ÇUKUROVA UNIVERSITY – FACULTY OF ECONOMICS AND ADMINISTRATIVE SCIENCES

2021-2022 ACADEMIC YEAR FALL SEMESTER

ECONOMETRICS I MIDTERM EXAM HOMEWORK ASSIGNMENT

QUESTIONS

PART 1: Fill in the blanks.

1. The sum of the residuals is zero for Classical Linear Regression Models (CLRM). But there is an exception that it is not required to be zero and it is furthermore possible to come across a negative R-squared value for **…………………………………** models.
2. *Other things remaining the same* principle is also called **…………………**
3. Standardized variables have **………** mean and **……….** standard deviation.
4. Total number of observations in the sample less the number of the independent (linear) constraints put on them is called **……………….**
5. The regression line is horizontal to the x-axis if the value of r^2 is equal to **…….**
6. Coefficients of the standardized regression model are also called as **……….** coefficients.
7. Coefficients directly show **…………** values in log-log (i.e. double log) models.
8. Least squares criterion depends on minimizing **……………………….**
9. Inexact relationships between economic variables are covered by **………………** models where the disturbance term also takes place.
10. The theoretical justification for the normality assumption is the **………………………** theorem.
11. In the regression on standardized variables, the intercept term is always **…………**
12. If β_2 and β_3 are the elasticities of the output in terms of labor and capital, the linear restriction $\beta_2 + \beta_3 = 1$ means testing the constant **……………………………**
13. For normally distributed variables, skewness value should be **………** and kurtosis value should be **………**
14. **…………………** data are data on one or more variables collected at the same point in time, such as the census of population conducted by the Census Bureau every 10 years.
15. As the sample size increases indefinitely, the estimators converge to their true population parameters and this property is called **………………**
16. **……………………** theorem is remarkable in that it makes no assumptions about the probability distribution of the random variable u_i, and therefore of Y_i. As long as the assumptions of CLRM are satisfied, the theorem holds.
17. The growth rate of certain economic variables is measured with **……………….** models in which the semielasticity is also defined.
18. Rejecting a true hypothesis is known as **…………………** error.
19. The precision of an estimate is measured by **…………………**

20. There are two conditions for comparing two R-squared values: **……………………………………**

21. Given the assumptions of the CLRM; the least-squares estimators, in the class of unbiased linear estimators, have minimum variance. That is, they are **…….**; and this is known as **………………** theorem.

22. The strength or degree of linear association between two variables is called **…………………………**

23. In regression theory, the **…………….** variable is stochastic; but the **……………….** variables are **…………** or nonstochastic.

24. Asymptotic criteria are also called **……………** sample properties.

25. 1 minus the significance level is known as **…………………………..**

26. A random or **……………** variable is a variable that can take on any set of values, positive or negative, with a given probability (i.e. a variable which has well-defined probabilistic properties).

27. "Linear" regression will always mean a regression which is linear in the **……………………**

PART 2: Answer the following questions.

1. Write the stages of an econometric research.
2. Write the assumptions of the "Classical Linear Regression Model" (CLRM).
3. Explain the justification of adding the error term into a linear regression model.
4. Prove that the least squares estimators may be computed from the sample regression function.
5. Derive the normal equations of least squares (LS).
6. Show that the OLS estimator $b = (X'X)^{-1} X'Y$ is unbiased (i.e. $E(b) = \beta$) using $\mathbf{Y} = \mathbf{X}\beta + \mathbf{u}$.
7. Explain the statement that zero correlation does not necessarily imply independence.
8. Show that the sum of deviation of a variable from its mean value is always zero **(**$\sum x_i = 0$, where $x_i = X_i - \bar{X}$ **)**.
9. State the difference between the population regression function (PRF) and sample regression function (SRF). Which function is not observable?
10. <u>Define the following concepts</u>:

- Estimator
- Estimate
- Residual
- Disturbance (error) term
- Standard error
- Principle of parsimony
- Regression analysis
- Difference between BLUE and BUE
- Standard deviation
- Goodness of fit
- Central limit theorem
- Type 2 error
- Standard error of the regression
- Determination coefficient (R-squared)

PART 3: Solve the following exercises (show your all calculations step by step**).**

1. Based on the following data where Y (dependent variable) represents the number of TV sets sold and X (independent variable) represents the number of sales representatives for totally 10 observations:

