Statistical modelling
1. Midterm test

Name:
NEPTUN:
Tutor:

GROUP A

Total number of points:......................................

Dear Students!

Please write the requested information in the header!

Mark T or F statements clearly in the right place! If the statement is false, you should justify the reason!

In the case of short essay tasks, strive for the essential parts, do not write a long essay.

In the case studies, clearly show which answer belongs to which part of the task. Write all the important answers here, on the paper. The R script should be clear and transparent.

In case of suspected cheating, the test score will be 0 point and an ethics investigation will be proposed.

You have 90 minutes!

Good luck!

                                                                    Laci & Zoli

Statistical modelling
1. Midterm test

Name:
NEPTUN:
Tutor:

1. Theory True or false (6 points): Write one of the true / false answers in the given frame! Each good answer gets 1, every wrong answer gets 0 points! In case of a false answer, please explain why the statement is false. False statement can only be evaluated by giving a correct justification.

1.1	In case of bootstrap we take n-element samples (generally 1000 or 10000) from the population to estimate the confidence interval of a given parameter.	
1.2	The integral of a distribution function is the density function.	
1.3	If we would like to calculate $P(a<X<b)$, the cumulative distribution function measures this probability by the difference between the function values.	
1.4	The likelihood of the whole sample is the probability of all the observed y_i values $(y_1, y_2,...,y_n)$ occurring all at once	
1.5	We have a proportionally stratified sample. The between variance is 0. The estimation of the PS sample is more efficient than the estimation of the random sample without replacement.	
1.6	In case of a one-sample t-test the p-value is 0.03. Our decision could be Type II error, if the significance level is 5%.	

Justification if necessary:

Statistical modelling　　　　　　　**Name:**
1. Midterm test　　　　　　　　　　**NEPTUN:**
　　　　　　　　　　　　　　　　　　Tutor:

2. <u>Short essay questions</u> (4 points)

2.1 Task

How can we decompose the sample variance in case of proportionally stratified sampling? How can we interpret the within standard deviation? (1 point)

2.2 Task

What does the Central Limit Theorem say? (1 point)

2.3 Task

Prove that, the variance of proportion is $\sqrt{\dfrac{p*(1-p)}{n}}$ in case of IID sample! The proportion is the estimator for the parameter of a Bernoulli distribution! (2 points)

3. <u>Short case study</u> (6 points)

a) Generate a population from exponential distribution for which the value of lambda is 0.02! Calculate the theoretical expected value and standard deviation! Compare them to the observed values! (2 points)

b) Take 100 IID samples (with replacement) from the population given in a). The number of observations is 100 in each! Calculate the sample mean for each of the samples! Depict the histogram of the sample means! What is your conclusion? (1.5 points)

c) Calculate the 95% confidence intervals of the means for the samples! (1 point)

d) Calculate the ratio of the samples in which the confidence interval covers the population mean! How does it connect to the confidence level? (1.5 points)

4. Long case study (9 points)

Import the Practice.xlsx file! Change the type of variables, it it is neccessary! It is an IID sample with replacement from the houses in Hungary!

Town	name of the town
Price	1000 HUF
Size	size in square metre
Condition	Condition of the house
Rooms	number of rooms

a) Estimate the 90% confidence interval of the proportion of the houses with excellent condition! (1 point)

b) Estimate the 98% confidence interval of the average price for each town! Depict a bar chart of the mean prices with the confidence intervals! (2 points)

c) Test the following statement: the average price is more than 25 million HUF! (alpha=0.05) Calculate the p-value manually! (2 points)

d) According to a survey, 48% of properties for sale are larger than 100 square metres. Test that fact in this sample! (alpha=0.05) Calculate the p-value manually! (2 points)

e) We assume that, the properties are cheaper on average in Abony compared to Aszód. Test that assumption! (alfa=0.05) Calculate the p-value manually! (2 points)