Øvelsestest til Statistik

Prøve 1

1. The director of a university department wishes to see whether there is a difference in the knowledge of students of a course, depending on the teaching method used (1 or 2), and professor that teaches the course (A or B). Four students from each professor and teaching method are randomly selected and they are asked to attend to a single exam. Is there a significant effect of the professor and the teaching method on the student´s results? Which professor and/or method provided the best results?

A. Is Method 1 better with Professor A
B. Is Method 1 better with Professor B
C. Is Method 2 better with Professor A
D. Is Method 2 better with Professor B

	Professor A	Professor B
Method 1	67, 63, 79, 89	91, 95, 89, 88
Method 2	54, 53, 76, 64	85, 65, 76, 79

2. Mumbai and Delhi, also known as the two major cities of the Republic of India have been known for their cosmopolitan attraction. During the times a survey, has been done where the cities have been testing for their traffic and infrastructure. In this case the following data has been collected, where you have to find the significance between the two cities and draw an interactions plot from them.

A. Delhi has a better Infrastructure than Mumbai.
B. Mumbai has better Infrastructure than Delhi.
C. Delhi has better Traffic than Mumbai.
D. Mumbai has better Traffic than Delhi.

Subject	Delhi	Mumbai
Traffic	1,2,3,4,5	16,17,18,19,20
Infrastructure	10,11,12,13,14	6,7,8,9,10

3. Lahore and Karachi also known as Pakistan's Economic and Cultural Capitals. Though Lahore is considered as the most financial and culturally prosperous capital of Pakistan, the country is now seeing a battlement of entitlement of being known as the "cultural capital and economic capital" between Lahore and Karachi. Use the following Data and find the significance between the cities and construct an interaction plot where you can conclude, which country is the cultural and economic capital of Pakistan.

- **Lahore is better with the title of being "Cultural Capital of Pakistan".**
- Karachi is better with the title of being "Cultural Capital of Pakistan".
- **Karachi is better with the title of being "Financial Capital of Pakistan".**
- Lahore is better with the title of being "Financial Capital of Pakistan".

Subject	Lahore	Karachi
Cultural	10,11,12,13,14	0,1,2,3,5
Economic	0,1,2,2,4	15,16,17,18,19

Øvelsestest til Statistik

4. The number of hours of study of the students of a course and the final grade of the students (out of 100), is shown in the table. Calculate the correlation coefficient and determine whether the correlation is significant. Obtain the regression line.

Hours of Study	Grade
74	87
59	63
45	50
29	39
20.8	21
19.1	28
13.4	14
8.5	15

- In this case we can see, that our correlations coefficient is equal to 0,98 which is a linear strong positive graph. We have also drawn a regression line, which works!
- Looking, at the p-value, we can see that the p-value is equal to 0,000004537 which is lower than 95%=0,05. Therefore, the correlation is significant.

5. A researcher wants to investigate the relationship between sweets and blood sugar. In this case 6 persons have been chosen for this experiment. Use the appropriate test and find out the relationship between Sweets and Blood Sugar and construct a plot graph with a regression line.

People	Sweets Eaten	Blood Sugar
1	5	4.9
2	6	5.4
3	7	5.7
4	8	6.5
5	9	6.9
6	10	7.5

- In this case we can see, that our correlations p-value is significant because it is 0,00004164, which is lower than 0,05. Whereas if we look closer towards the graph, we can see that there is a linear positive strong regression.

6. Indian Punjabis and Pakistani Punjabis eat a lot of mustard leaves blended with spinach, and along with that they drink a lot of phirni and lassis. According to a Professor from the University of Cambridge, has conducted research where he has selected 10 persons to eat the dish of mustard leaves. From the research it is said that “sarso ka saag” or mustard leaves, has the ability to remove chance of getting alopecia areata. Use the significant test and construct an interaction plot, where a regression line can be added.

Personer	Saag	Alocpecia Areata
1	**10**	**10**
2	**11**	**9**
3	**12**	**8**

Øvelsestest til Statistik

4	**13**	**7**
5	**14**	**6**
6	**15**	**5**
7	**16**	**4**
8	**17**	**3**
9	**18**	**2**
10	**19**	**1**

- In this case we can see that the correlations coefficient is equal to -1, and that it is significant because the p-value is lower than 0,05.

