STATISTICS II

Bachelor's degrees in Economics, Finance and Management

2nd year/2nd Semester
2025/2026

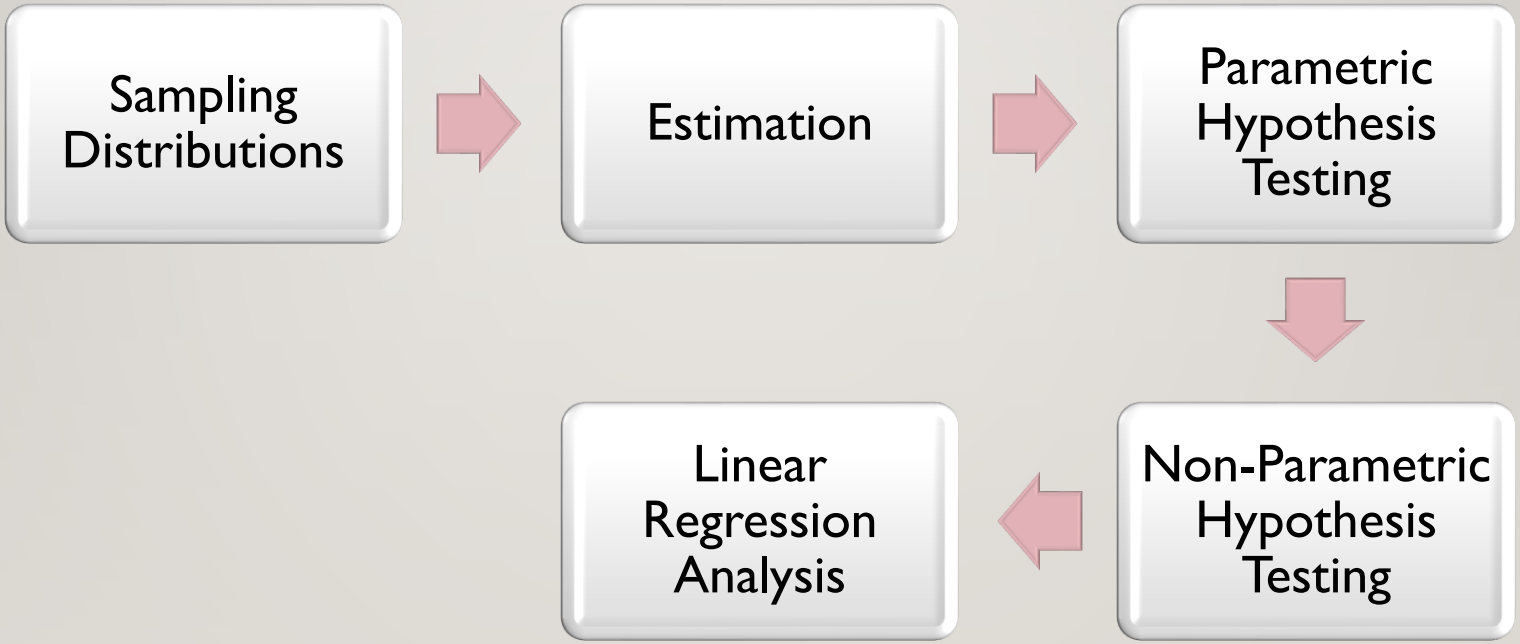
CONTACT

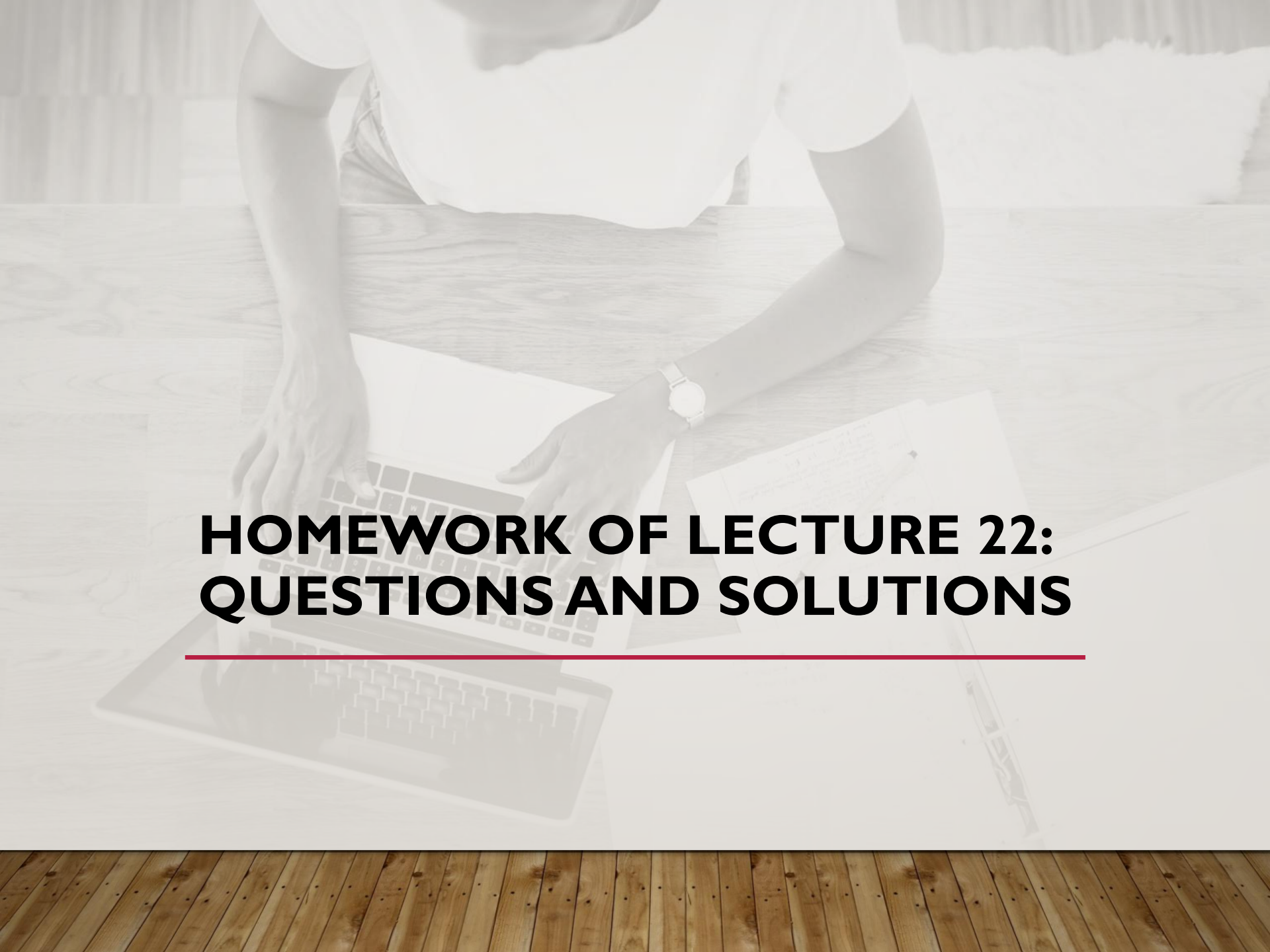
Professor: Elisabete Fernandes
E-mail: efernandes@iseg.ulisboa.pt

