Midterm Exam

(April 21, 2023, **1 hour 15 minutes**)

Econometrics (Spring 2023)

Professor: Wonmun Shin

Part I

True or False

(Total 30 points) Read each statement below carefully. Write a **T** if you think the statement is True. Write an **F** if you think the statement is false. Do not forget to state a reason (or concise explanation) for your choice.

1. When we set a simple two-variable regression model, $Y_i = \beta_1 + \beta_2 X_i + e_i$, β_1 and β_2 are unknown random variables capturing some characteristics of population. This is the reason we should conduct hypothesis testing after we obtain OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$.

2. Sample regression function (or fitted line) $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ obtained by least square estimation must pass through $(\bar{X}, \bar{Y})$.

3. If error terms e_i are not normally distributed, the OLS estimators are not BLUE.

4. The greater the variance of explanatory variable X_i is, the lower the accuracy of OLS estimator $\hat{\beta}_2$ is.

5. Large size of test means that the estimate is significantly different from zero.

6. We cannot compare the coefficients of determination (R^2) between two models, $Y_i = \beta_1 + \beta_2 X_i + e_i$ and $Y_i = \beta_1 + \beta_2 \ln X_i + e_i$ because regressors take different forms each other.

Part II

Short Questions

1. **(10 points)** In the simple linear regression $Y_i = \beta_1 + \beta_2 X_i + e_i$, we can obtain the OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$. *Student A* decides to use the demeaned values for X_i, considering X_is have large numbers. That is, *Student A* runs the following regression:

$$Y_i = \alpha_1 + \alpha_2 \left(X_i - \bar{X} \right) + u_i$$

where $\bar{X}$ is a mean of X_is. Obtain OLS estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$, and discuss whether they are different from $\hat{\beta}_1$ and $\hat{\beta}_2$ or not.

2. **(10 points)** From a sample of 10 observations, the following results were obtained:

$$\sum X_i = 80, \quad \sum Y_i = 60, \quad \sum X_i Y_i = 660$$

$$\sum X_i^2 = 1,000, \quad \sum Y_i^2 = 610$$

Student B wants to obtain the OLS estimators for a simple regression model, $Y_i = \beta_1 + \beta_2 X_i + e_i$. Using the above results, <u>(i) calculate $\hat{\beta}_1^B$ and $\hat{\beta}_2^B$</u>. Meanwhile, *Student C* re-checks the sample, and finds that one pair of observations is incorrect. Specifically, the last observation (X_{10}, Y_{10}) is not (0, 10) but (20, 0). In other words, *Student C* corrects one pair of observations:

From

X	Y
0	10

To

X	Y
20	0

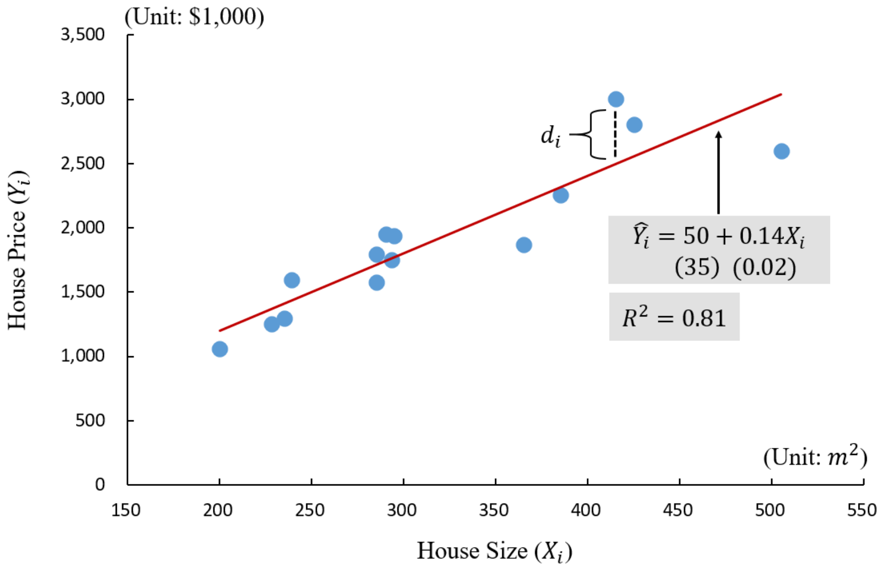
Reflecting the above correction of observations, <u>(ii) calculate the OLS estimators $\hat{\beta}_1^C$ and $\hat{\beta}_2^C$.</u>

3. **(10 points)** Consider a regression model passing through the origin, $Y_i = \beta X_i + e_i$. Let us assume that the regression errors are independent and identically distributed. Specifically, they follow normal distribution, *i.e.* $e_i \sim N\left(0, \sigma^2\right)$. Let $\hat{\beta}$ be the OLS estimator of the slope coefficient. Compute $Bias\left(\hat{\beta}\right)$ and $Var\left(\hat{\beta}\right)$, and discuss efficiency and consistency of $\hat{\beta}$.