7. A study has been conducted to determine the relationship, between the genetics of Danes and Germans. The following data are shown below:

   Use the appropriate test to see, whether there is a negative significant correlation?

Danes	20	19	18	17	16	15
German	5	6	7	8	9	10

- In this case we can see, that there is linear strong negative relationship, between the numbers. And that is because the two number in rows are going up and down.

8. A study has been conducted to determine the relationship, between the genetics of Indians and Pakistanis. The following data are shown below:

   Use the appropriate test to see, whether there is a negative significant correlation?

Indian	16	17	18	19	20	21
Pakistani	6	7	8	9	10	11

- In this case we can see that there is strong linear positive relationship, and that is because again the numbers are the opposite of eachother in the dataset.

9. A researcher wants to test, if Zink Included in the diet changes the vitamin D inside the body. Six subjects were pretested at Week 0 and they took the zink-medication during 7 weeks. After the 7-weeks period, their vitamin D level was measured again. Using R, can we conclude with 95% confidence level that the cholesterol level has been changed? Assume the variable is approximately normally distributed.

Subject	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Uge 0	220	244	213	195	177	249	355
Uge 7	189	165	206	193	174	224	325

What is the interval?

- We can see that the confidence interval is equal to 0,48 < 25,28 < 50,083

Is it significantly different?

- Whereas the p-value is equal to 0,04684, which is lower than 95%=0,05 and that is significant different.

10. According to the Harvard 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 times/years, that 7 employees in a company check their emails:

5460 5900 6090 6310 7160 8440 9930

Use R to find out: which one of the following statements is correct?

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.

11. We find 100 women with similar lung capacity who do not take Vitamin Pills. We randomly assign them to two groups (A and B). One group with 50 women will take Vitamin pills for 1 year, and the other 50 women will NOT take Vitamin pills for 1 year.

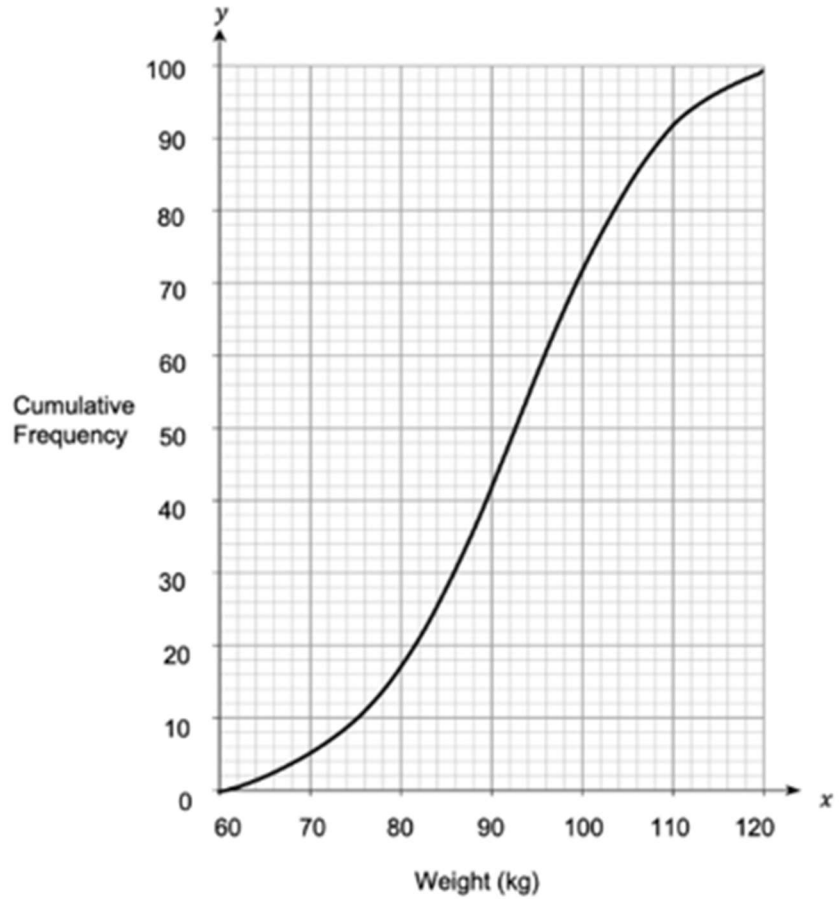
- Because we are controlling the groups throughout a year, then we can see that it is an experiment.