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



HOMEWORK OF LECTURE 22: QUESTIONS AND SOLUTIONS

EXERCISE 11.32 A)

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.

a. A random sample of size $n = 38$ with
$b_1 = 5 \quad s_{b_1} = 2.1$

Newbold et al (2013)

EXERCISE 11.32 A): SOLUTION

✅ Answer:

Given

- Sample size: $n = 38$
- Estimated slope: $\hat{b}_1 = 5$
- Standard error of slope: $s_{b_1} = 2.1$
- Significance level: $\alpha = 0.05$

Degrees of freedom:

$$df = n - 2 = 36$$

1) Hypothesis test for the slope

Hypotheses

$$H_0 : \beta_1 \leq 0$$

$$H_1 : \beta_1 > 0$$

Test statistic

$$t = \frac{\hat{b}_1 - 0}{s_{b_1}} = \frac{5}{2.1} \approx 2.38$$

EXERCISE 11.32 A): SOLUTION

✅ Answer:

Critical value

For a **one-sided test** at $\alpha = 0.05$ with $df = 36$:

$$t_{0.95,36} \approx 1.69$$

Decision

$$2.38 > 1.69$$

👉 Reject H_0

RR = [1.69; +∞[

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 A): SOLUTION

✅ Answer:

2) Two-sided 95% confidence interval for β_1

Critical value:

$$t_{0.975,36} \approx 2.03$$

$$\hat{b}_1 \pm t \cdot s_{b_1} = 5 \pm 2.03(2.1) = 5 \pm 4.26$$

$$\boxed{(0.74,\ 9.26)}$$

3) Two-sided 99% confidence interval for β_1

Critical value:

