PS03

Due date: Friday, June 28th @ 11:59pm

Instructions

Complete the following questions below and show all work. You may either type or hand write your answers. However, you must submit your problem set to Canvas as an html or pdf. Meaning any handwritten solutions are to be scanned and uploaded. The onus is yours to deliver an organized, clear, and/or legible submission of the correct file type that loads properly on Canvas. Do not simply change the file type and submit (e.g. edit name of document from .jpg .pdf) Double-check your submissions! Integrity: If you are suspected of cheating, you will receive a zero—for the assignment and possibly for the course. Cheating includes copying work from your classmates, from the internet, and from previous problem sets. You are encouraged to work with your peers, but everyone must submit their own answers. Remember, the problem sets are designed to help you study for the midterm. By cheating, you do yourself a disservice.

Questions

Q01. Goodness of Fit

Recall the following definitions:

$$TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$ESS = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

$$RSS = \sum_{i=1}^{n} \hat{u}_i^{\,2}$$

(a) In your own words, what do TSS, ESS, and RSS, describe?

(b) Show that $TSS = ESS + RSS$.

Q02. Wage Regression

Consider a dataset obtained from a labor economics study that investigates the impact of years of education on individual's wages. The dataset includes a random sample of workers in a specific region. The following regression equation estimates the relationship between wages (measured in thousands of dollars) and years of education:

$$\text{Wage}_i = \beta_0 + \beta_1 \cdot \text{Eduction}_i + u_i$$

From the regression output, you have the following estimates:

$$\text{Wage}_i = 10 + 2 \cdot \text{Eduction}_i$$

- Standard error of β_1, $\sigma_{\beta_1} = 0.5$
- $R^2 = 0.12$
- Number of observations, $n = 150$

(a) Interpret the Estimates:

Interpret the intercept and slope coefficients in the context of the model.

(b) Predicted Outcome:

If an individual has 12 years of education, what is the predicted wage according to the model?

(c) Effect of Changing :

Suppose an individual worker is deciding whether or not to complete her associates degree (two years). What would the model predict her change in wage would be? In other words, what is her expected increase in wage if she completes her associates degree?

(d) Hypothesis Test:

Conduct a simple hypothesis test to determine if there is a statistically significant relationship between education and wage. State the null and alternative hypothesis and calculate the t-statistic to determine your conclusion. The critical value at the 5% significance level is approximately 1.98. Use this critical value to test the significance of the slope coefficient.

(e) Confidence Intervals

Calculate the 95% confidence interval for β_1. Does your confidence interval satisfy the hypothesis test from part d?

(f) Interpret the R^2:

Explain what the R^2 value tells us about the model's fit to the data.

Q03. R^2

(a) What is R^2?

(b) Why is R^2 bounded between 0 and 1?

Explain using the following equations:

$$R^2 = \frac{ESS}{TSS}$$
$$R^2 = 1 - \frac{RSS}{TSS}$$

Q04. Gauss-Markov Theorem

(a) What does unbiasedness mean? What does efficiency refer to?

Feel free to make a visualization to answer this question.

(b) What does the Gauss-Markov theorem say about unbiasedness and efficiency of the OLS estimator?

Explain in your own words.

(c) In the context of the wage equation, what must we assume to be true regarding the error term u_i for the OLS estimator to be unbiased.

Specifically, I am interested in the third assumption of the classical linear regression model. Give an example of a violation of this assumption in the context of the wage equation.