Bachelor's Degrees in Management and Finance and Accounting

Statistics 2

Confidence Intervals

Exercises with SPSS outputs

Academic year 2024/2025.

Problem 1:

A machine produces screws. The diameter of these screws is normally distributed, and the standard deviation is unknown. A random sample of 25 screws was obtained and the following SPSS output was obtained.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Diameter (mm)	25	100,0%	0	0,0%	25	100,0%

Descriptives

			Statistic	Std. Error
Diameter (mm)	Mean		12,01	,10000
	99% Confidence Interval for Mean	Lower Bound	11,73	
		Upper Bound	12,29	
	5% Trimmed Mean		12,01	
	Median		12,04	
	Variance		,25	
	Std. Deviation		,50	
	Minimum		10,80	
	Maximum		13,19	
	Range		2,38	
	Interquartile Range		,59	
	Skewness		-,054	,464
	Kurtosis		,843	,902

Interpret the 99% confidence interval for the average diameter of screws produced by this machine and the values involved in its computation that are presented in the output.

Problem 2

Two different Sports Centres decided to compare the average weight of its athletes. Two random samples were collected and using SPSS (**Analyse → Compare Means → Independent samples T-test**, see ***SPSS support sheet_Part3***) the following output was obtained

Group Statistics

	centro	N	Mean	Std. Deviation	Std. Error Mean
Weight	Centro A	10	77,00	6,32456	2,00000
	Centro B	10	68,00	10,54093	3,33333

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									90% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Weight	Equal variances assumed	2,743	,115	2,315	18	,033	9,00	3,89	2,26	15,74
	Equal variances not assumed			2,315	14,737	,035	9,00	3,89	2,18	15,82

Indicate and interpret the 90% confidence interval for the difference of mean weights between Sports Centres, based on these samples. (Assume weights at both centres to be normally distributed and equality of variances between populations)

Problem 3

A random sample of 25 students was taken to carry out a study on students in the field of Management Sciences. One of the characteristics analyzed was the age of the students attending the courses. The descriptive results obtained are summarized in the following table:

			Statistic	Std. Error
Age	Mean		21,40	**a**
	95% Confidence Interval for Mean	Lower Bound	**b**	
		Upper Bound	**c**	
	5% Trimmed Mean		21,23	
	Median		21,00	
	Variance		5,667	
	Std. Deviation		2,380	
	Minimum		18	
	Maximum		28	
	Range		10	
	Interquartile Range		3	
	Skewness		1,193	,464
	Kurtosis		1,233	,902

Complete the table by determining the values **a**, **b** and **c**. Interpret the meaning of the estimated interval.

Answer:

X - Age of students, in years, attending Management Sciences courses

n = 25 observations

We know that the CI for the mean is of the form $\bar{X} \pm t \frac{s'}{\sqrt{n}}$

(a) is the standard error of the mean: $\frac{s'}{\sqrt{n}} = \frac{2,380}{\sqrt{25}} = 0,476$

(b) and (c) are the lower and upper limits of a 95% CI for the population mean

Since $\bar{x} = 21,40$, the CI is given by

$$21,40 \pm t \times 0,476$$

We need to find the value t → In this case*, the Pivotal Variable is

$$\frac{\bar{X} - \mu}{\frac{s'}{\sqrt{n}}} \sim t_{(n-1)}$$

Since n=25,

$$\frac{\bar{X}-\mu}{\frac{s'}{\sqrt{n}}} \sim t_{(24)}$$

and we want to determine a 95% CI, we will look for $t_{(24;0.975)}$

Consulting the t-student distribution table, we find $t = 2{,}064$ and obtain

$$(a) = 21{,}40 - 2{,}064 \times 0{,}476 = 20{,}42$$

and

$$(b) = 21{,}40 + 2{,}064 \times 0{,}476 = 22{,}38$$

Interpretation:

With 95% confidence, the true average age is expected to be between 20,42 years and 22,38 years.

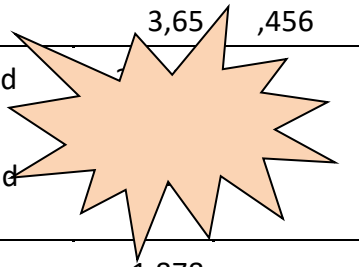
*It should be added that in order to use this Pivotal Variable with the aforementioned t-distribution, it is necessary to assume the normality of the population distribution (age in years).

Problem 4

The owner of Bar Oceano wants to conduct a study on the expenses that his customers make in his bar. The local university's Statistics professor decided to help him and discreetly recorded the expenses of 17 customers (randomly selected). The results obtained are presented in the following table.

