

# Mathematics and Statistics Part II

## Statistics

### Introduction to Statistics

### Statistics for European Studies

---

August 12<sup>th</sup> 2024

There are **7** pages inclusive front page and a data file in Excel format.

1. Assigner: Nils Karl Sørensen
2. Hand-out: 10:00
3. Hand-in: 14:00
4. All pages, incl. the front page, should include the following: Exam no.
5. All pages must be numbered.
6. You can hand in: One pdf-document, and several Excel-files.

The internet may only be used to access Digital Exam in order to retrieve and download the exam questions, and to hand-in your exam paper in Digital Exam.

Note that comprehensive explanations are important, when assessing your answers. Assignments that only includes output without comments and theoretical explanations are not accepted.

This present assignment represents half of the full grade in the course “mathematics and statistics”.  
The grade 02 cannot be reached by rounding up. The grade –3 must not appear.

The problems have the following weights:

Problem 1: 30 %  
Problem 2: 15 %  
Problem 3: 15 %  
Problem 4: 20 %  
Problem 5: 20 %

## Problem 1 (30%)

A shop is selling fancy clothing and accessories. Over a period of 30 days the number of customers has been recorded. The folder *Problem 1* in the Excel file *Stat-Ex-08-2024.xlsx* brings the statistics. The statistics is also found in the enclosure to this exercise.

- A. Set up a nice histogram for the distribution of daily shoppers. Characterize the shape of the distribution.
- B. Define and calculate the mean number of daily shoppers per day. Define and find also the median.
- C. Define and find the lower and upper quartile. Denote the Inter Quartile Range
- D. At one day there was a sale, and 100 customers were at the shop. Is this an extreme observation?
- E. Define and find the standard deviation and the coefficient of variation?

Assume that the data set of daily shoppers is Normal Distributed with mean and standard deviation as just found.

- F. What is the probability that there on a random selected day will be between 60 and 70 customers in the shop?

## Problem 2 (15%)

Do females multitask more than males? Maybe – conventional wisdom may be “yes”! However, the concept of multitasking originated in computing and, in that domain; it appears that males are more likely to multitask.

Oxford Internet Surveys asked by phone a sample of 632 males and 756 females if they multitasked while online e.g. listening to music, using the phone or similar. Out of the males 405 said that they multitasked, whereas 446 females said that they multitasked.

- A. Examine a hypothesis stating that the share of males multitasking is higher than the share of females. Set up the hypotheses. Conduct the test and evaluate.
- B. Define the p-value. What is the p-value in the present exercise?

**Problem 3 (15 %)**

According to Eurostat Denmark has highest rate of job shifts in Europe. However, Danes are not very motivated for using many hours every day in order to commute between home and work.

In a questionnaire undertaken for the Danish job webpage “Jobzonen” this issue was examined by gender. A total of 377 persons were asked about their preferences for commuting time. There were 219 males and 158 females.

A total of 146 persons claimed that their maximal total commuting time were less than 30 minutes. There were 69 males and 77 females.

A total of 150 persons claimed that their maximal total commuting time were between 30 and 60 minutes. There were 98 males and 52 females.

A total of 53 persons claimed that their maximal total commuting time were between 1 and 2 hours. There were 32 males and 21 females.

A total of 28 persons claimed that their maximal total commuting time were higher than 2 hours. There were 20 males and 8 females.

- A. Set up a cross-tabulation displaying the information given.
- B. Set up the hypotheses and examine by use of a relevant test if a relation is present between the gender and the preference for commuting time. Comment on the outcome of the test.

### Problem 4 (20%)

In a market survey the price of a specific basket of good were examined in four different cities in Denmark. The results are summarized in the table below. All data are in DKK.

Aarhus	Odense	Copenhagen	Esbjerg
750	800	810	680
780	790	790	740
820	830	870	750
840	850	860	820
790	770	780	730
740	770	800	750

Data in Excel format can also be found in the file *Stat-Ex-08-2024.xlsx* folder *Problem 4*.

- Examine a hypothesis stating that the prices differ by cities. Set up the hypotheses for the test, calculate the ANOVA table. List the assumptions and set up the hypotheses for the test; conduct and explain the test. Provide an interpretation of the outcome of the test.
- If necessary, provide a supplementary analysis to identify the differences. Which city is the most expensive?

### Problem 5 (20%)

The health service of the Danish Region of Southern Denmark has conducted a pilot study in order to clarify the preferences for teaching programs in self-control for breast cancer.

