Econometrics-2022-2023.
Home assignment 1. Simple Linear Regression Model.
To be submitted by October 5, 23:55 (following instructions in HA Rules)
to shakhnazarov.ar@gmail.com

Requirements:

1) Single file in pdf format no more than 10 pages of standard printed text (3000 words) or equivalent legible handwritten text.

2) Use right scan options, check the quality of the file before sending.
3) The first page should be titled like this "Vadim Ivanov, gr. 13, HA01, data07, Eviews12".

1. [30 marks] The 3rd year student of HSE who have just started to study econometrics, is trying to estimate linear regression model describing the work of food deliverers during the COVID-19 epidemic. The model

$$M_i = \beta K_i + u_i, i = 1,\ 2,\ \ldots,\ n.$$

describes the relationship between the number of orders K_i completed by the deliverer per day and money M_i earned. The model does not include intercept as there is no constant part of earnings for delivery. Variable K_i is assumed nonstochastic, disturbance term has zero expectation, constant variance σ_u^2 with different values not correlated each other.

1.1. [10 marks] A fourth-year student who had already studied econometrics advised him to use for estimation coefficient well known estimator $b = \frac{\widehat{\mathrm{Cov}}(M,K)}{\widehat{\mathrm{Var}}(K)}$. Comment on the idea of the fourth year student. What is OLS estimator b_{OLS}^* of β? Explain. Show that b_{OLS}^* is linear and unbiased. Does b also possess these properties?

1.2. [10 marks] Obtain the expression for the variance of b_{OLS}^*. Show that the variance of the estimator b is generally speaking greater than the variance of estimator b_{OLS}^* (except for some special cases). What conclusions about the properties of the estimators under consideration can be drawn from here? Can this result be obtained from the Gauss-Markov theorem from the lecture?

1.3. [10 marks] Assume that under assumptions of model A the number of observations n of the sample tends to infinity. Use sufficient condition for consistency to show that the estimator β_{OLS}^* is consistent.

2. [20 marks]. The researcher estimated (OLS) a relationship between number of students $\boldsymbol{X_i}$ enrolled a certain school and the cost of operating school $\boldsymbol{Y_i}$ for a sample of n schools

$$\boldsymbol{Y_i = \beta_1 + \beta_2 X_i + u_i}\ (1)$$

2.1. [10 marks]. After performing the calculation, he noticed that he forgot to include one more school in the sample, but did not recalculate. Tell the researcher how $\boldsymbol{TSS}$ would change if he did include an additional school in the sample. Give reasons for your judgment. And what about change in $\boldsymbol{R^2}$ before and after adding new observation?

2.2. [10 marks]. The additional school differs significantly from the others in the sample: it has many more students, and the operating cost value for this school is far from the regression line plotted for the rest of the schools. Should a researcher simply throw this school away as an outlier? How do these circumstances affect the validity of Gauss-Markov's conditions for the relationship in question? What consequences could it have? What would you suggest to do to ensure high quality of estimates?

Computer Practice. *Consult **'Distribution of datasets'** to find your personal file.*

3. **[50 marks]** For your data set **ha01_data__** use simple linear regression models to investigate whether earnings of respondents (variable ***EARNINGS***) are significantly dependent on the following variables - ***AGE, ASVABC, HIGHT, HOURS, S, SIBLIBGS, TENURE, WEIGHT, WEXP*** **(**refer to **ha01_data description.pdf).** See detailed questions below.

3.1. [10 marks] **Preliminary analysis of the relationship between earnings and tenure.**

□ Consider data on ***EARNINGS*** and ***TENURE.*** Construct a scatter-diagram for the data under consideration without a regression line and comment on it. Add regression line on the scatter diagram and give additional comments.

□ Build a residual graph. What information can you get out of it?

□ Regress ***EARNINGS*** on ***TENURE*** **– regression (1)**. Present the resulting equation in a standard form (with standard errors under each coefficient and specifying the value of R-squared).

□ Comment on the obtained value of R-square?

□ Give interpretation of coefficients of estimated equation.

□ Estimate the significance of the coefficients and the equation as a whole, based on p-values from the regression printout. Explain the logic of analysis.

3.2. [10 marks] **Detailed analysis of the statistical quality of the regression (1).**

□ Perform statistical tests on significance of regression coefficients using *t*-statistics. Explain the logic of t-test (pair of hypotheses, statistic, critical values, conclusion).

□ Construct confidence intervals for the regression coefficients. Use these confidence intervals to find whether regression coefficients are significant.

□ Perform statistical test on significance of regression using *F*-statistics. Comment on.

□ Compare the conclusions obtained with analysis results in 3.1.

3.3. [10 marks] **Analysis of the relationship between earnings and schooling.**

□ Regress ***EARNINGS*** on ***S*** **– regression (2).** Give interpretations to the coefficients of estimated equation.

□ Generate new variable ***H*** **=** ***S*****–12**, explain its meaning, regress ***EARNINGS*** on ***H*** **– regression (3)** and give interpretations to the coefficients of estimated equation.

□ Compare regressions **(2)** and **(3)** on the base of their printouts: explain mathematically why some values are the same while some others are different.

3.4. [10 marks] **Analysis of the relationship between earnings and other available variables.**

□ Regress *EARNINGS* in turn on ***ASVABC, HOURS, WEXP AGE, SIBLIBGS, HIGHT, WEIGHT*** **(only simple linear regressions both with and without intercept).** Present and discuss briefly **only two or three of them** that seem most interesting to you because they demonstrate some strange or unusual econometric phenomena. Suggest some explanation to the observed phenomena.

3.5. [10 marks] **Relationships between the elements of regression output table.**

□ You are invited to experiment with the regression results. Consider printout table of the regression **(1)** evaluated in 3.1. Delete some data from the table and try to restore it based on the relationships between between different values from the regression table. Try to make your experiment interesting and meaningful.