$$t_{0.995,36} \approx 2.72$$

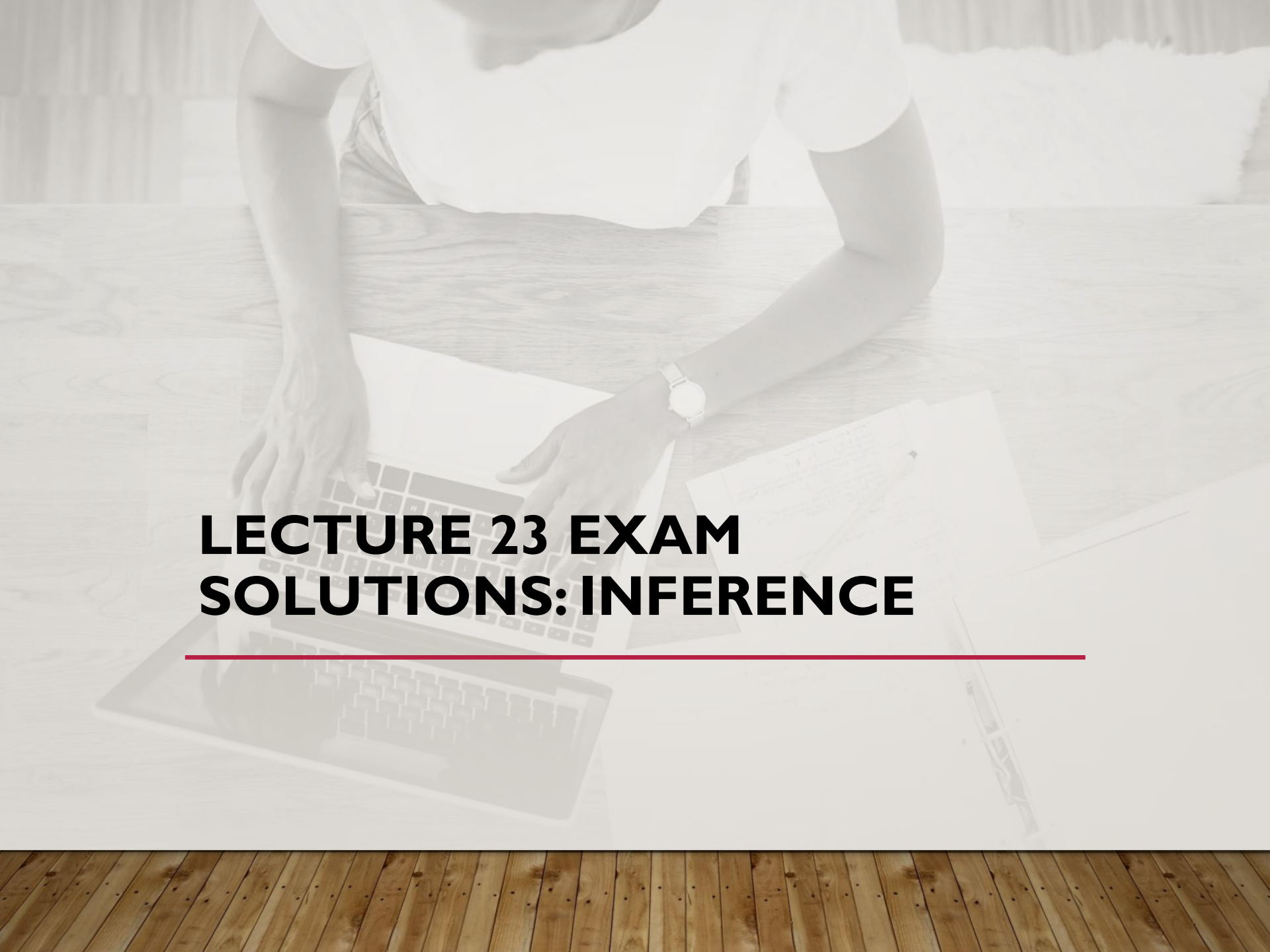
$$5 \pm 2.72(2.1) = 5 \pm 5.71$$

$$\boxed{(-0.71,\ 10.71)}$$

✅ **Final summary**

- The slope is **significantly greater than zero** at the 5% level.
- 95% CI: $(0.74,\ 9.26)$ → excludes 0
- 99% CI: $(-0.71,\ 10.71)$ → includes 0

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



LECTURE 23 EXAM SOLUTIONS: INFERENCE

EXERCISE I

Exercise | · Correlation vs Causation

A study finds a positive correlation between **number of cafés in a city** and **number of tourists**.

Which statement is correct?

a) Cafés cause tourism

b) Tourism causes cafés

c) The variables are correlated, but causation is not implied

d) There is no relationship between the variables

EXERCISE 3: SOLUTION

✅ Answer:

✅ **Exercise | - Correlation vs Causation**

Correct answer: c) The variables are correlated, but causation is not implied

Justification:

Correlation only measures association.

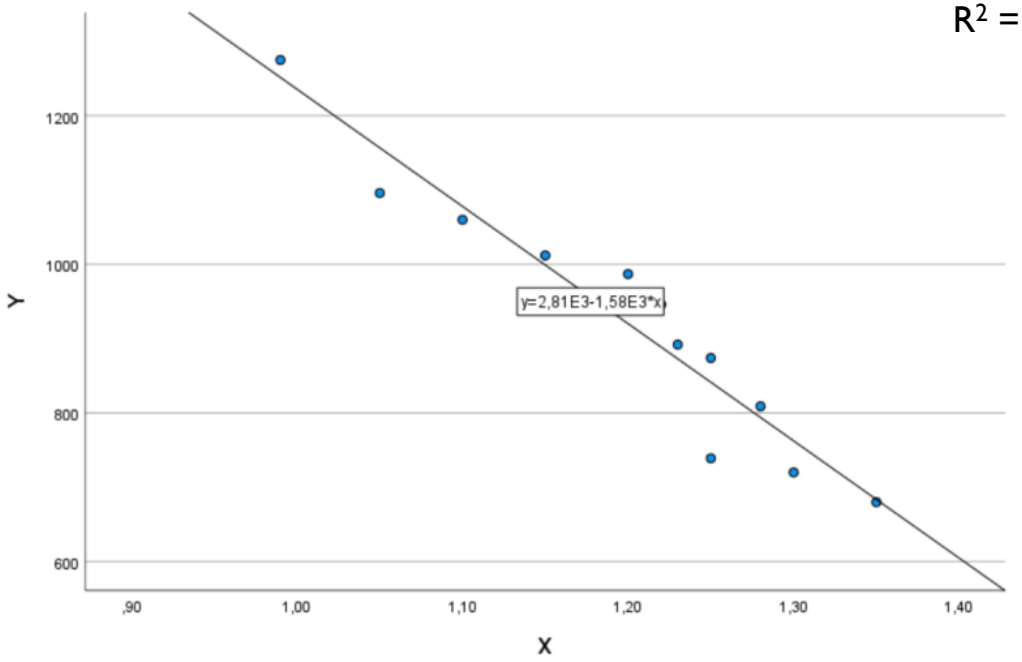
It does **not** prove that one variable causes the other.

EXERCISES 2 TO 3

2 to 3. It is believed that the number of packages sold of a certain generic drug (Y) depends on its price (X, in euros). For this purpose, the values of these variables were observed over 12 weeks, and the following results were obtained:

y	892	1012	1060	987	680	739	809	1275	946	874	720	1096
x	1,23	1,15	1,10	1,20	1,35	1,25	1,28	0,99	1,22	1,25	1,30	1,05

Later, a researcher analysed these data and obtained the following graph:



EXERCISE 2

2. It can be concluded that there is a linear correlation:

a) Moderate positive correlation between X and Y.

b) Strong positive correlation between X and Y.

c) Moderate negative correlation between X and Y.

d) Strong negative correlation between X and Y.

e) Perfect negative correlation between X and Y.

EXERCISE 2: SOLUTION

✅ Answer:

Question 2

Correct answer: d) Strong negative correlation between X and Y

$R^2 = 0.921$
Then R = -0.960

Justification:
From the scatterplot (price vs. number of packages sold), as price increases, sales decrease in a strong linear pattern, indicating a strong negative linear correlation.

EXERCISE 3

3. Based on the results presented, it can be stated that:

a) 92.10% of the variability of Y is explained by X.

b) 92.10% of the variability of X is explained by Y.

c) 96.0% of the variability of Y is explained by X.

d) 96.0% of the variability of X is explained by Y.

e) The variability of Y is not explained by X.

EXERCISE 3: SOLUTION

✅ Answer:

Question 3

Correct answer: a) 92.10% of the variability of Y is explained by X

Justification:
The coefficient of determination is $R^2 = 0.9210$, meaning 92.10% of the variability of the dependent variable $Y$is explained by X.