X	Y
1	3
1	6
1	10
2	5
2	10
2	12
3	5
3	10
3	10
2	8

a. Estimate the coefficients of the simple linear regression model: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ using normal equations.

b. Estimate the regression coefficients for $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ using the formula for the deviation-from-mean slope estimator (Hint: $\hat{\beta}_1 = \frac{\sum xy}{\sum x^2}$). Show that the results found in both (a) and (b) are the same.

c. Write the resulting estimated regression line and interpret the coefficients.

d. Find and interpret the coefficient of determination.

e. Find the residuals and show that the sum of the residuals is equal to zero.

f. Find the sum of squares for the residuals (RSS).

g. Find the standard error of the regression.

h. Find the standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$.

i. Test the significance of β_0 and β_1 at 95% confidence level (Hint: t-statistic).

j. Find the 95% confidence intervals for β_0 and β_1 (and state your conclusions about the significance of β_0 and β_1). Do the conclusions confirm with the results found in the previous question (i)?

k. Find the 95% confidence interval for the mean prediction for a given value of $X_0 = 5$.

l. Test the significance of the model for 5% significance level (Hint: Set up an ANOVA table).

m. What is the forecast value of the number of TV sets sold when the number of sales representatives is 4?

n. Estimate the regression through the origin model: $\hat{Y} = \hat{\beta}_1 X$.

o. Find the residuals and calculate their sum for the model $\hat{Y} = \hat{\beta}_1 X$.

p. Based on the model $\hat{Y} = \hat{\beta}_1 X$, calculate the standard error of $\hat{\beta}_1$ and test the significance of β_1 using t-statistic for 5% significance level.

q. When compared to the simple linear regression model which includes the intercept term given in Question (a), does this regression through the origin model measure the slope coefficient with a smaller or greater precision? By what criterion can we understand this situation? Explain.

r. Calculate the raw R^2 value.

s. Could we compare this raw R^2 value with the one that is computed in Question (d)?

2. Based on the following dataset (totally 20 observations),

Y	X2	X3
6	1	1
6	2	1
11	2	2
9	1	3
16	3	3
16	1	5
4	1	1
8	3	1
11	2	2
13	3	2
13	1	4
9	1	2
17	3	3
17	2	4
12	4	1
6	1	1
5	1	1
12	3	2
8	1	2
9	2	2

a. Estimate the three-variable linear regression line $\hat{Y} = b_1 + b_2 X_2 + b_3 X_3$ by using $b = (X'X)^{-1} X'Y$ (Show all calculations step by step).

b. Based on the results in (a), find the elasticity of Y with respect to X_2 at the point $X_2 = \bar{X}_2$ (Hint: $\varepsilon = \frac{\partial Y}{\partial X_i} . \frac{X_i}{Y}$).

c. Based on the results in (a), find the elasticity of Y with respect to X_3 at the point $X_3 = \bar{X}_3$.

d. Estimate the three-variable regression line by using "*model in deviation form*" matrix method ($b_* = (X_*' X_*)^{-1} X_*' Y_*$) and verify that the results are the same as the one found in (a).

e. Based on the results in (d); calculate the values for **s.e.** (b_2), **s.e.** (b_3), **cov** (b_2, b_3).

f. Given the information in (e), test the significance of β_2 and β_3 by using t-test.

g. Show the way that you calculate the R^2 value based on "the *model in deviation form* method" and interpret the result.

3. For the linear regression model $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$ [where $u \sim N(0, \sigma^2)$]; define R matrix and r vector which formulate the following null hypotheses in the form of $H_0 : R\beta = r$:

