Statistisk Dataanalyse 2023

Instruktør: Vivek Misra

Exercise Class NR4

Solutions to the Tasks

Task 1 - Description

- According to the Havard Business Review (in the article: ”How to Spend Way Less Time on Email Every Day”), the average professional checks his/her emails 15 times per day.
- The data represent a sample of the number of the number of times/year, that 7 employees in a company, ech their emails.

Emals of 7 Employees	5460	5900	6090	6310	7160	8440	9930

a) We can be 99% confident that the mean number of times that the employees of this company check their email each year is between 4785 and 9298.

b) We can be 99% confident that the mean number of times that the employees of this company check their email is not significantly different from that of the ”average professional”.

c) None of the previous responses is correct.

d) A and B are correct.

Task 1 - Solution

- Solution: To solve this tasks, we will first find the mean and then the standard-deviation. Thereafter we will use the formula to find if the mean lies between the given confidence intervals.

Sum of the Sample (7 Employees)	$\mathbf{5460 + 5900 + 6090 + 6310 + 7160 + 8440 + 9930 = 49290}$
Mean of the Sample	$\mathbf{\frac{49290}{7} = 7041,429}$

Range of Mean and Values	$\mathbf{7041,429 - 5460}$	$\mathbf{7041,429 - 5900}$	$\mathbf{7041,429 - 6090}$	$\mathbf{7041,429 - 6310}$	$\mathbf{7160 - 7041,429}$	$\mathbf{8440 - 7041,429}$	$\mathbf{9930 - 7041,429}$
Results from Substraction	1581,429	1141,429	951,429	731,429	118,571	1398,571	2888,571
Squaring the values	$1581,429^2$	$1141,429^2$	$951,429^2$	$731,429^2$	$118,571^2$	$1398,571^2$	$2888,571^2$

Task 1 - Solution

- Solution: Now here we have found the standard-deviation

Sum of the Squared Values	$\mathbf{2900918 + 1302860 + 905217,1 + 534988,4 + 14059,08 + 1956001 + 8343842 \approx 1,555789 \cdot 10^7}$
Mean of the Sample	$\mathbf{\sqrt[2]{\frac{1,555789 \cdot 10^7}{7-1}} = 1610,274}$

- Now we will calculate degrees of freedom, along with the critical value. Here we have the value N-1, which is for our case 7-1=6.
- We have already defined a confidence level which is 3,7074 from Student-T-Test.
- Now we just need to implement the formula, shown on the next slide and see whether there is a difference between the populations mean and the sample mean.

Task 1 - Solution

$$\overline{x} - t_{N-1} \cdot \frac{s}{\sqrt[2]{N}} < \mu < \overline{x} + t_{N-1} \cdot \frac{s}{\sqrt[2]{N}}$$

- Now we just need to add the values inside the formula.

$$7041,429 - 3,7074 \cdot \frac{1610,274}{\sqrt[2]{7}} = 7041,429 - \frac{5969,93}{\sqrt{7}} \approx 4785,008$$

$$7041,429 + 3,7074 \cdot \frac{1610,274}{\sqrt[2]{7}} = 7041,429 + \frac{5969,93}{\sqrt{7}} = 9297,85$$

- We can be confident that 99% of the employees of the company check their email each year between 4785 < 7041,429 < 9298.

Task 2 – Description / Solution

- A job advisor claims that the average salary for engineers is 24.000 euros/year. Ten engineers are randomly selected and the mean of their salaries is 23.450 euros/year and a standard deviation of 400 euros/year. Is there evidence (at 95% confidence level) to reject the statement of the job advisor?

- Short-Solution: We can just by looking at the statement say that there is 95% confidence level for rejecting the statement of the job advisor. The reason is clear, and that the small sample group has almost the same average salary for all engineers

Task 2 - Solution

- Long-Solution: We will use the one-sample t-test to see whether there is a difference between the population mean and the sample mean.

$$\overline{x} - t_{N-1} \cdot \frac{s}{\sqrt[2]{N}} < \mu < \overline{x} + t_{N-1} \cdot \frac{s}{\sqrt[2]{N}}$$

- We have found the following values from the description of the task.

$$\overline{x} = 23.450$$
$$t_{N-1} = 2,26$$
$$s = 400\,€\ pr.year$$
$$N = 10$$

Task 2 - Solution

$$\overline{x} - t_{N-1} \cdot \frac{s}{\sqrt[2]{N}} < \mu < \overline{x} + t_{N-1} \cdot \frac{s}{\sqrt[2]{N}}$$

- We are inserting the values now:

$$23450 - 2,6222 \cdot \frac{400}{\sqrt[2]{10}} = -\frac{904,88}{\sqrt[2]{10}} + 23450 \approx 23165,85$$

$$23450 + 2,6222 \cdot \frac{400}{\sqrt[2]{10}} = \frac{904,88}{\sqrt[2]{10}} + 23450 \approx 23736,15$$

- Now we will calculate the confidence level of the values:

$$23165{,}85 < 23450 < 23736{,}15$$

- There is a difference between the populations mean and sample and therefore the alternative hypothesis is accepted.