12. We find 50 women that have been taking Vitamin pills for 1 year. 50 women that have NOT been taking Vitamin pills for 1 year.

- Because we have 50 women, where we have not been involved inside it. We can say that this experiment is observational.

13. The below cumulative frequency graph shows the weight of 100 people who attend the program, Weight Watchers. Based on the given graph, evaluate the following sentences as TRUE or FALSE:

Øvelsestest til Statistik



i. The intertquartile range of the data is 60 kg.
ii. The median is 100 kg
iii. 143 kg is an outlier

a. I) FALSE, ii) FALSE, iii) TRUE
b. i)FALSE, ii) TRUE, iii) TRUE
c. i) TRUE, ii) FALSE, iii) FALSE
d. i) FALSE, ii) FALSE, iii) FALSE

14. Find the standard deviation of the following dataset: 1,2,3,4,6,7,8,9,10

- Our standard deviation, would become 3,027.

15. Find the mean of the following dataset: 999,856,789,234,987,234,333,555

- Our mean would become 623,375.

16. When evaluating the performance of students in a certain exam, the teacher calculated the standard deviation of students' scores and found a standard deviation of 4.5 points. In the following exam, the students performed better, and the score obtained by each student in class increased by 3 points compared to the previous exam. In this second exam, the standard deviation will:

A. Remain the same
B. Increase by the square root of 3 points
C. Increase by 9 points

D. Increase by 3 points

17. The US. Government wants to check if children in their country are getting the appropriate nutrition according to their age. Three groups have been selected for this task, which in this case are children from the age 0 to 5, the other age, which is 5 to 10, and the third one which is 10-17. **Use the appropriate statistical test and see if there is a significant difference among the three children's groups, in terms of having the same amount of nutrition at different ages or the same age.**

Newborn Children	Small Children	Young Children
500	500	500
700	700	700
1000	1200	1300
2000	1250	1250
2500	1280	1400
2575	1300	1700

- In this case we can see that the p-value is equal to 0,04684 which is lower than 0,05 and that tells us that it is significant.
- Whereas if we look closer towards the means of each row, we can see that Newborn=1545,833, Small=1038,333, Young=1141,667 and that tells us also that there is not any difference among newborn and young children. Whereas if we look towards small children, then there is a significant difference.

18. Pakistan wants to increase its military budget, but also wants to double the salary income of its army chief. The Government of Pakistan does not want to ask the army chief directly about the wages. Therefore, the Pakistani Government looks at the income of the three Dictators, they had while they were ruling the country. **Using the following data, use or conduct the appropriate test to see if there were a significant difference between the salaries of the 3 previous military dictators!**

Ayub Khan	Zia Ul-Haq	Pervez Musharraf
10	100	1000
20	120	2000
30	150	3000
40	170	5000
50	200	8000
60	210	10000

- In this case, we can look at the p-value and conclude that because it is 0,04684 which is lower than 0,05. Then it can be said that there is a significant difference.

Øvelsestest til Statistik

19. The Indian National Cricket Team wants to compare the scores obtained by three players who have either been or is currently captain of the Team, during the World Cup Finals. The INCT has chosen Virat Kohli, Mahindra Singh Dhoni and Sachin Tendulkar as primary objectives to collect the data. **When using the appropriate statistical test, is there a significant difference among the scores obtained by each of the players?**

Virat Kohli	Mahindra Singh Dhoni	Sachin Tendulkar
1	**2**	**3**
4	**5**	**6**
7	**8**	**9**
10	**11**	**12**
13	**14**	**15**
16	**17**	**18**

- In this case, we can see that the p-value is lower than 95%=0,05, which is 0,04684. Therefore there is a significant difference among the scores obtained by each of the players!

