EC 320: Introduction to Econometrics

Problem Set 4

Total: 40 points
Due: Friday 21st February, at 5 pm

Learning Outcomes:

- Understanding hypothesis testing

- Understanding multiple regression models

- Understanding goodness of fit

Checklist Before Handing In:

- Did you answer all questions?

- Did you answer all parts for each question?

- Were your answers too vague? If so, make them more precise to make sure they really answer the question being asked.

Instructions: You are encouraged to work with other students in the class, but you must provide original responses. To receive full credit, justify your answers and list your collaborators. For full credit on the computational exercises, include your code and output in addition to your answers. You will turn in digital copies of your responses on Canvas. Please note the list of acceptable file types on the submission page.

Name:

Collaborator 1:

Collaborator 2:

Collaborator 3:

Analytical Questions

1. Suppose that you run a regression of Y_i on X_i with 102 observations and obtain an estimate for the slope (*i.e.*, $\hat{\beta}_2$). Your estimate for the standard error of $\hat{\beta}_2$ is 1. You are considering two different hypothesis tests.

   The first is a **one-sided test**:

   $$\text{H:}_0 \ \beta_2 = 0, \quad \text{H:}_a \ \beta_2 > 0, \quad \alpha = 0.05.$$

   The second is a **two-sided test**:

   $$\text{H:}_0 \ \beta_2 = 0, \quad \text{H:}_a \ \beta_2 \neq 0, \quad \alpha = 0.05.$$

   (a) What values of $\hat{\beta}_2$ would lead you to reject the null hypothesis in the **one-sided** test?

   (b) What values of $\hat{\beta}_2$ would lead you to reject the null hypothesis in the **two-sided** test?

   (c) What values of $\hat{\beta}_2$ would lead you to reject the null hypothesis in the **one-sided** test, **but not** the **two-sided** test?

   (d) What values of $\hat{\beta}_2$ would lead you to reject the null hypothesis in the **two-sided** test, **but not** the **one-sided** test?

2. Suppose that you are studying the effect of police officers on crime rates in several large American cities. When you estimate the model

   $$\text{Crime}_i = \beta_1 + \beta_1 \text{Police}_i + u_i,$$

   you obtain a positive slope estimate.

   (a) What does your slope estimate imply about how the number of police officers affects crime?

   (b) Provide an example of an omitted variable that could explain why the slope estimate is positive.

3. Suppose that your friend wrote a computer program that runs both simple and multiple linear regressions using OLS. Your friend asks you to test the software. After importing a dataset with 2000 observations and three cryptically named variables—Y, X_1, and X_2—you run two regressions. The first is a regression of Y on X_1, which gives you an intercept estimate of 10.4, a slope estimate of -3.8, and $R^2 = 0.215$. The second is a regression of Y on X_1 and X_2, which gives you an intercept estimate of 9.3, an X_1 slope estimate of -2.9, an X_2 slope estimate of -0.07, and $R^2 = 0.198$. You tell your friend that they must made a mistake somewhere in their code. Why?

4. A useful application of multiple regression analysis is *Hedonic modeling.* Hedonic models seek to explain the price of a good—such as a house—in terms of its attributes (*e.g.,* number of bedrooms, square footage, or distance from the nearest toxic waste dump). Consider the following Hedonic model of home sale prices:

$$\text{Price}_i = \beta_0 + \beta_1(\text{Square footage})_i + \beta_2\text{Bathrooms}_i + \beta_3\text{Bedrooms}_i + u_i.$$

Using data from 37 home sales, you estimate the model and obtain $\hat{\beta}_0 = 90000$, $\hat{\beta}_1 = 1100$, $\hat{\beta}_2 = 16000$, $\hat{\beta}_3 = 35000$, $\text{SE}(\hat{\beta}_1) = 650$.

(a) Interpret each coefficient.

(b) What is the model's forecasted sale price for a 2500-square-foot home with 3 bedrooms and 2.5 bathrooms?

(c) In a remodeling frenzy, a homeowner adds an additional bedroom and an additional bathroom by splitting up existing rooms. What is the forecasted change in the price of her home?

(d) A homeowner adds a 450-square-foot bedroom and a 75-square-foot bathroom by extending the footprint of his home into an area that used to be a driveway. What is the forecasted change in the price of his home?

(e) Conduct two-sided tests of the hypothesis that square footage has no effect on sale price at the 10, 5, and 1 percent levels.

(f) Construct a 95 percent confidence interval for β_1.

Computational Problems

For this portion of the problem set, you will use the `apple.csv` file in the `Problem Set 4` folder on Canvas. The file contains data from an experimental survey. The survey presented participants with **randomly determined** prices for "eco-labeled" apples and regular apples and then asked how many eco-labeled and regular apples they would buy at those prices. For reference, eco-labeling helps consumers identify sustainably-produced (or "green") products and helps firms command higher prices for their products. You will estimate the demand for eco-labeled and regular apples by running regressions of apple quantity on prices. The fact that the prices were randomly assigned means that the exogeneity assumption holds—so long as both prices are included in the model. To complete this assignment, you will need to load the `tidyverse`, `stargazer`, and `broom`.

Variable Name	Description
reglbs	Pounds of regular apples demanded
ecolbs	Pounds of eco-labeled apples demanded
regprc	Price of regular apples (per pound)
ecoprc	Price of eco-labeled price (per pound)

1. Run a regression of `reglbs` on `regprc`. Interpret the slope coefficient. Is the sign of the slope consistent with what you know about demand curves?

2. Run a regression of `ecolbs` on `ecoprc`. Interpret the intercept coefficient.

3. Run a regression of `reglbs` on `regprc` and `ecoprc`. How does the estimated coefficient on `regprc` change? What does this tell you about the correlation between `regprc` and `ecoprc`? Justify your answer and then use `R` to verify.

4. Run a regression of `ecolbs` on `regprc` and `ecoprc`. Identify and interpret R^2.

5. Summarize your regression results in a table. Which two of the four regressions have the highest R^2? Why?

6. Construct a 99 percent confidence interval for the `ecoprc` coefficient from the regression described in exercise 4.