method of maximum likelihood cannot be applied at this stage and will be introduced later in the course.

EXERCISE 3: SOLUTION

Answer:

$$\mu_1' = E(X) = \alpha\theta$$

$$Var(X) = E(X^2) - E(X)^2 \Leftrightarrow E(X^2) = Var(X) + E(X)^2$$

$$\Leftrightarrow \mu_2' = Var(X) + E(X)^2$$

$$\begin{cases} \mu_1' = \dfrac{\sum_{i=1}^{320} x_i}{320} \\ \mu_2' = \dfrac{\sum_{i=1}^{320} x_i^2}{320} \end{cases} \Leftrightarrow \begin{cases} \alpha\theta = \bar{x} \\ \alpha\theta^2 + (\alpha\theta)^2 = \sum_{i=1}^{320} x_i^2/320 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} \alpha = \dfrac{\bar{x}}{\theta} \\ \dfrac{\bar{x}}{\theta}\theta^2 + \left(\dfrac{\bar{x}}{\theta}\right)^2\theta^2 = \sum_{i=1}^{320} x_i^2/320 \end{cases} \Leftrightarrow$$

EXERCISE 3: SOLUTION

Answer:

$$\Leftrightarrow \left\{ \bar{x}\theta + \bar{x}^2 = \sum_{i=1}^{320} x_i^2/320 \right. \Leftrightarrow$$

$$\Leftrightarrow \left\{ \bar{x}\theta = \sum_{i=1}^{320} x_i^2/320 - \bar{x}^2 \right. \Leftrightarrow \left\{ \theta = \frac{\sum_{i=1}^{320} x_i^2/320 - \bar{x}^2}{\bar{x}} \right. \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} \alpha = \dfrac{\bar{x}^2}{s^2} \\ \theta = \dfrac{s^2}{\bar{x}} \end{cases}$$

EXERCISE 3: SOLUTION

Answer:

Conclusion: :

$$\tilde{\alpha} = \left(\frac{22.2}{320}\right)^2 / \left[\frac{535.8}{320} - \left(\frac{22.2}{320}\right)^2\right] = 0.00288$$

$$\tilde{\theta} = \left[\frac{535.8}{320} - \left(\frac{22.2}{320}\right)^2\right] / \left(\frac{22.2}{320}\right) = 24.06576$$

EXERCISE 4

There are θ balls in a bag, numbered from 1 to θ. A sample of three balls was drawn **at random, with replacement**, and the following values were observed: 13, 5, and 9.

a) Compute an estimate of the number of balls in the bag using the **method of moments**.

b) Obtain the **maximum likelihood estimator** of θ.

c) Based on the estimates obtained in the previous parts, comment on the estimators that generated them.

Murteira (2015), Chapter 7

EXERCISE 4 A): SOLUTION

Answer:

θ bolas

Random Sample (X_1, X_2, X_3)

① ② ⋯

⋯ $(\theta-1)$ (θ)

(Amostra com reposição) $\longrightarrow$ $\begin{cases} x_1 = 13 \\ x_2 = 5 \\ x_3 = 9 \end{cases}$

$$f_X(x|\theta) = \frac{1}{\theta} \quad (x = 1, 2, \cdots, \theta)$$

$X \sim U(1, \theta)$ Discrete Uniform Distribution

$$E(X) = \frac{\theta+1}{2} = \mu_1'$$

a) $\mu_1' = \frac{\sum_{i=1}^{3} X_i}{3} \Leftrightarrow \frac{\theta+1}{2} = \bar{X} \Leftrightarrow \theta = 2\bar{X} - 1$ Estimator

Estimate: $\tilde{\theta} = 2\bar{x} - 1 = 2\left(\frac{13+5+9}{3}\right) - 1 = 17$ bolas

EXERCISE 4 A): SOLUTION

Answer:

Com a amostra observada o método dos momentos resulta numa estimativa admissível para θ uma vez que não se observou nenhuma bola com numero superior a 17. No entanto, não é garantido que isto se verifique em todas as amostras. Por exemplo:

$$(x_1, x_2, x_3) = (1,2,9) \Rightarrow \hat{\theta} = 2\,\frac{(1+2+9)}{3} - 1 = 7.$$

Com esta amostra o método dos momentos já não resultava numa estimativa admissível para θ uma vez que já se observou a bola nº 9, logo não podem existir apenas 7 bolas.