Task 3 - Description

- The number of children born in 7 towns in a region is:

Number of Children	**7540**	**8421**	**8560**	**7412**	**8953**	**7859**	**6098**

- Find the 99% confidence interval for the mean number of children born annually per town.

Task 3 - Solution

- Solution: To solve this tasks, we will first find the mean and then the standard-deviation. Thereafter we will use the formula to find if the mean lies between the given confidence intervals.

Sum of the Sample (7 Employees)	$\mathbf{7540 + 8421 + 8560 + 7412 + 8953 + 7859 + 6098 = 54842}$
Mean of the Sample	$\mathbf{\frac{54843}{7} = 7834,714}$

Range of Mean and Values	$\mathbf{7834,714 - 7540}$	$\mathbf{8421 - 7834,714}$	$\mathbf{8560 - 7834,714}$	$\mathbf{7834,714 - 7412}$	$\mathbf{8953 - 7834,714}$	$\mathbf{7859 - 7834,714}$	$\mathbf{7834,714 - 6098}$
Results from Substraction	294,714	586,29	725,29	422,714	1118,286	24,286	1736,714
Squaring the values	$29.714^2 \approx 882,9218$	$586,29^2 \approx 343736$	$725,29^2 \approx 526045,6$	$422,7149^2 \approx 178687,1$	$1118,286^2 \approx 1250564$	$24,286^2 \approx 589,8098$	$1736,714^2 \approx 3016176$

Task 3 - Solution

- Solution: Now here we have found the standard-deviation.

Sum of the Squared Values	$\mathbf{882,9218 + 343736 + 526045,6 + 178687,1 + 1250564 + 589,8098 + 3016176 \approx 5316681}$
Mean of the Sample	$\sqrt[2]{\frac{\mathbf{5316681}}{\mathbf{7-1}}} = \mathbf{941,336}$

- Now we will calculate degrees of freedom, along with the critical value. Here we have the value N-1, which is for our case 7-1=6.
- We have already defined a confidence level which is 3,7074 from Student-T-Test.
- Now we just need to implement the formula, shown on the next slide and see whether there is a difference between the populations mean and the sample mean.

Task 3 - Solution

$$\overline{x} - t_{N-1} \cdot \frac{s}{\sqrt[2]{N}} < \mu < \overline{x} + t_{N-1} \cdot \frac{s}{\sqrt[2]{N}}$$

- Now we just need to add the values inside the formula.

$$7834,71 - 3,7074 \cdot \frac{941,336}{\sqrt[2]{7}} = 7834,71 - \frac{3489,909}{\sqrt{7}} \approx 6515,652$$

$$7834,71 + 3,7074 \cdot \frac{941,336}{\sqrt[2]{7}} = 7834,71 + \frac{3489,909}{\sqrt{7}} \approx 9153,776$$

- We can be 99% confident, that the null-hypothesis is true and the interval is: 4785 < 7041,429 < 9298.

Task 4 - Description

- A survey of 50 students found that the mean age of their bicycles was 5.6 years. Assuming the standard deviation of the population is 0.8 years, which of these statements is correct.

a) Based on the sample of 50 students, we can be 90% confident that the mean age of all bicycles is between 5.297 and 5.903.

b) Based on the sample of 50 students, we can be 99% confident that the mean age of all bicycles is between 5.297 and 5.903.

c) A and B are correct.

d) None of them is correct.

Task 4 - Solution

- Solution: We will use the same one-sample t-test formula as before. But this time, we just need to gather all the information in which has been given to us in the text.
- N = amount of students = 50
- S = amount of years = 0.8
- X = Mean of the bicycles age = 5.6
- MU = Not described in the text!
- In the next page, we will tell you about the result which we got from the student-t test.

