WESTMINSTER
INTERNATIONAL UNIVERSITY IN TASHKENT
An Accredited Institution of the University of Westminster (UK)

FUNDAMENTALS OF STATISTICS (4ECON006C)

PORTFOLIO OF INDIVIUAL EXERCISES 2 (POE 2)

Semester: Spring 2020

Weight: 10%

Deadline of submission: 17:00 on March 17th, 2020

E-mail: o.isakov@wiut.uz

Office hours: Thursdays, 11:00-13:00

Learning outcomes to be covered

1. Apply a variety of methods for explaining, summarizing and presenting data and interpreting results clearly using appropriate diagrams, titles and labels when required.
2. Explain the fundamentals of statistical inference and be able to apply these principles to justify the use of an appropriate model and perform tests in a number of different settings.
3. Have grounding in probability theory and some grasp of the most common statistical methods and use inference to test the significance of common measures such as means and proportions.

Instructions for submission

1. This is an individual assignment. Students are required to submit their work by 17:00 on March 17th, 2020. Paper version must be submitted to Registrar's Office. Electronic version must be submitted through Intranet (only MS Excel formats are acceptable for electronic version).
2. There are 3 tasks in this assignment. Total score is 50 marks. This assignment has 10% weight on the overall mark for the module.
3. There is Excel data set "housing" that must be used to answer all the tasks. That file has two worksheets: first one for data description and second one for actual dataset.
4. **Paper version** (MS Word document). The students must provide Student ID on the document. Each question and part number (letter) should be written in the answer document. There is no need to provide the questions in the submission form. Excel outputs for relevant exercises must be copied and pasted into the submission (paper) document.
5. **Electronic version** (MS Excel spreadsheet). The file must be named as student's ID (e.g. 1234.xlsx). Electronic version must contain only dataset, appropriate test results and plots. There is no need to provide models, hypotheses, conclusions, etc. (they must be in paper version).

Instructions on how to obtain dataset (Please read carefully)

In order to avoid plagiarism and assessment offence, each student is required to delete two random (read next paragraph on how to generate random numbers) rows from dataset "housing.xlsx" before solving the exercises. That way each student will have his/her own unique dataset (**note that the probability of having the same dataset for two students is (2/57)*(1/57), which is approximately equal to 0.0006**).

Two numbers must be generated using the following Excel function: =randbetween(1,57) or randbetween(1;57) depending on Excel version. Students should run this function twice (more if generated numbers are identical) to get two different random numbers *(e.g. Suppose student obtained 5, 10 as random numbers)*. Then follow the instructions below:

Step 1. Student must delete the row with larger of two numbers. *(Ex. 10th row should be deleted).*

Step 2. The row with smaller number must be deleted next. *(Ex. 5th row).*

Student **is required** to write down these two randomly generated numbers on front page of submission paper.

Failing to follow these instructions may result in mark reduction down to 0.

This is an instruction how to delete two rows for generated random numbers: 5, 10 are used for instructional purposes.

Step 1.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Data Row	Price ($)	Size (m^2)	Rooms	Bathrooms	Region	Structure	Condition	Floor				
2	1	28,500	56	2	1	Chilanzar	Masonry	Good	Middle				
3	2	27,500	50	2	1	Chilanzar	Concrete	Good	First				
4	3	28,999	62	2	1	Sergeli	Concrete	Excellent	Middle				
5	4	19,500	32	1	1	Chilanzar	Concrete	Poor	Top				
6	5	24,500	32	1	1	Chilanzar	Concrete	Excellent	Middle				
7	6	47,999	85	4	1	Chilanzar	Masonry	Good	First				
8	7	42,000	52	2	1	Yunusabad	Concrete	Good	Middle				
9	8	25,500	50	2	1	Chilanzar	Concrete	Good	Top				
10	9	33,500	56	2	1	Chilanzar	Masonry	Good	Top				
11	10	23,500	34	1	1	Chilanzar	Concrete	Excellent	Middle				
12	11	46,498	70	3	1	Chilanzar	Masonry	Excellent	First				
13	12	25,499	52	2	1	Chilanzar	Concrete	Good	Top				
14	13	90,000	72	3	1	Mirabad	Concrete	Excellent	Middle				
15	14	37,000	63	2	1	Mirabad	Concrete	Excellent	First				
16	15	25,000	30	1	1	Chilanzar	Concrete	Excellent	Middle				
17	16	55,000	64	2	1	Mirabad	Concrete	Excellent	Middle				
18	17	33,000	60	2	1	Yunusabad	Concrete	Good	Middle				
19	18	81,000	150	4	2	Yunusabad	Masonry	Excellent	Middle				
20	19	46,000	92	4	1	Yunusabad	Concrete	Excellent	Middle				