EXERCISE 4

Exercise 4 – t-Test (When to Use)

When should a **t-test for the mean** be used instead of a z-test?

a) When the population variance is known

b) When the sample size is very large

c) When the population variance is unknown and the sample is small

d) When data are categorical

EXERCISE 4: SOLUTION

✅ Answer:

✅ **Exercise 4 – t-Test vs z-Test**

Correct answer: c) When the population variance is unknown and the sample is small

Justification:

The t-test is used when σ is unknown and the sample size is small.

EXERCISE 5

5. In a statistical test, it is known that:

a) The rejection region does not depend on the significance level.

b) The p-value does not depend on the significance level.

c) The probability of a Type I error is equal to the power of the test.

d) The probability of a Type II error is equal to the power of the test.

e) If the p-value is greater than the significance level (α), then the null hypothesis is rejected at α.

EXERCISE 5: SOLUTION

✅ Answer:

Question 5

Correct answer: b) The p-value does not depend on the significance level

Justification:
The p-value is computed from the sample data. The significance level $\alpha$is chosen by the researcher independently.

EXERCISE 6

6. A sample of 30 students was analysed for hours of study and exam scores. The Pearson correlation coefficient was r = 0.65. Which statement is correct?

a) Approximately 42% of the variance in exam scores can be explained by hours of study.

b) The correlation implies causation.

c) The coefficient of determination is 0.35.

d) There is a strong negative linear relationship.

e) The mean of exam scores equals the mean of hours studied.

EXERCISE 6: SOLUTION

✅ Answer:

Question 6

Correct answer: a) Approximately 42% of the variance in exam scores can be explained by hours of study

Justification:

$$R^2 = r^2 = (0.65)^2 = 0.4225$$

Thus, about 42% of the variability in exam scores is explained by hours studied.

EXERCISE 7

7. A company reports that the population mean delivery time is 12 days, with population standard deviation 3 days. A random sample of 36 deliveries is taken. The sampling distribution of the mean is approximately:

a) $N(12, 5)$

b) $N(12,3)$.

c) $t(35)$.

d) $N(12,0.25)$.

e) $N(12,1)$.

EXERCISE 7: SOLUTION

✅ Answer:

Question 7

Correct answer: d) $N(12,0.25)$.

Justification:

$$\sigma_{\bar{X}} = \frac{3}{\sqrt{36}} = 0.5$$

EXERCISE 8

8. A random sample of 100 voters found that 62 would vote for candidate A. Which is the 95% confidence interval for the population proportion?

a) (0.518; 0.722)

b) (0.50; 0.74)

c) (0.52; 0.72)

d) (0.48; 0.76)

e) (0.55; 0.70)

EXERCISE 8: SOLUTION

✅ Answer:

Question 8

Correct answer: c) (0.52; 0.72)

Justification:

$$\hat{p} = 0.62, SE = \sqrt{\frac{0.62(0.38)}{100}} \approx 0.048$$

$$CI = 0.62 \pm 1.96(0.048) \approx (0.52, 0.72)$$

EXERCISE 9

9. A researcher wants to estimate the proportion of students who approve a new teaching method. He wants a 95% confidence interval and estimates the proportion to be 0.6. What is the approximate margin of error for a sample of 100 students?

a) 0.098

b) 0.096

c) 0.092

d) 0.085

e) 0.080

EXERCISE 9: SOLUTION

✅ Answer:

Question 9

Correct answer: b) 0.096

Justification:

$$ME = 1.96\sqrt{\frac{0.6(0.4)}{100}} \approx 0.096$$

EXERCISE 10

10. A dataset of 100 transactions has observed frequencies:

Pagment Type	Cash	Card	Mobile
Observed Frequency	25	60	15

Historical proportions are 0.2, 0.6 and 0.2, respectively. Which statement is correct at the significance level $\alpha = 0.05$?

a) The observed frequencies are consistent with historical proportions.

b) The observed frequencies are inconsistent with historical proportions.

c) Only Cash proportion is consistent.

d) Only Card proportion is inconsistent.

e) Cannot conclude without more data.

EXERCISE 10: SOLUTION

✅ Answer:

Question 10

Correct answer: a) The observed frequencies are consistent with historical proportions

Justification:

Expected frequencies: $(20, 60, 20)$.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{25-20)^2}{20} + 0 + \frac{(15-20)^2}{20} = 2.5$$

With 2 degrees of freedom, $2.5 < 5.99$. Fail to reject H_0.

EXERCISE 11

11. A random sample of 20 daily temperatures has a sample mean of 22°C and a sample standard deviation of 3°C.
Test whether the mean temperature differs from 20°C at a significance level of α = 0.05.
Use the t-statistic formula:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

The critical value for a two-tailed test with 19 degrees of freedom is $t_{0.025,19} = -2.093$.
Question: Based on this information, which conclusion is correct?

a) Reject H0; mean temperature is significantly different from 20°C.
b) Fail to reject H0; no significant difference.
c) Accept H1; mean equals 22°C.
d) Reject H1; mean equals 20°C.
e) Cannot perform test; population variance unknown.

EXERCISE 11: SOLUTION

✅ Answer:

Question 11

Correct answer: a) Reject H_0; mean temperature is significantly different from 20°C

Justification:

$$t = \frac{22 - 20}{3/\sqrt{20}} \approx 2.98$$

Since $| t | > 2.093$, reject H_0.

EXERCISE 12

Exercise 12 Chi-Square Test (Purpose)

The chi-square goodness-of-fit test is mainly used to:

a) Compare two population means

b) Test if observed frequencies match expected proportions

c) Test correlation between two variables

d) Estimate a confidence interval

EXERCISE 12: SOLUTION

✅ Answer:

✅ **Exercise 12 Chi-Square Test**

Correct answer: b) Test if observed frequencies match expected proportions

Justification:

The chi-square goodness-of-fit test compares **observed** and **expected** frequencies.