O método dos momentos é bom para estimar parâmetros referentes a momentos de variáveis aleatórias. Neste exercício, o parâmetro que estamos a estimar é o limite do suporte da variável aleatória (distribuição da população). A qualidade de um estimador não se deve avaliar olhando apenas para estimativas particulares (estejam estas muito próximas ou muito afastadas do verdadeiro valor do parâmetro). Interessa sim estudar o comportamento do estimador em sucessivas amostras, ou seja, estudar a sua distribuição por amostragem. Neste caso seria melhor usar outro estimador (o EMV).

EXERCISE 4 B): SOLUTION

Answer:

b) $L(\theta) = f_X(x_1, x_2, x_3 \mid \theta) = \prod_{i=1}^{3} f(x_i \mid \theta) = \frac{1}{\theta^3} \quad (\theta = x_{(3)}, x_{(3)}+1, x_{(3)}+2, \dots)$

$\Theta = \{x_{(3)}, x_{(3)}+1, x_{(3)}+2, \dots\} \rightarrow$ espaço-parâmetro discreto

$\rightarrow$ não existe derivada em ordem a θ

Como maximizar L(θ) em ordem a θ ?

$L(\theta) = \frac{1}{\theta^3} \quad (\theta = x_{(3)}, x_{(3)}+1, x_{(3)}+2, \dots)$

- $L(\theta)$ é função decrescente de θ
- Vamos escolher o menor valor $\theta \in \Theta$
- $L(\theta)$ é maximizada em $\theta = x_{(3)}$
- $\hat{\theta} = \arg\max_{\theta} \{L(\theta)\} = X_{(3)}$ (Máximo da amostra)

$\hookrightarrow$ Estimador da M.V.

EXERCISE 4 B): SOLUTION

✅ Answer:

Conclusão: A estimativa da máxima verosimilhança para θ é $\hat{\theta} = x_{(3)} = 13$

Comentário: O método da máxima verosimilhança escolhe o maior valor observado como estimativa o numero de bolas no saco. Isto significa que, no contexto desde problema, e independentemente da amostra observada, o método da máxima verosimilhança nunca estima valores não admissíveis para θ. Vimos em a) que tais estimativas absurdas podem acontecer com o método dos momentos.

EXERCISE 7

7. The time a student takes to answer a question on an exam is a random variable with an **exponential distribution** with parameter λ. In a random sample of 40 observations, a total of 480 minutes was recorded.

a) Obtain the **maximum likelihood estimator (MLE)** for λ.

b) Determine the **maximum likelihood estimate** of the proportion of questions answered in **less than 15 minutes**.

c) If an exam has eight questions, obtain an estimate of the **probability of answering all of them**, given that the total duration of the exam is 2 hours.

Murteira (2015), Chapter 7

EXERCISE 7 A): SOLUTION

$X \equiv$ Tempo que o aluno demora a responder a uma pergunta do exame (em minutos)

$X \sim ex(\lambda),\ \lambda > 0 \qquad f_X(x \mid \lambda) = \lambda e^{-\lambda x} \quad (x > 0)$

Amostra casual ($n = 40$): $(x_1, \dots, x_{40})$

$$\sum_{i=1}^{40} x_i = 480 \qquad \bar{x} = \frac{480}{40} = 12$$

a)

$$L(\lambda) = f_X(x_1, \dots, x_{40} \mid \lambda) = \prod_{i=1}^{40} f_X(x_i \mid \lambda) =$$

$$= \prod_{i=1}^{40} \lambda e^{-\lambda x_i} = \lambda^{40} e^{-\lambda \sum_{i=1}^{40} x_i}$$

$$\ell(\lambda) = \ln[L(\lambda)] = \ln\left(\lambda^{40} e^{-\lambda \sum_{i=1}^{40} x_i}\right) =$$