Step 1. Delete the 10th row from Dataset rows, **not Spreadsheet's 10th row.**

Step 2.

	A	B	C	D	E	F	G	H	I
1	Data Row	Price ($)	Size (m^2)	Rooms	Bathrooms	Region	Structure	Condition	Floor
2	1	28,500	56	2	1	Chilanzar	Masonry	Good	Middle
3	2	27,500	50	2	1	Chilanzar	Concrete	Good	First
4	3	28,999	62	2	1	Sergeli	Concrete	Excellent	Middle
5	4	19,500	32	1	1	Chilanzar	Concrete	Poor	Top
6	5	24,500	32	1	1	Chilanzar	Concrete	Excellent	Middle
7	6	47,999	85	4	1	Chilanzar	Masonry	Good	First
8	7	42,000	52	2	1	Yunusabad	Concrete	Good	Middle
9	8	25,500	50	2	1	Chilanzar	Concrete	Good	Top
10	9	33,500	56	2	1	Chilanzar	Masonry	Good	Top
11	11	46,498	70	3	1	Chilanzar	Masonry	Excellent	First
12	12	25,499	52	2	1	Chilanzar	Concrete	Good	Top
13	13	90,000	72	3	1	Mirabad	Concrete	Excellent	Middle

Step 2. Delete the 5th row from Dataset rows, **not Spreadsheet's 5th row.**

Tasks

Students are required to create their unique housing dataset by following the instructions given on the previous page. This dataset is used for all the tasks provided below.

1. The response variable is "Price" expressed in US dollars. Use $\alpha = 0.10$ for significance tests. **[20 marks]**

a. Using numerical and graphical summaries, describe the distribution of each quantitative variable. Identify if there are any outliers on dependent variable by plotting the histogram. **[5 marks]**

b. Using numerical or graphical summaries, describe the relationship between each pair of numerical variables. **[3 marks]**

c. Consider fitting a simple linear model using **one** of the following variables as **the only** independent variable: Size or Rooms. **[2 marks]**

d. Run the regression and find the estimated regression coefficients**.[3 marks]**

e. Provide interpretations for regression coefficients in part d. **[3 marks]**

f. Based on regression results in part d, conduct t-test whether there is a significant relationship between Price and independent variable. State your hypothesis, test results and conclusions. **[4 marks]**

2. Consider a statistical model using the following four independent variables: Size, Rooms and pick two more categorical variables. The outcome variable is Price. Use $\alpha = 0.01$ for significance tests. **[20 marks]**

a. Write down the general and fitted linear models. **[3 marks]**

b. Run the regression and provide the estimates for intercept and partial slopes. **[3 marks]**

c. What percent of variability in Price is explained by this model? **[1 mark]**

d. Conduct a test (tests) for overall significance of the model in part a. State the hypotheses, test results and conclusions. **[4 marks]**

e. Conduct a test (tests) for significance(s) of partial slopes of Size and Rooms. State the hypotheses, test results and conclusions. **[8 marks]**

f. Provide the independent variable(s) that is (are) statistically significant in part e. **[1 mark]**

3. Design a Two-Way ANOVA table using two categorical factors from the following variables: District, Structure, Condition, and Floor. Number of replications must be at least 2 (two). Dependent variable is Price. Use $\alpha = 0.10$. **[10 marks]**

a. Provide your designed table. The observations for each treatment groups can be randomly chosen from the dataset at student's discretion. **[3 marks]**

b. State your hypotheses whether selected factors are significant in determining the housing price. **[3 marks]**

c. Provide relevant output table(s) from MS Excel and your conclusions. **[4 marks]**