75 females participated in the pilot study. The folder *Problem 5* in the file *Stat-Ex-08-2024.xlsx* contains the data for the pilot study. The statistics is also found in the enclosure to this exercise.

The data set contains the following variables:

<i>Variable</i>	<i>Description</i>
<i>PREF</i>	Index for the preference (utility) of the program (range 0 to 0.5)
<i>INCOME</i>	Household income in 100,000 DKK
<i>AGE</i>	Age in years
<i>ALONE</i>	1 = Living alone, and 0 otherwise
<i>BTIME</i>	Time in seconds it takes to perform the self-control
<i>BCHANGE</i>	1 = Has experienced breast changes before, and 0 otherwise
<i>BREG</i>	1 = Performing regular self-control herself, and 0 otherwise
<i>BCORRECT</i>	1 = Participant can perform the self-control correct, and 0 otherwise

- A. Use Excel to estimate the model for PREF explained by the seven x-variables.  
Provide a nice presentation of the result of the regression output.

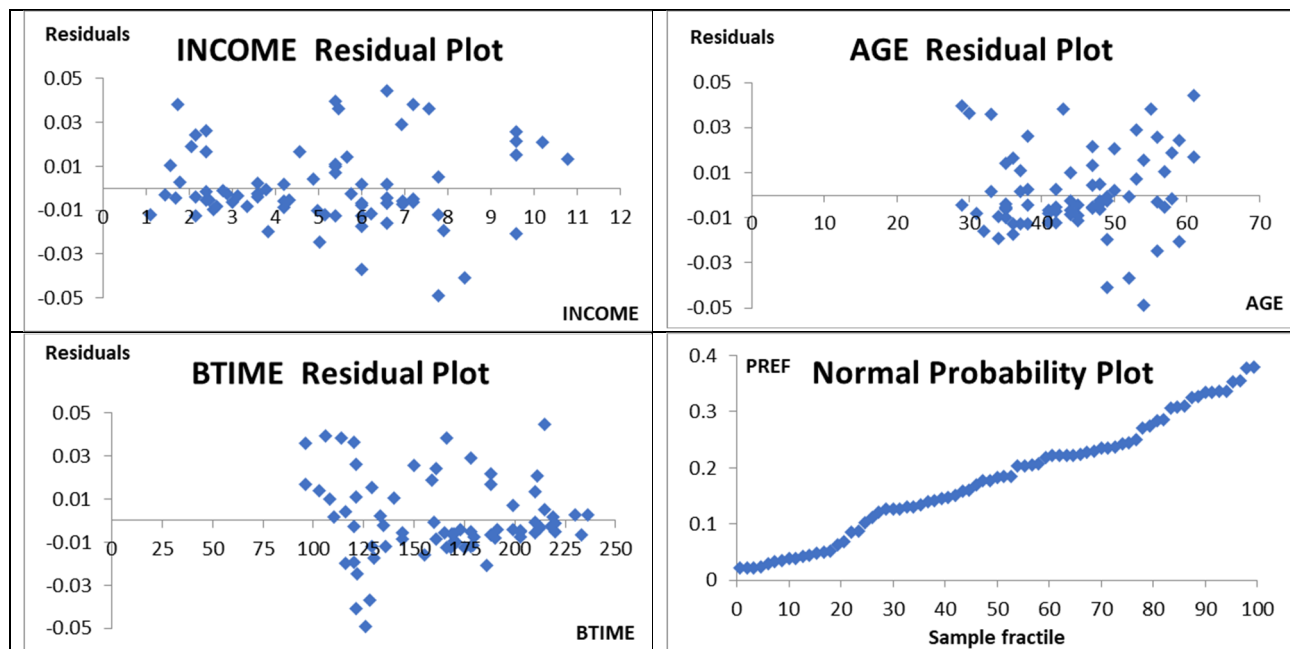
Based on the output that you have provided, answer the following short questions:

- B. Which variables are significant at the 5 percent level?  
C. How is the t-statistic defined?  
D. The variable BCORRECT is negative significant. Is this expected?  
E. What is the R Square equal to, and what does it tell?  
F. What hypothesis is considered in the ANOVA-table in the regression output.  
G. List the assumptions underlying the regression model.

The graphs below show the most interesting plots for model estimated.

- H. Inspect the plots in order to see if the assumptions of the regression model are fulfilled.