EXERCISE 7 A): SOLUTION

$$= 40 \ln(\lambda) - \lambda \sum_{i=1}^{40} x_i$$

$$\frac{d}{d\lambda} \ell(\lambda) = \frac{40}{\lambda} - \sum_{i=1}^{40} x_i = 0 \Leftrightarrow$$

$$\Leftrightarrow \frac{40}{\lambda} = \sum_{i=1}^{40} x_i \Leftrightarrow \hat{\lambda} = \frac{40}{\sum_{i=1}^{40} x_i} = \frac{1}{\bar{x}}$$

Condição de 2^{a} ordem:

$$\frac{d^2}{d\lambda^2} \ell(\hat{\lambda}) = \frac{d}{d\lambda}\left(\frac{40}{\lambda} - \sum_{i=1}^{40} x_i \right)\Bigg|_{\lambda = \frac{1}{\bar{x}}} = -\frac{40}{\lambda^2}\Bigg|_{\lambda = \frac{1}{\bar{x}}} < 0$$

Conclusão: $\hat{\lambda} = \frac{1}{\bar{X}}$ é o estimador da máxima verosimilhança para λ.

EXERCISE 7 B): SOLUTION

b) Percentagem = Probabilidade × 100

$$P(X<15) = \int_0^{15} f_X(x|\lambda)\,dx = \int_0^{15} \lambda e^{-\lambda x}\,dx =$$

$$= \left[-e^{-\lambda x}\right]_0^{15} = -e^{15\lambda} - (-e^{-0}) = 1 - e^{-15\lambda} =$$

$= t(\lambda) \rightarrow$ função estritamente crescente de λ

$t(\lambda)$ é função biunívoca de λ, logo podemos usar a propriedade de invariância do EMV:

$$\widehat{t(\lambda)} = t(\hat{\lambda}) = 1 - e^{-15\hat{\lambda}} = 1 - e^{-\frac{15}{12}} = 0.7135$$

$$\hat{\lambda} = \frac{1}{\bar{x}} = \frac{1}{12} \qquad \text{exame} = 2\text{ h} = 120\text{ min}$$

EXERCISE 7 C): SOLUTION

c) $X \equiv$ Tempo 1 pergunta (min.) $\sim ex(\lambda)$

$S = \sum_{i=1}^{8} X_i \equiv$ Tempo 8 perguntas (min) $\sim G(8,\lambda)$

$X \sim G(n;\lambda) \Leftrightarrow 2\lambda X \sim \chi^2(2n)$ $\qquad Q = 2\lambda S \sim \chi^2(16)$

$$P(S \leqslant 120) = P(2\lambda S \leqslant 2 \times \lambda \times 120) =$$
$$= P(Q \leqslant 240\lambda)$$
$$= t(\lambda) \rightarrow$$ função estritamente crescente de λ em $\lambda > 0$

Assim sendo podemos usar a propriedade de invariância do EMV:

$$\widehat{P(S \leqslant 120)} = P(\widehat{Q \leqslant 240\lambda}) = P(Q \leqslant 240\hat{\lambda}) =$$
$$= P\left(Q \leqslant \frac{240}{12}\right) = P(Q \leqslant 20) \rightarrow$$ não está na tabela
$$= 0.78 \leftarrow$$

```
> pchisq(20 , 16) |> round(2)
[1] 0.78
```

A1 f_x =ARRED(DIST.CHIQ(20;16;VERDADEIRO); 2)

	A	B	C	D	E	F	G
1	0,78						
2							

Usando só as tabelas, o melhor que se podia fazer era usar o valor mais próximo de 20 na tabela da $\chi^2(16)$:

$$P(Q \leqslant 20) = 1 - P(Q > 20) \approx 1 - P(Q > 19.369) = 1 - 0.25 =$$
$$= 0.75$$

EXERCISE 8

8. It is assumed that the repair time of a certain type of machine, X, follows a **normal distribution** with unknown parameters. To estimate these parameters, a random sample of repair times (in minutes) was collected. The data are as follows:

$$n = 10, \quad \frac{1}{10}\sum_{i=1}^{10} x_i = 846, \quad \frac{1}{10}\sum_{i=1}^{10} x_i^2 = 71607.$$

Estimate the **probability that the repair time of a machine is less than 83 minutes**.