Table 1: Descriptives

			Statistic	Std. Error
Expense at the bar	Mean		3,65	,456
	95% Confidence Interval for Mean	Lower Bound **(a)**		
		Upper Bound **(b)**		
	Std. Deviation		1,878	
	Minimum		1,30	
	Maximum		6,50	



However, in the hustle and bustle of the bar, some beer drops fell on some of these numbers (indicated by **(a)** and **(b)** in the table) making them unrecognizable. Help the owner of Bar Oceano.

Calculate the values of **(a)** and **(b)** and interpret the Confidence Interval in the context of the problem.

Answer:

X – Expenses incurred at the Bar, in euros

(a) and (b) are the lower and upper limits of a 95% CI for the population mean

n = 17 observations

We know that the CI for the mean is of the form $\bar{X} \pm t \frac{s'}{\sqrt{n}}$

$$\frac{s'}{\sqrt{n}} = standard\ error\ of\ the\ mean = 0{,}456$$

$$\bar{x} = 3{,}65$$

Thus, the CI is given by

$$3{,}65 \pm t \times 0{,}456$$

We need to find the value t → In this case*, the Pivotal Variable

$$\frac{\bar{X} - \mu}{\frac{s'}{\sqrt{n}}} \sim t_{(n-1)}$$

Since n=17,

$$\frac{\bar{X}-\mu}{\frac{s'}{\sqrt{n}}} \sim t_{(16)}$$

and we want to determine a 95% CI, we will look for $t_{(16;0.975)}$

Consulting the t-student distribution table, we find $t = 2{,}12$ and get,

$$(a) = 3{,}65 - 2{,}12 \times 0{,}456 = 2{,}68$$

and

$$(b) = 3{,}65 + 2{,}12 \times 0{,}456 = 4{,}62$$

Interpretation:

With 95% confidence, the true average expenditure at Bar Oceano is expected to be between 2,68 € and 4,62 €.

* It should be added that in order to use this Pivotal Variable with the aforementioned t-distribution, it is necessary to assume that the distribution of the population is normal (Expenditure incurred at the Bar, in €)

Problem 5 (Midterm test from 2018/2019)

As part of a market study on buying behavior, participants were asked about the probability (measured in percentage, from 0% to 100%) of making online purchases in the next 6 months. For this purpose, it was considered important to stratify the population into two age groups:

Group 1: up to 35 years old (inclusive) and

Group 2: over 35 years old

To evaluate whether, on average, the probability of making an online purchase in the next few months differs between the two age groups, a confidence interval was constructed for the parameter related with this objective. The available outputs are as follows

Group Statistics

	Grupo etário	N	Mean	Std. Deviation	Std. Error Mean
probability of making online purchases in the next 6 months	**1**-up to 35 years	50	45,40	28,317	4,005
	2-over 35 year	38	43,29	18,754	3,042

Independent Samples Test

		Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
				Lower	Upper
probability of making online purchases in the next 6 months	Equal variances assumed	2,111	5,308	-8,440	12,661
	Equal variances not assumed	2,111	5,029	**A**	**B**

Note: In solving questions a) and b), assume that there is no significant difference in variability between the groups under analysis

a) Of the following expressions, which is the appropriate choice for the Pivotal Variable to be used in the construction of the required confidence interval?

(A) $\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$

(B) $\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

(C) $\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$

(D) $\frac{\bar{X}_1-\mu_1}{S_1/\sqrt{n_1}}-\frac{\bar{X}_2-\mu_2}{S_2/\sqrt{n_2}}$

Answer: B

b) Indicate, justifying your answer, whether the following statement is true or false: *"In the population under study and for the next 6 months, the difference in the (average) probability of making an online purchase between the two age groups is between -8.440 and 12.111 with a probability of 0.95."*

Answer:
False. In the population under study and for the next 6 months, the difference in the (average) probability of making an online purchase between the two age groups is between -8.440 and 12.661, with 95% confidence.

c) In the problem under study, assume that the 95% confidence interval constructed on the basis of the assumption that there are significant differences in variability between the groups under analysis is more accurate than the 95% confidence interval constructed on the basis of the assumption that there are no significant differences. Given this information and the presented Output, indicate which of the following options could correspond to a possible value for **A** and a possible value for **B**.

(A) $\mathbf{A} = -7{,}899$ e $\mathbf{B} = 12{,}111$

(B) $\mathbf{A} = -8{,}899$ e $\mathbf{B} = 13{,}111$

(C) $\mathbf{A} = -9{,}899$ e $\mathbf{B} = 10{,}111$

(D) None of the above options corresponds to a possible value for A and a possible value for B.

Answer: A

For the same percentage of confidence (95%), as the precision increases, the amplitude of the CI decreases, and both extremes of the CI get closer.