20. What is the definition of a null hypothesis?

a. It is a statistical hypothesis which says that there is not a statistical difference between the means of two populations.
b. A statistical test based on the F-distribution, which can be used for comparing the means of three or more populations.
c. It is an analysis of Variance.

Øvelsestest til Statistik

Prøve 2

1. Happy town has collected data in collaboration with M&M's. The data has been used to find out, how many people visit Legoland in Billund City in Denmark! **Use the following data and find the outlier and the interquartile range.**

55, 28, 29, 30, 40, 45, 60, 90, 80, 75, 78, 79, 69, 41, 42, 46, 47, 43, 103, 105

2. Denmark has recently made a statement about the weather condition in Billund. The reason behind this statement is to determine, whether flights can depart or land in the Runway at Billund Airport. **Use the following data below and determine the mean of the rainfall.**

20, 30, 40, 50, 65, 77, 88, 99, 11, 14, 16, 17, 19, 28, 28, 29, 56, 88, 89, 98, 102, 104

3. What is the definition of ANOVA?

a. Is a statistical method that allows us to determine if differences in mean value between three or more sample groups are by chance, or if the populations are indeed significantly different.
b. Is a statistical method that allows us to determine if differences in mean value between three or more sample groups are by chance, or if the populations are indeed significantly different.
c. Is a statistical hypothesis that states that there is a statistical difference between the population mean and a specific value.

4. Lufthansa, a member of the StarAlliance group has recently launched a non-stop service between Billund Airport and Munich Airport. Use R and construct a histogram, from the following dataset and conclude in which condition the skewness can be found.

Percentile	Data
100^{th}	90
80^{th}	70
70^{th}	60
60^{th}	50
50^{th}	40
40^{th}	30
30^{th}	20
20^{th}	10
10^{th}	5

a. Positive Skewed
b. Negative Skewed
c. Symmetrical Bell-Shaped
d. Uniform Shaped

e. Reverse J-Shaped
f. J-Shaped

5. In the upcoming years, Vistara has decided to merge with Air India and increase its fleet. The following dataset has been given. **Use the appropriate test and find out how many fleets the merged Airline will have in year 2030?**

Years	Fleets
2014	99
2015	98
2016	97
2017	100
2018	150
2019	200
2020	250
2021	500
2022	1000

6. In Denmark, Copenhagen and Aarhus have been through time competing each other for having the status as the financial and cultural capital of Denmark. In this case the following data have been collected on the basis of Economics and Cultural Aspects. **Use the appropriate test and find which city can be called the financial capital or the cultural capital of Denmark.**

X	Copenhagen	Aarhus
Financial Capital	11,12,13,14,15	5,6,7,8,9
Cultural Capital	9,10,11,12,13	16,17,18,19,20

7. A study has been conducted to determine the relationship, between the genetics of Danes and Germans. The following data are shown below:

Danes	20	19	18	17	16	15
German	5	6	7	8	9	10

Øvelsestest til Statistik

8. The Middle eastern Airlines, such as Emirates, Oman Air and Qatar Airways have been compared to see whether there are significantly different from each other in terms of on time departure and arrival. The following data has been collected. **Use the appropriate test and conclude whether there is a significant difference between the three airlines in terms of on time performance!**

Emirates	Oman Air	Qatar Airways
10	20	30
40	50	60
70	80	90
100	110	120
130	140	150
160	170	180
190	200	210

9. Three Airlines in the Indian Aviation Sector, which are Air India, Vistara and IndiGo wants to compare each other in terms of level of service. In this case, we can see that we have got in total 10 scores from the airlines which have been have collected by the Aviation Ministry of India. **Using the appropriate statistical test, and conclude if there is a significant difference among the scores obtained by each of the airlines?**

Air India	Vistara	IndiGo
100	500	1000
120	510	1100
140	520	1200
160	530	1300
170	540	1400
180	550	1500
200	570	1700
205	580	1800
207	605	2000
210	710	3150

10. A farmer wants to inspect whether his crops are growing fine in a about four weeks of time. The farmer has collected samples of data at the first week and the fourth week. The result has been as followed, shown on the table below. Using the right test to conclude, whether there is a significant difference in the growing of crops.