a. $H_0 : 3\beta_4 - 2\beta_1 = 5\beta_3$

b. $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$

c. $H_0 : \beta_1 + \beta_2 + \beta_3 = 1$

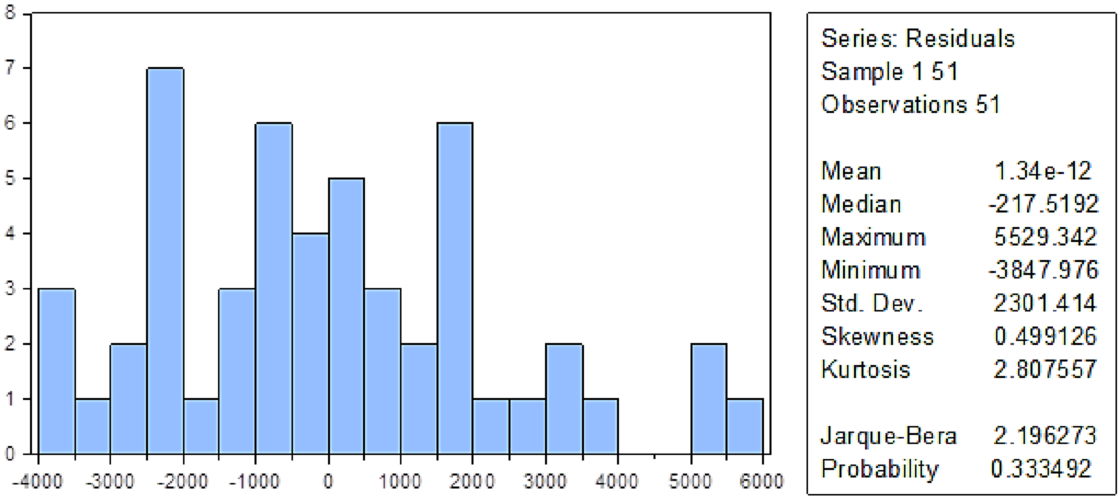
4. In order to estimate the parameters of the linear regression model $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$; data for the sample which includes 10 observations are given as follows:

$$\sum Y = 20,\ \sum X_2 = 30,\ \sum X_3 = 40,\ \sum Y^2 = 88.2,\ \sum X_2^{\ 2} = 92$$
$$\sum X_3^{\ 2} = 163,\ \sum X_2 Y = 59,\ \sum X_3 Y = 88,\ \sum X_2 X_3 = 119$$

Test the linear hypothesis that $H_0 : \beta_2 - \beta_1 = 2$ for the 5% significance level using the *model in deviation form* (Note: Round your results to three decimal places).

Hint:
$$\sum x_i x_j = \sum X_i X_j - n\bar{X}_i \bar{X}_j$$
$$\sum x_i^{\ 2} = \sum X_i^{\ 2} - n\bar{X}_i^{\ 2}$$
$$\sum y^2 = \sum Y^2 - n\bar{Y}^2$$
$$\sum x_i y = \sum X_i Y - n\bar{X}_i \bar{Y}$$

5. Interpret the following histogram plot:



6. Interpret the following regression output where Y shows consumption expenditures and X shows income:

$$\hat{Y} = 18.1581 + 0.5509X$$
$$\text{s.e.} \quad (........) \quad (0.00447)$$
$$t = (13.06037) \; (.......) \qquad r^2 = (........) \qquad \mathbf{d.f}.=92$$
$$\text{prob} \; (0.0000) \; (0.0000)$$
$$F_{1,92} = (.......)$$

a. Fill in the missing numbers and show your calculations.

b. What is the number of observations (i.e. sample size) in this model?

c. Interpret the given regression output (i.e. coefficients, F, R^2 etc.)

d. Is the estimated slope coefficient in accordance with theoretical or prior expectations?

7. Consider the following regression where “Expdur” represents expenditure on durable goods and “t” represents time:

$$\ln Expdur_t = 6.2217 + 0.0154t$$
$$\text{s.e.} \quad (0.0076) \; (0.000554)$$
$$r^2 = 0.9737$$

Calculate the instantaneous and compound rates of growth of expenditure on durable goods. What is the estimated semielasticity, and how would you interpret the slope coefficient in this case?

8. Interpret the following regression output and test the significance of intercept and slope parameters. Also test the hypothesis that $\beta_2 = 1$.

$$\ln \hat{Y} = 1.55476 + 0.89540 \ln X$$
$$\text{s.e.} \ (0.14646) \ (0.00753)$$
$$\mathbf{d.f.} = 92$$

9. Interpret the coefficients of the following regression where the dependent variable is salary (in dollars) and the independent variable is total spending on education:

$$\hat{Y} = -85027 + 13368 \ln X$$
$$t = (-7.3791) \ (9.4972)$$

Good Luck.

Important Guidelines

- **ASSIGNMENT FINAL SUBMISSION DATE: 25.12.2021 - 23:59**

- For exercises, you are required to show all your calculations clearly.
- Your mid-term exam homework should be written by hand. Write all your solutions on blank paper sheets, take their photos and convert into a single PDF file. Do not submit any homework files written by using computer.
- **At the every page of your papers; your name, surname, student number and signature should be available.**
- You do not have to write the questions, write only your answers by stating the question number.
- Upload your pdf file over Microsoft Teams in a timely manner.