EXERCISE 13

13. The 90% confidence interval for the average exam score of students in Class A is (34.18; 56.35). This means that:

a) There is a 90% probability that the true mean exam score is between 34.18 and 56.35.

b) A point estimate of the mean exam score is 35.

c) With 90% confidence, the true mean exam score is between 34.18 and 56.35.

d) There is a 10% probability that the true mean exam score is below 34.18.

e) There is a 5% probability that the true mean exam score is above 56.35.

EXERCISE 13: SOLUTION

✅ Answer:

Question 13

Correct answer: c) With 90% confidence, the true mean exam score is between 34.18 and 56.35

Justification:
This is the correct frequentist interpretation of a confidence interval.

EXERCISE 14

Exercise 14 Hypothesis Testing (Intuition)

In hypothesis testing, a **small p-value** indicates:

a) Strong evidence in favour of the null hypothesis

b) Strong evidence against the null hypothesis

c) The null hypothesis is always true

d) The significance level is large

EXERCISE 14: SOLUTION

✅ Answer:

✅ **Exercise 14 Hypothesis Testing (p-value)**

Correct answer: b) Strong evidence against the null hypothesis

Justification:

A small p-value indicates that the observed result is unlikely under H_0, so it provides evidence **against** the null hypothesis.

EXERCISE 15

Exercise 15 Sampling Distribution (Idea)

Which statement about the sampling distribution of the sample mean is correct?

a) It always has the same variance as the population

b) Its standard deviation decreases as sample size increases

c) It depends on the sample mean only

d) It is never approximately normal

EXERCISE 15: SOLUTION

✅ Answer:

✅ **Exercise 14 Sampling Distribution**

Correct answer: b) Its standard deviation decreases as sample size increases

Justification:

The standard deviation of the sample mean is:

$$\frac{\sigma}{\sqrt{n}}$$

As n increases, the variability decreases.

EXERCISE 16

Exercise 16 Margin of Error (Conceptual)

Which action will **reduce the margin of error** of a confidence interval?

a) Decreasing the sample size

b) Increasing the confidence level

c) Increasing the sample size

d) Increasing the standard deviation

EXERCISE 16: SOLUTION

✅ Answer:

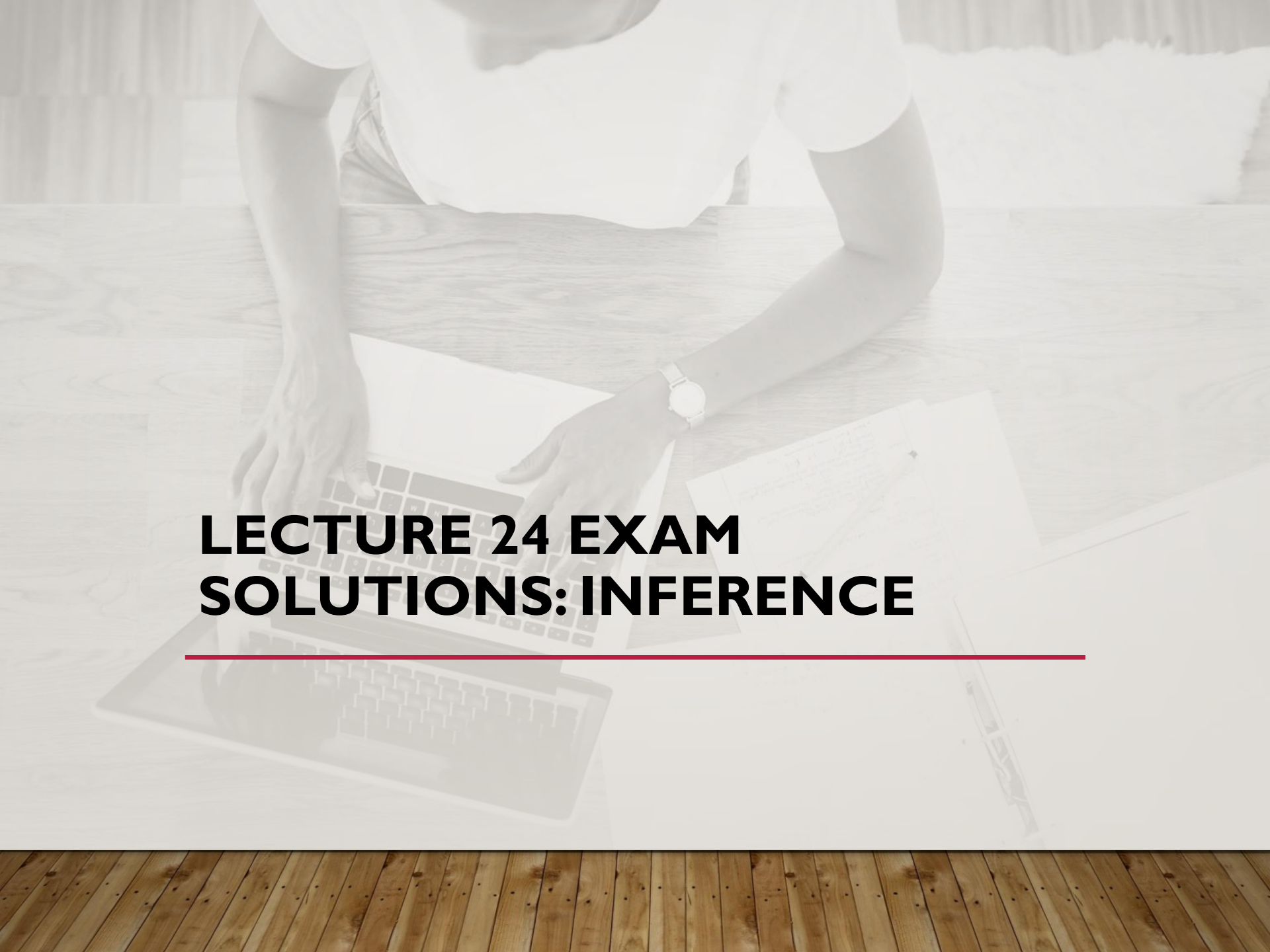
✅ **Exercise 16 Margin of Error**

Correct answer: c) Increasing the sample size

Justification:

The margin of error is inversely proportional to $\sqrt{n}$.