Part III

Long Questions

1. **(Total 40 points)** Suppose you want to analyze the effect of the size of houses on house prices, and you run a simple regression $Y_i = \beta_1 + \beta_2 X_i + e_i$ where Y_i is house price and X_i is house size. You have 14 observations where the mean of Y_i $(= \bar{Y})$ is 1,900, the mean of X_i $(= \bar{X})$ is 318, and the variation of Y_i around its mean $(= \sum (Y_i - \bar{Y})^2)$ is 100,000. As a result, you obtained the resulting figure below:



Notice that each dot is each observation of data, the linear line on the plane is the fitted line yielded by OLS methodology. The values in the parentheses of equation represent the standard errors of estimates.

(a) **(4 points)** Interpret the regression result. (*You should use proper units in interpretation!*)

(b) **(4 points)** Is the elasticity of house prices with respect to house size CONSTANT? If yes, what is it?

(c) **(4 points)** Compute the sample correlation coefficient between house price and house size.

(d) **(4 points)** Let us denote d_i as the difference between the actual house price (Y_i) and the fitted house price $(\hat{Y}_i)$. That is, the vertical gap between each dot and the fitted line is d_i (Please refer to the above figure). Then, what is the value of sum of all d_i $(= d_1 + d_2 + \cdots + d_{14})$?

(e) **(4 points)** Let us continue to use the definition of d_i in the above question. Calculate the value of the sum of squares of d_i $(= d_1^2 + d_2^2 + \cdots + d_{14}^2)$.

(f) **(4 points)** Discuss the significance of the slope coefficient $(\hat{\beta}_2)$ at 5% significance level. *(Note: If you need, please refer to the probability table on the last page.)*

(g) **(4 points)** Discuss how to calculate the OLS estimate of variance of regression error $(\hat{\sigma^2})$ using the given information. What is the value of $\hat{\sigma^2}$ in the above regression? *(Tip: Express in the form of fraction.)*

(h) **(4 points)** Is the p-value (for two-sided test) of the intercept coefficient $(\hat{\beta}_1)$ is lower than 0.1 or not? *(Note: If you need, please refer to the probability table on the last page.)*

(i) **(4 points)** Imagine that you want to change the measurement unit of house price from \$1,000 to \$10,000, holding the unit for house size m^2. How do your estimates ($\hat{\beta}_1$ and $\hat{\beta}_2$) change? Present the values of new estimates and interpret the new slope coefficient.

(j) **(4 points)** Imagine that you want to change the measurement unit of house size from m^2 to ft^2 (square feet), holding the unit for house price \$1,000. Note that 1 square feet is equal to 0.1 square meter. How do your estimates ($\hat{\beta}_1$ and $\hat{\beta}_2$) change? Is R^2 in the new regression greater than before, or lower than before?

- **Probability table for Midterm Exam**:
  - $P(F(1,+200) \geq 3.84) = 0.05$, $P(F(1,+200) \geq 6.63) = 0.01$ for F distribution with degrees of freedom $(1,+200)$. ($+200$ indicates for more than 200 degrees of freedom.)
  - $P(F(2,+200) \geq 3.00) = 0.05$, $P(F(2,+200) \geq 4.61) = 0.01$ for F distribution with degrees of freedom $(2,+200)$. ($+200$ indicates for more than 200 degrees of freedom.)
  - $P(F(3,+200) \geq 2.60) = 0.05$, $P(F(3,+200) \geq 3.78) = 0.01$ for F distribution with degrees of freedom $(3,+200)$. ($+200$ indicates for more than 200 degrees of freedom.)
  - $P(Z \geq 1.28) = 0.1$, $P(Z \geq 1.645) = 0.05$, $P(Z \geq 1.96) = 0.025$, $P(Z \geq 2.58) = 0.005$ where Z is a standard normal random variable.

(End of Exam, Total 4 Pages)