

Instruction:

- 1) Use 3 decimal places for all numerical answers.
- 2) Manage your time spending on each question carefully.
- 3) All questions are related to the simple (two-variable) linear regression model for which the population regression equation can be written in conventional notation as:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

where Y_i and X_i are observable variables, β_1 and β_2 are true parameters and u_i is a disturbance term.

The Ordinary Least Squares (OLS) sample regression equation corresponding to the above equation is

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \quad (i = 1, 2, \dots, N)$$

where $\hat{\beta}_1$ is the OLS estimator of the intercept coefficient β_1 , and $\hat{\beta}_2$ is the OLS estimator of the slope coefficient β_2 , $\hat{u}_i$ is the OLS residual for the i -th sample observation.

1. (30 points)

1.1 (15 points) When we adopt the ordinary least square criterion (OLS), why do we choose the criterion of minimizing $\sum \hat{u}_i^2$ instead of the criterion of minimizing $\sum \hat{u}_i$? Logically explain with support of a diagram.

1.2 (15 points) Stating explicitly all the assumptions of the classical linear regression model (CLRM). What happens if the assumptions of CLRM are violated?

2. (20 points)

2.1 (10 points) Why do we need to perform the hypothesis testing of the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$? Give a detailed explanation of applying the t-statistic for the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$. State all assumptions about u_i required for the derivation.

2.2 (10 points) State the concept of the analysis of variance (ANOVA) from the regression viewpoint? Explain fully what its application.

4. **(45 points)** A researcher is using data for a sample of 30 companies to investigate the relationship between annual business taxes paid Y_i (measured in millions of dollars per year) and gross business income X_i (measured in millions of dollars per year). Preliminary analysis of the sample data produces the following sample information:

$$\sum Y_i = 28.730 \quad \sum X_i = 166.045 \quad \sum Y_i^2 = 33.170$$

$$\sum X_i^2 = 1095.526 \quad \sum X_i Y_i = 190.128 \quad \sum x_i y_i = 31.112$$

$$\sum y_i^2 = 5.657 \quad \sum x_i^2 = 176.80 \quad N = 30$$

where $x_i = X_i - \bar{X}$, $y_i = Y_i - \bar{Y}$, and $\hat{y}_i = \hat{Y}_i - \bar{Y}$

Use the above sample information to answer all the following questions. **Show explicitly all formulas and calculations**

4.1 **(5 points)** Use the above information to compute OLS estimates of the intercept coefficient β_1 and that of the slope coefficient β_2 .

4.2 **(5 points)** Interpret the slope coefficient estimate you calculated in part(a)–i.e., explain in words what the numeric value you calculated for $\hat{\beta}_2$ means.

4.3 **(5 points)** Calculate an estimate of σ^2 .

4.4 **(5 points)** Compute the value of r^2 . Briefly explain what the calculated value of r^2 means.

4.5 **(5 points)** Compute the estimated variance of $\hat{\beta}_1$ and the estimated variance of $\hat{\beta}_2$.

4.6 **(5 points)** Compute the two-sided 90% confidence interval for the β_1 . Briefly explain what the two-sided 90% confidence interval means.

4.7 **(5 points)** Perform a test of the null hypothesis $H_0 : \sigma^2 = 0.01$ against the alternative hypothesis $H_1 : \sigma^2 > 0.01$ at the 5 % significance level (i.e., for significance level $\alpha = 0.05$). Show how you calculated the test statistic. State the decision rule you use, and the inference you would draw from the test. Briefly explain what the test outcome means.

4.8 **(10 points)** Perform a test of the null hypothesis $H_0 : \beta_2 = 0$ against the alternative hypothesis $H_1 : \beta_2 \neq 0$ at the 1 % significance level (i.e., for significance level $\alpha = 0.05$). Fill in all information in table1. Show how you calculated the analysis of variance or F test. State the decision rule you use, and the inference you would draw from the test. Briefly explain what the test outcome means.

Table 1. ANOVA Table for the two-variable regression model

Source of variation	Sum of Square SS	df	Mean Sum of Square MSS
Due to regression (ESS)			
Due to residuals (RSS)			
TSS			