Larger samples → smaller margin of error.



LECTURE 24 EXAM SOLUTIONS: INFERENCE

EXERCISE 1

Question

Which of the following statements about the **sampling distribution of the sample mean** is **correct**?

A. The sampling distribution of the sample mean is always normal, regardless of the population distribution and sample size.

B. As the sample size increases, the standard deviation of the sampling distribution of the sample mean increases.

C. The mean of the sampling distribution of the sample mean is equal to the population mean.

D. The sampling distribution of the sample mean has the same variance as the population.

E. The sampling distribution of the sample mean becomes more skewed as the sample size increases.

EXERCISE I: SOLUTION

✅ Answer:

A. ❌ *The sampling distribution of the sample mean is always normal, regardless of the population distribution and sample size.*

This is **false**.
The sampling distribution of the sample mean is:

- normal **only if** the population is normal, **or**
- approximately normal when the sample size is sufficiently large (Central Limit Theorem).

For small samples from non-normal populations, the distribution is **not necessarily normal**.

B. ❌ *As the sample size increases, the standard deviation of the sampling distribution of the sample mean increases.*

This is **false**.
The standard deviation (standard error) of the sample mean is:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

As the sample size n increases, $\sqrt{n}$ increases, so the standard deviation **decreases**, not increases.

EXERCISE I: SOLUTION

✅ Answer:

D. ❌ *The sampling distribution of the sample mean has the same variance as the population.*

This is **false**.

The variance of the sampling distribution of the sample mean is:

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

which is **smaller than the population variance** for $n > 1$.

E. ❌ *The sampling distribution of the sample mean becomes more skewed as the sample size increases.*

This is **false**.

As the sample size increases, the sampling distribution of the sample mean becomes **less skewed** and more symmetric, approaching a normal distribution (Central Limit Theorem).

Final remark (exam style)

Only statement **C** is always true.

All other statements contradict fundamental properties of sampling distributions and the Central Limit Theorem.

EXERCISE 2

2. A simple linear regression analysis was performed. It can be assumed that: [1.0]

a) All variables are qualitative.

b) A histogram was constructed.

c) The coefficient of determination was calculated.

d) The coefficient of variation was calculated.

e) A contingency table was constructed.

EXERCISE 2: SOLUTION

✅ Answer:

a) **All variables are qualitative.**

❌ False — regression requires **quantitative** variables.

b) **A histogram was constructed.**

❌ False — a histogram is not required for regression analysis.

c) **The coefficient of determination was calculated.**

✅ True — R^2 **is always computed** in simple linear regression.

d) **The coefficient of variation was calculated.**

❌ False — regression does not require or guarantee a CV calculation.

e) **A contingency table was constructed.**

❌ False — contingency tables are used for qualitative variables, not regression.

✅ **Correct answer: c) The coefficient of determination was calculated.**

EXERCISE 3

3. In a financial company, the profit generated by each investment over the past year has a population mean of €8,200 and a population standard deviation of €3,600. Consider a random sample of 100 investments.

3.1. The sampling distribution of the mean of the 100 investments is approximately: [2.5]

a) $t_{(99)}$.
b) $N(0\ ,\ 1)$.
c) $N(\mu = 8200\ ,\ \sigma = 3600)$.
d) $N(\mu = 8200\ ,\ \sigma = 360)$.
e) None of the above.

3.2. The probability that the sample mean differs from the population mean by more than €806.76 is: [2.5]

a) 0.0125.
b) 0.025.
c) 0.9575.
d) 0.9899.
e) None of the above.

EXERCISE 3.1: SOLUTION

✅ Answer:

3.1. Sampling distribution of the sample mean

- Population mean $\mu = 8200$.
- Population standard deviation $\sigma = 3600$.
- Sample size $n = 100$.
- By the Central Limit Theorem (and because σ is known), the sampling distribution of the sample mean $\bar{X}$ is approximately normal with

$$\bar{X} \sim N\left(\mu = 8200,\ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3600}{\sqrt{100}} = 360\right).$$

✔ **Correct answer: d)** $N(\mu = 8200,\ \sigma = 360)$.

EXERCISE 3.2: SOLUTION

✅ Answer:

3.2. Probability that $|X - \mu| > 806.76$

We compute the Z–score for the deviation:

$$Z = \frac{806.76}{\sigma_{\bar{X}}} = \frac{806.76}{360} = 2.241.$$

This is a two–sided probability:

$$P\big(|\bar{X} - \mu| > 806.76\big) = 2\big(1 - \Phi(2.241)\big).$$

Using the standard normal CDF,

$$\Phi(2.241) \approx 0.98749 \quad \Rightarrow \quad 2(1 - \Phi(2.241)) \approx 0.02503 \approx 0.025.$$

✔ **Correct answer: b) 0.025.**

EXERCISE 4

4. A pharmaceutical company produces tablets in which the variability of the active ingredient from one tablet to another must be very small. The population standard deviation is supposedly 1 milligram. Inspectors from the Ministry of Health selected a random sample of 16 tablets. Assuming the population is normally distributed, the probability that the corrected sample variance exceeds 0.736 mg^2 is: [2.5]

a) 0.25.
b) 0.30.
c) 0.75.
d) 0.78.
e) None of the above.

