Introductory Econometrics
Lecture 2: Introduction to linear regression model
Suggested Solution

by Hieu Nguyen

Fall 2024

1.

Derive the 'summation formula' for the OLS estimator $\hat{\beta}_1$ in the linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.$$

2.

Suppose you are a university director considering canceling the entrance exams and searching for new criteria for selecting good students for your school. Assume you consider using the grades from high school (in the US called the American College Testing [ACT] scores, 1–36).

To decide, some data about current students seem to be helpful. Let us say you have a sample of 15 students. The following table contains the ACT scores and the GPA (Grade Point Average, 0–4) for these 15 students. Grade Point Average represents the student's performance at the university; it is based on a four-point scale and has been rounded to one digit after the decimal.

Student	GPA	ACT
1	2.8	21
2	3.4	24
3	3.0	26
4	3.5	27
5	3.6	29
6	3.0	25
7	2.7	25
8	3.7	30
9	3.2	23
10	3.0	28
11	3.5	30
12	2.5	20
13	3.8	32
14	2.6	26
15	2.7	23

(a)

Estimate the relationship between GPA and ACT using OLS; that is, obtain the intercept and slope estimates in the equation: $\text{GPA} = \beta_0 + \beta_1 \text{ACT} + \varepsilon$. Use Excel for the computation:

1. Compute the intercept and slope estimates using the summation formulas for $\hat{\beta}_0$ and $\hat{\beta}_1$.
2. Compute the intercept and slope estimates using the matrix formula for $\hat{\beta}$.

(b)

Comment on the direction of the relationship. Does the intercept have a meaningful interpretation here? Explain. How much higher is the GPA predicted if the ACT scores increase by 5 points?

(c)

Find and list the fitted values and the residuals of the model.

(d)

What is the predicted value of GPA when ACT = 20?

3. BONUS EXERCISE (very similar to 2.) to take home and master OLS in Excel:

This exercise illustrates the crucial distinction between the stochastic error term and the residual. Usually, we can never observe the error term, but we can get

around this difficulty if we assume values for the true parameters. Calculate values of the error term and the residual for each of the following six observations given that the true β_0 equals 0, the true β_1 equals 1.5.

	y_i	x_i
1	2	1
2	6	4
3	3	2
4	8	5
5	5	3
6	4	4