Murteira (2015), Chapter 7

EXERCISE 8: SOLUTION

$X \equiv$ Tempo de reparação (em minutos) $\sim N(\mu, \sigma^2)$

$$\frac{X-\mu}{\sigma} \sim N(0,1)$$

Amostra casual: $(x_1, \cdots, x_{10})$ $\quad n = 10$

$$\sum_{i=1}^{10} x_i = 846$$

$$\sum_{i=1}^{10} x_i^2 = 71607$$

$$f_X(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$$L(\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Nota: Já determinados os EMV no exemplo anterior, denotado por exercício 2.

EXERCISE 8: SOLUTION

$$\ell(\mu,\sigma) = \ln L(\mu,\sigma^2) = -\frac{n}{2}\ln\ 2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 =$$

$$= -\frac{n}{2}\left[\ln(2\pi) + \ln(\sigma^2)\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 =$$

$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2, \quad (\mu,\sigma^2)\in\Theta,$$

$$\Theta = \{(\mu,\sigma^2):\ \mu\in\mathbb{R} \wedge \sigma^2\in\mathbb{R}^+\}.$$

EXERCISE 8: SOLUTION

$$\begin{cases} \dfrac{\partial}{\partial \mu} \ell(\mu, \sigma^2) = -\dfrac{1}{2\sigma^2} \displaystyle\sum_{i=1}^{n} 2(x_i - \mu)(-1) = \dfrac{\sum_{i=1}^{n} (x_i - \mu)}{\sigma^2} = 0 \\ \dfrac{\partial}{\partial \sigma^2} \ell(\mu, \sigma^2) = -\dfrac{n}{2\sigma^2} + \dfrac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\sigma^4} = 0 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} \displaystyle\sum_{i=1}^{n} (x_i - \mu) = 0 \\ \dfrac{n}{2\sigma^2} = \dfrac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\sigma^4} \end{cases} \Leftrightarrow \begin{cases} \displaystyle\sum_{i=1}^{n} x_i = n\mu \\ \sigma^2 = \dfrac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} \mu = \dfrac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \\ \sigma^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = s^2 \end{cases}$$

EXERCISE 8: SOLUTION

$$P(X < 83) = \Phi\left(\frac{83-\mu}{\sigma}\right) = \Phi\left(\frac{83-\mu}{\sqrt{\sigma^2}}\right) = t(\mu, \sigma^2)$$

A função $t(\mu, \sigma^2)$ é biunívoca (separadamente) em relação a μ e a σ^2. Assim sendo, podemos usar a propriedade de invariância do EMV:

$$\widehat{t(\mu, \sigma^2)} = t(\hat{\mu}, \hat{\sigma}^2)$$

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{846}{10} = 84.6$$

$$\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^{10} x_i^2}{n} - \bar{x}^2 = \frac{71607}{10} - 84.6^2 = 3.54$$

EXERCISE 8: SOLUTION

$$\underline{\text{então}}: \widehat{P(X<83)} = \Phi\left(\frac{83-\hat{\mu}}{\sqrt{\hat{\sigma}^2}}\right) = \Phi\left(\frac{83-\bar{x}}{\sqrt{s^2}}\right) =$$

$$= \Phi\left(\frac{83-84.6}{\sqrt{3.54}}\right) \simeq \Phi(-0.85)$$

$$= 1 - \Phi(0.85) \quad = \underset{\substack{\uparrow \\ \text{Tabela 4}}}{1} - 0.8023$$

$$= 0.1977$$

THANKS!

Questions?