EXERCISE 4: SOLUTION

✅ Answer:

We use the fact that for a normal population the corrected sample variance S^2 satisfies

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Given $\sigma = 1$, $n = 16$ and the threshold 0.736, compute the chi-square value:

$$\chi^2_{\text{obs}} = \frac{(n-1)\,0.736}{1^2} = 15 \times 0.736 = 11.04.$$

We want

$$P\big(S^2 > 0.736\big) = P\big(\chi^2_{15} > 11.04\big).$$

Using the chi-square distribution (df = 15),

$$P(\chi^2_{15} > 11.04) \approx 0.7498 \approx 0.75.$$

✔ **Correct answer: c) 0.75.**

EXERCISE 5

5. The daily price (€) per liter of crude oil on the stock market is a random variable, X, with distribution $N(\mu, \sigma)$. To better understand this price, a random sample of 30 days was selected, and the corresponding daily prices per liter of crude oil were recorded. From this data, the sample mean and corrected sample variance were obtained: $\bar{x}$ = €1.16 and s^2 = 0.1567 €². A 95% confidence interval for the true mean daily price per liter of crude oil on the stock market is given by: [3.5]

a) (1.0183; 1,3017).
b) (1.0122; 1.3078).
c) (1.0148; 1.3052).
d) (1.0411; 1.2789).
e) None of the above.

EXERCISE 5: SOLUTION

✅ Answer:

Given: $\bar{x} = 1.16,\ s^2 = 0.1567,\ n = 30$, confidence level 95%.

Since σ is unknown and $n = 30$, use the t-distribution with $df = n - 1 = 29$.

1. Compute the sample standard deviation s:

$$s = \sqrt{0.1567} = 0.3958535083 \text{ (approx.)}$$

2. Standard error of the mean:

$$SE = \frac{s}{\sqrt{n}} = \frac{0.3958535083}{\sqrt{30}} = 0.07227263198 \text{ (approx.)}$$

3. t-critical value for a two-sided 95% CI with 29 degrees of freedom:

$t_{0.975;\,29} \approx 2.045.$

EXERCISE 5: SOLUTION

✅ Answer:

4. Margin of error:

$$\mathrm{ME} = t_{0.975;\,29} \times SE = 2.045 \times 0.07227263198 \approx 0.1477975324$$

5. 95% confidence interval:

$$\bar{x} \pm \mathrm{ME} = 1.16 \pm 0.1477975324$$

$$\Rightarrow (1.0122024676,\ 1.3077975324)$$

Rounded to the 4 decimals shown in the choices, this is approximately **(1.0122; 1.3078)**.

✔ **Correct answer: b) (1.0122; 1.3078).**

EXERCISE 6

6. In a confidence interval for the population proportion, p, with 95% confidence, if the sample size decreases, the margin of error: [1.0]

a) Increases.

b) Decreases.

c) Does not change.

d) Becomes zero.

e) Becomes negative.

EXERCISE 6: SOLUTION

✅ Answer:

The **margin of error (ME)** for a proportion is given by:

$$ME = z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Where:

- $z_{1-\alpha/2}$ the critical value from the standard normal distribution (for 95% confidence, $z_{0.975} \approx 1.96$),
- $\hat{p}$ is the sample proportion,
- n is the sample size.

Step 1: Look at the formula

$$ME \propto \frac{1}{\sqrt{n}}$$

- This means **as n increases, the margin of error decreases.**
- Conversely, **if n decreases, the margin of error increases.**

EXERCISE 6: SOLUTION

✅ Answer:

Step 2: Analyze the options

a) **Increases.** ✅ This matches the formula.

b) **Decreases.** ❌ This is the opposite of what happens.

c) **Does not change.** ❌ The margin of error depends on n, so it changes.

d) **Becomes zero.** ❌ It cannot become zero unless n is infinite, which is not the case.

e) **Becomes negative.** ❌ Margin of error is always positive.

✅ **Step 3: Correct answer**

a) Increases.

EXERCISE 7

7. In a hypothesis test H_0: μ ≥ 25 versus H_1: μ < 25, the p-value obtained was 0.057. For a significance level of α = 0.05, which of the following conclusions is appropriate? [1.0]

a) There is statistical evidence to conclude that the population mean is 25.

b) There is no statistical evidence to conclude that the population mean is less than 25.

c) There is statistical evidence to conclude that the population mean is less than 25.

d) There is no statistical evidence to conclude that the population mean is 25.

e) There is no statistical evidence to conclude that the population mean is greater than or equal to 25.

EXERCISE 7: SOLUTION

✅ Answer:

We are given:

- Null hypothesis: $H_0 : \mu \geq 25$
- Alternative hypothesis: $H_1 : \mu < 25$ (this is a **left-tailed test**)
- p-value = 0.057
- Significance level: $\alpha = 0.05$

Step 1: Compare p-value with α

- **Rule:** If $p\text{-value} \leq \alpha$, reject H_0.
- Here: $p = 0.057 > 0.05$ → **do not reject H_0.**

Step 2: Interpret the conclusion

- "Do not reject H_0" means **there is not enough statistical evidence to support H_1.**
- H_1 is $\mu < 25$, so we **cannot conclude the mean is less than 25.**

EXERCISE 7: SOLUTION

✅ Answer:

Step 3: Check the options

a) **There is statistical evidence to conclude that the population mean is 25.** ❌

- We never "conclude" a specific value; we only fail to reject H_0.

b) **There is no statistical evidence to conclude that the population mean is less than 25.** ✅

- This matches our reasoning.

c) **There is statistical evidence to conclude that the population mean is less than 25.** ❌

- p-value > 0.05 → no evidence for H_1.

d) **There is no statistical evidence to conclude that the population mean is 25.** ❌

- The test does not prove equality; failing to reject H_0 does **not prove** $\mu = 25$.

e) **There is no statistical evidence to conclude that the population mean is greater than or equal to 25.** ❌

- Actually, we **do not reject** H_0, so there is no evidence against $\mu \geq 25$.

✅ **Step 4: Correct answer**

b) There is no statistical evidence to conclude that the population mean is less than 25.