Task 4 - Solution

- Solution: We can see, that our student t-test in two-tailed shows a value of +-1,67 with a confidence level of 90%, whereas if 99% then we have a value of +-2,68.
- Now, we just need to use the formula of one-sample t-test.

$$\overline{\boldsymbol{x}} - \boldsymbol{t}_{\boldsymbol{N-1}} \cdot \frac{\boldsymbol{s}}{\sqrt[2]{\boldsymbol{N}}} < \boldsymbol{\mu} < \overline{\boldsymbol{x}} + \boldsymbol{t}_{\boldsymbol{N-1}} \cdot \frac{\boldsymbol{s}}{\sqrt[2]{\boldsymbol{N}}}$$

$$\mathbf{5,6} - \mathbf{1,6766} \cdot \frac{\mathbf{0,8}}{\sqrt[2]{\mathbf{50}}} = \mathbf{5,6} - \frac{\mathbf{0,268256}}{\sqrt{\mathbf{2}}} \approx \mathbf{5,410314}$$

$$\mathbf{5,6} + \mathbf{1,6766} \cdot \frac{\mathbf{0,8}}{\sqrt[2]{\mathbf{50}}} = \mathbf{5,6} + \frac{\mathbf{0,268256}}{\sqrt{\mathbf{2}}} \approx \mathbf{5,789686}$$

- We can see, that our interval is: 5,41<5,6<5,78.
- We can accept the null-hypothesis.

Task 4 - Solution

- Now, we will also calculate the 99% and here we will just change the critical value to +-2,68.

$$\overline{x} - t_{N-1} \cdot \frac{s}{\sqrt[2]{N}} < \mu < \overline{x} + t_{N-1} \cdot \frac{s}{\sqrt[2]{N}}$$

$$5,6 - 2,68 \cdot \frac{0,8}{\sqrt[2]{50}} = 5,6 - \frac{0,4288}{\sqrt{2}} \approx 5,296793$$

$$5,6 + 2,68 \cdot \frac{0,8}{\sqrt[2]{50}} = 5,6 + \frac{0,4288}{\sqrt{2}} \approx 5,903207$$

- **We have got the interval: 5,29<5,6<5,90 where the null-hypothesis is accepted!**

- **This means, that the null-hypothesis for both 90% and 99% are accepted.**

Task 5 - Description

- The following data represent a sample of the assets (in millions of euros) of 30 motherboards computer hardware manufacturers in Europe. Find the 90% confidence interval of the mean.

12.23	16.56	4.39
2.89	1.24	2.17
13.19	9.16	1.42
73.25	1.91	14.64
11.59	6.69	1.06
8.74	3.17	18.13
7.92	4.78	16.85
40.22	2.42	21.58
5.01	1.47	12.24
2.27	12.77	2.76

Task 6 - Solution

- Solution: We will use the same steps as before in the previous tasks. We have simplified some NOTE-Informations, such as: **We have already calculated the mean to be: 11,09. The variance has been calculated to be: 6017,991.**
- We just need to calculated the Standard Deviation for now:

$$sd = \sqrt[2]{\frac{6017{,}991}{30-1}} \approx 14{,}40$$

Task 6 - Solution

- Solution: Just to make things clear, we are working with a value of confidence level equal to 90%=0,10. The critical value has been found to be: +-1,6991.

$$\overline{\boldsymbol{x}} - \boldsymbol{t}_{\boldsymbol{N-1}} \cdot \frac{\boldsymbol{s}}{\sqrt[2]{\boldsymbol{N}}} < \boldsymbol{\mu} < \overline{\boldsymbol{x}} + \boldsymbol{t}_{\boldsymbol{N-1}} \cdot \frac{\boldsymbol{s}}{\sqrt[2]{\boldsymbol{N}}}$$

$$\mathbf{11,09067 - 1,6991} \cdot \frac{\mathbf{14,40}}{\sqrt[2]{\mathbf{30}}} = \mathbf{11,09067} - \frac{\mathbf{24,46704}}{\sqrt{\mathbf{30}}} \approx \mathbf{6,62362}$$

$$\mathbf{1,09067 + 1,6991} \cdot \frac{\mathbf{14,40}}{\sqrt[2]{\mathbf{30}}} = \frac{\mathbf{24,46704}}{\sqrt{\mathbf{30}}} + \mathbf{11,09067} \approx \mathbf{15,55772}$$

- The interval is: 6,62<11,09<15,55, where the null-hypothesis is accepted.

Tak for i dag!

Instruktør: Vivek Misra