**Econ 11020: Introduction to Econometrics**

**Winter 2023 – Instructor: Oscar Galvez-Soriano**

**Problem Set 1**

Deadline: 10/10/2023, 4:30pm Central Time during the discussion section.

Make sure that your answers are clear and easily readable.

In this problem set you will run regressions “by hand”. Later you will learn how to use statistical software for this purpose.

1. Consider the following population model:
2. Explain what each term in this model represents. You may use an example to illustrate your answer
3. If we estimate the following sample model: , how and are different from and ? (Hint: which one is a constant number and which one a random variable)
4. Given a random sample , we can produce a fitted regression line based on this cloud of points. Show that minimizing the sum of squared vertical distances between the points and the line gives the formulas in class:

You might use the following steps[[1]](#footnote-1):

1. Start by stating the minimization problem
2. Then calculate the FOCs like we did in class. You should get and
3. Work with the first FOC to obtain the formula for the intercept
4. Work with the second FOC (you can get rid of the -2), distribute the product and notice that and
5. Solve for and notice that and that . Make sure that you can derive these equalities yourself
6. Suppose you estimate a model as in question 1(b), using a regression with observations. If you are given a value of the regressor as (that did not appear in any observation in your original sample), what would be your prediction for the outcome ? Present your answer algebraically and graphically.
7. Show that the FOC with respect to the intercept parameter implies that the sample regression line passes through the point . Give an intuitive explanation for this feature of OLS.
8. Assume you have the following dataset:

|  |  |  |
| --- | --- | --- |
| **Observation** |  |  |
| i=1 | 3 | 1 |
| i=2 | 5 | 1 |
| i=3 | 7 | 2 |

1. Run the regression of y on x by computing (manually) the OLS estimates and ) for the parameters and .
2. Fill out the table below, based on your estimates from part (a)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Observation** |  |  |  |  |
| i=1 | 3 | 1 |  |  |
| i=2 | 5 | 1 |  |  |
| i=3 | 7 | 2 |  |  |

1. Based on the table in (b), draw a scatter plot and add the sample regression line. Label the fitted values and residuals.
2. Suppose denotes a person’s hourly wages in dollars, and their education in years. Using the formulas in question 2, explain how your estimates  and   would change if:
3. Wages are measured in cents instead of dollars
4. Education is measured in months instead of years.
5. What did you learn from this question?

1. For better learning: try to understand every step of this proof and be able to reproduce it without needing to check your lecture notes. [↑](#footnote-ref-1)