Subject	1	2	3	4	5	6	7
Week0	9	12	18	23	29	36	44
Week4	12	23	29	36	44	50	59

Øvelsestest til Statistik

11. The Coronavirus has been known as global pandemic back in the year 2020 and 2021. Researchers have through the research of the outcome of the virus, collected sample data to see if there is a relationship between the CPC and the virus. The coronavirus numbers show the rising cases in China, and numbers of CPC shows the disappearance of people. The following data has been discovered. **Use the appropriate test and determine the correlations coefficient and the regressionsline.**

Coronavirus	201	302	403	506	607
CPC	101	92	83	76	67

12. **The number of hours of study of the students of a course and the final grade of the students (out of 100), is shown in the table. Calculate the correlation coefficient and determine whether the correlation is significant. Obtain the regression line.**

Hours of Study	Grade
74	87
59	63
45	50
29	39
20.8	21
19.1	28
13.4	14
8.5	15

13. The External Ministry of Denmark wants to see, the number of travelers going through Copenhagen Airport in Kastrup. The ministry has sent one of its groups researchers to make research at week 0 and the other research at week 10.

Subject	1	2	3	4	5	6	7	8	9	10
Uge 0	110	120	130	140	150	160	170	180	190	200
Uge 10	210	220	230	240	250	260	270	280	290	300

What is the interval?

Is there a significance difference?

Øvelsestest til Statistik

14. The External Ministry of Denmark wants to see, the number of travelers going through Billund Airport in Syddanmark. The ministry has sent one of its groups researchers to make research at week 0 and the other research at week 10.

Subject	1	2	3	4	5	6	7	8	9	10
Uge 0	110	120	130	140	150	160	170	180	190	200
Uge 10	210	220	230	240	250	260	270	280	290	300

What is the interval?

Is there a significance difference?

15. The Indian National Cricket Team wants to compare the scores obtained by three players who have either been or is currently captain of the Team, during the World Cup Finals. The INCT has chosen Virat Kohli, Mahindra Singh Dhoni and Sachin Tendulkar as primary objectives to collect the data. **When using the appropriate statistical test, is there a significant difference among the scores obtained by each of the players?**

Virat Kohli	Mahindra Singh Dhoni	Sachin Tendulkar
1	**2**	**3**
4	**5**	**6**
7	**8**	**9**
10	**11**	**12**
13	**14**	**15**
16	**17**	**18**

Øvelsestest til Statistik

16. You are a researcher interested in social factors that influence heart disease. You survey 15 towns and gather data on the percentage of people in each town who smoke, the percentage of people in each town who bike to work, and the percentage of people in each town who have heart disease.

Town	Heart.disease	Smoking
A	2.9	69.4
B	3.1	65.7
C	4.1	54.4
D	6.4	65.1
E	6.7	55.9
F	6.8	51.8
G	8.6	53.1
H	8.6	62.8
I	9.6	48.8
J	12.1	35.3
K	15.9	4.8
L	14.2	2.0

a. What is the dependent variable and the independent variables in this study? What would you expect about the relationship between the dependent variable and each of the independent variables?
b. Determine the regression line for the model and the corresponding R^2.
c. Are all independent variables significant to the model? Consider a 95% confidence level.

17. According to Havard University, it has been discovered that the average person in the world uses his/her telephone 20 times per day. The data represents a sample of the number of times/years, that 7 employees in a company check their emails:
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.

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

7540 8421 8560 7412 8953 7859 6098

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

Øvelsestest til Statistik

19. The number of populations in different provinces in a region, are as followed:

7000, 5666, 9877, 1234, 3456, 2345, 6277

Find the mean of the population data, and thereafter find the confidence level interval where you can include the mean number and conclude whether there is a 99% significance from the average population in different provinces.

20. The following data has been given from the ministry of external affairs of India, to check the rising capacity of Delhi Airport.

10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200

Find the median and the 10^{th} percentile of the dataset and the 90^{th} percentile of the dataset.