EXERCISE 8 TO 10

8 to 10. The store "Grand Sales" has observed in recent years that 15% of its customers pay for their purchases by check, 38% by credit card, 32% by debit card, and 15% in cash. A sample of 160 sales made during the week before Christmas revealed the following results:

Payment Type	Check	Credit Card	Debit Card	Cash
Number of Sales	27	65	48	20

Is the type of payment used by "Grand Sales" customers during the Christmas period consistent with the store's historical information (α = 10%)? To answer this question, a statistician performed a statistical analysis and obtained the following outputs from the SPSS software package (Asymp. Sig. = p-value):

	Observed N	Expected N	Residual
Cheque	27	24,0	3,0
Cartão de crédito	65	60,8	4,2
Cartão de débito	48	51,2	-3,2
Dinheiro	20	24,0	-4,0
Total	160		

	Tipo de pagamento
Chi-Square	1,532[a]
df	3
Asymp. Sig.	,675

EXERCISE 8 TO 10

8. The null hypothesis (H_0) of the hypothesis test is (where p_1, p_2, p_3, and p_4 are the proportions of customers who paid by check, credit card, debit card, and cash, respectively): [1.0]

a) H_0: $p_1 = 0.25$, $p_2 = 0.25$, $p_3 = 0.25$, $p_4 = 0.25$.

b) H_0: $p_1 = 0.24$, $p_2 = 0.608$, $p_3 = 0.512$, $p_4 = 0.24$.

c) H_0: $p_1 = 0.15$, $p_2 = 0.38$, $p_3 = 0.32$, $p_4 = 0.15$.

d) H_0: $p_1 = 0.50$, $p_2 = 0.50$, $p_3 = 0.50$, $p_4 = 0.50$.

e) H_0: $p_1 = 0.25$, $p_2 = 0.50$, $p_3 = 0.50$, $p_4 = 0.25$.

EXERCISE 8: SOLUTION

✅ Answer:

We are given:

- Historical proportions of payment types:
  - Check = 15% → $p_1 = 0.15$
  - Credit card = 38% → $p_2 = 0.38$
  - Debit card = 32% → $p_3 = 0.32$
  - Cash = 15% → $p_4 = 0.15$
- Sample size $n = 160$ and observed counts (from table)

Question 8: Null hypothesis H_0

The **null hypothesis** for a chi-square goodness-of-fit test is:

$$H_0 : \text{The observed proportions match the historical proportions.}$$

This means:

$$H_0 : p_1 = 0.15,\ p_2 = 0.38,\ p_3 = 0.32,\ p_4 = 0.15$$

✅ Correct answer: c)

EXERCISE 9

9. The rejection region (RR) of the hypothesis test mentioned above is ($\alpha = 10\%$): [2.5]

a) RR = $]-\infty; 6.251]$.

b) RR = $]-\infty; -6.251]$.

c) RR = $]-\infty; -6.251] \cup [6.251; +\infty[$.

d) RR = $[-6.251; +\infty[$.

e) RR = $[6.251; +\infty[$.

EXERCISE 9: SOLUTION

✅ Answer:

Step 1: State the hypotheses

- **Null hypothesis H_0:** The observed proportions match historical proportions.

$$H_0 : p_1 = 0.15,\ p_2 = 0.38,\ p_3 = 0.32,\ p_4 = 0.15$$

- **Alternative hypothesis H_1:** At least one proportion differs from historical.

$$H_1 : \text{At least one } p_i \neq \text{historical } p_i$$

Step 2: Identify the test statistic

The **chi-square statistic** is:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where:

- O_i = observed frequency
- $E_i = n \cdot p_i$ = expected frequency

EXERCISE 9: SOLUTION

✅ Answer:

Step 3: Determine the rejection region (RR)

- For a chi-square goodness-of-fit test:
  - The test statistic χ^2 **cannot be negative**.
  - RR is in the **right tail** of the chi-square distribution.
  - Degrees of freedom: $df = k - 1 = 4 - 1 = 3$
- At $\alpha = 0.10$, the **critical value** $\chi^2_{0.10,3} \approx 6.251$ (from chi-square table).
- **Rejection region:**

$$RR = \{\chi^2 > 6.251\} \quad \text{or in interval notation: } [6.251, +\infty[$$

Step 4: Match with options

- e) RR = [6.251;+∞[✅ Correct.

Step 5: Notes

- Options a–d involve negative numbers, which are impossible for chi-square.
- Chi-square values are always ≥ 0, so only a right-tail region makes sense.

EXERCISE 10

10. Observing the SPSS outputs (above), it is known that the p-value = 0.675. Therefore, it can be stated that: [1.5]

a) At a 10% significance level, the type of payment used by customers during the Christmas period is consistent with the store's historical information.

b) At a 10% significance level, the type of payment used by customers during the Christmas period is not consistent with the store's historical information.

c) The type of payment used by customers during the Christmas period is consistent with the store's historical information.

d) The type of payment used by customers during the Christmas period is not consistent with the store's historical information.

e) No conclusion can be drawn with the data available.

EXERCISE I0: SOLUTION

✅ Answer:

We are given:

- **p-value = 0.675**
- **Significance level α = 0.10**
- Hypothesis test: chi-square goodness-of-fit (testing whether observed payment types match historical proportions)

Step 1: Compare p-value with α

- **Rule:**
  - If $p \leq \alpha$, reject H_0 → evidence that the distribution differs
  - If $p > \alpha$, **do not reject** H_0 → insufficient evidence to say the distribution differs

Here: $p = 0.675 > 0.10$ → **do not reject** H_0

EXERCISE I0: SOLUTION

✅ Answer:

Step 2: Interpret conclusion

- **Failing to reject H_0** means:
  - The observed payment types are **consistent with historical proportions**
  - **We cannot say they are inconsistent**

EXERCISE I0: SOLUTION

✅ Answer:

Step 3: Check the options

a) **At a 10% significance level, the type of payment used by customers during the Christmas period is consistent with the store's historical information.** ✅ Correct

b) At a 10% significance level, the type of payment used ... is **not consistent** ❌ Wrong (we did not reject H_0)

c) The type of payment used ... is consistent ... ❌ Slightly overgeneralized; significance level must be mentioned

d) ... is not consistent ❌ Wrong

e) No conclusion ... ❌ Wrong (we can conclude at the given significance level)

✅ **Step 4: Correct answer**

a) At a 10% significance level, the type of payment used by customers during the Christmas period is consistent with the store's historical information.

THANKS!

Questions?