## Plots of Residuals and Plot of Normality



## Enclosure

Problem 1: Daily Shoppers at a Fashion Shop						
Daily Shoppers						
Day	Number					
1	67					
2	80					
3	43					
4	57					
5	59					
6	39					
7	28					
8	55					
9	54					
10	80					
11	52					
12	55					
13	62					
14	100					
15	63					
16	79					
17	41					
18	54					
19	63					
20	65					
21	58					
22	41					
23	57					
24	55					
25	65					
26	45					
27	54					
28	57					
29	55					
30	58					

**Problem 5: Data on the preference for the program for self-control for breast cancer.**

PREF	INCOME	AGE	BTIME	ALONE	BCHANGE	BREG	BCORRECT
0.02	5.46	30	120	0	0	1	1
0.22	4.32	47	203	0	0	0	0
0.17	2.64	41	190	1	0	0	0
0.33	9.6	59	186	0	0	0	0
0.23	2.4	58	220	0	1	0	0
0.36	9.6	56	150	0	1	1	0
0.18	6.6	32	155	0	0	1	0
0.06	3.84	49	116	0	0	1	1
0.22	3.6	48	199	1	0	0	0
0.02	1.728	43	114	0	1	1	1
0.22	6.96	42	203	0	0	0	0
0.20	6	36	130	0	0	1	0
0.29	4.2	50	133	1	0	1	0
0.27	3.78	49	160	0	0	1	0
0.18	7.92	34	120	0	0	0	0
0.04	5.4	37	121	0	0	1	1
0.22	2.4	57	178	1	1	0	0
0.31	2.4	61	188	0	0	0	0
0.34	2.16	59	161	1	0	1	0
0.33	2.064	58	159	1	0	1	0
0.31	1.56	57	140	0	0	1	0
0.34	9.6	54	129	0	1	1	0
0.10	7.8	54	126	0	0	1	1
0.28	5.4	53	199	0	0	0	0
0.35	6.93418	53	178	0	0	1	0
0.24	2.796	52	210	0	0	0	0
0.09	6	52	128	0	0	1	1
0.33	10.2	50	211	0	0	0	0
0.09	8.4	49	121	0	0	1	1
0.22	2.88	49	212	1	0	0	0
0.27	7.8	48	215	0	0	0	0
0.05	5.76	48	135	0	1	1	1
0.25	7.2	48	169	0	1	1	0
0.05	6.96	47	144	0	1	1	1
0.34	9.6	47	188	0	0	1	0
0.04	4.896	47	116	0	1	1	1
0.31	10.8	47	210	0	0	0	0
0.19	6.24	45	179	0	1	0	0
0.24	6.6	45	203	0	0	0	0
0.25	4.2	45	144	0	0	1	0
0.04	5.4	44	108	0	1	1	1
0.18	6.6	44	233	0	1	0	0
0.22	7.2	42	220	0	0	0	0
0.24	5.16	42	129	0	0	1	0
0.15	1.10786	41	136	0	1	1	0
0.21	7.8	41	178	0	1	1	0
0.24	6	41	188	0	0	1	0
0.14	2.16	38	191	1	1	1	0
0.18	2.16	38	168	0	0	1	0
0.04	6.6	37	110	0	0	1	1
0.16	5.4	37	166	0	1	1	0
0.03	4.56	36	96	0	0	1	1
0.13	3	35	170	0	1	1	0
0.04	5.66371	35	103	0	0	1	1
0.13	2.46	35	165	0	1	1	0
0.15	4.2	35	210	0	0	0	0
0.14	3.12	35	212	0	0	0	0
0.16	2.568	34	170	1	0	1	0
0.02	7.56	33	96	0	1	1	1
0.02	5.4	29	106	0	0	1	1
0.38	6.6	61	215	0	0	0	0
0.38	7.2	55	166	0	0	1	0
0.21	1.44	56	218	1	1	0	0
0.15	4.98	35	171	0	1	1	0
0.13	1.68	29	173	0	0	1	0
0.12	6	33	219	0	1	0	0
0.11	6	59	118	0	0	1	1
0.05	3.6	44	120	0	0	1	1
0.03	2.4	38	121	0	0	1	1
0.20	6	36	175	0	0	1	0
0.14	6	31	180	0	1	1	0
0.07	5.04	56	122	0	1	1	1
0.13	1.8	42	236	0	1	0	0
0.13	3.6	38	230	0	1	0	0
0.23	3.36	44	161	1	0	1	0