Prøve 3 - Kort

1. Nazi Germany wants to increase its military budget, but also wants to double the salary income of its army chief. The Germans do not want to ask the army chief directly about the wages. Therefore, the German People look at the income of the three Dictators, which were ruling other states. **Using the following data, use or conduct the appropriate test to see if there were a significant difference between the salaries of the 3 previous military dictators!**

Mussolini	Hitler	Hirohito
10	100	1000
20	120	2000
30	150	3000
40	170	5000
50	200	8000
60	210	10000

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

7540 8421 8560 7412 8953 7859 6098

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

Øvelsestest til Statistik

3. You are a researcher interested in social factors that influence heart disease. You survey 15 towns and gather data on the percentage of people in each town who smoke, the percentage of people in each town who bike to work, and the percentage of people in each town who have heart disease.

Town	Heart.disease	Smoking	Biking
A	2.9	69.4	
B	3.1	65.7	
C	4.1	54.4	
D	6.4	65.1	
E	6.7	55.9	
F	6.8	51.8	
G	8.6	53.1	
H	8.6	62.8	
I	9.6	48.8	
J	12.1	35.3	
K	15.9	4.8	
L	14.2	2.0	

d. What is the dependent variable and the independent variables in this study? What would you expect about the relationship between the dependent variable and each of the independent variables?
e. Determine the regression line for the model and the corresponding R^2.
f. Are all independent variables significant to the model? Consider a 95% confidence level.

4. The External Ministry of Denmark wants to see, the number of travelers going through Billund Airport in Syddanmark. The ministry has sent one of its groups researchers to make research at week 0 and the other research at week 10.

Subject	1	2	3	4	5	6	7	8	9	10
Uge 0	110	120	130	140	150	160	170	180	190	200
Uge 10	210	220	230	240	250	260	270	280	290	300

What is the interval?

Is there a significance difference?

5. A researcher wants to test, if Zink Included in the diet changes the vitamin D inside the body. Six subjects were pretested at Week 0 and they took the zink-medication during 7 weeks. After the 7-weeks period, their vitamin D level was measured again. Using R, can we conclude with 95% confidence level that the cholesterol level has been changed? Assume the variable is approximately normally distributed.

Subject	1	2	3	4	5	6	7
Uge 0	220	244	213	195	177	249	355
Uge 7	189	165	206	193	174	224	325

Øvelsestest til Statistik

What is the interval?

Is it significantly different?

6. The Middle eastern Airlines, such as Emirates, Oman Air and Qatar Airways have been compared to see whether there are significantly different from each other in terms of on time departure and arrival. The following data has been collected. **Use the appropriate test and conclude whether there is a significant difference between the three airlines in terms of on time performance!**

Emirates	Oman Air	Qatar Airways
10	20	30
40	50	60
70	80	90
100	110	120
130	140	150
160	170	180
190	200	210

7. In Denmark, Copenhagen and Aarhus have been through time competing each other for having the status as the financial and cultural capital of Denmark. In this case the following data have been collected on the basis of Economics and Cultural Aspects. **Use the appropriate test and find which city can be called the financial capital or the cultural capital of Denmark.**

X	Copenhagen	Aarhus
Financial Capital	11,12,13,14,15	5,6,7,8,9
Cultural Capital	9,10,11,12,13	16,17,18,19,20

8. A study has been conducted to determine the relationship, between the genetics of Danes and Germans. The following data are shown below:

Danes	20	19	18	17	16	15
German	5	6	7	8	9	10

9. **The number of populations in different provinces in a region, are as followed:**

7000, 5666, 9877, 1234, 3456, 2345, 6277

Find the mean of the population data, and thereafter find the confidence level interval where you can include the mean number and conclude whether there is a 99% significance from the average population in different provinces.

Øvelsestest til Statistik

10. **The number of hours of study of the students of a course and the final grade of the students (out of 100), is shown in the table. Calculate the correlation coefficient and determine whether the correlation is significant. Obtain the regression line.**

Hours of Study	Grade
74	87
59	63
45	50
29	39
20.8	21
19.1	28
13.4	14
8.5	15