# Week 1 - Introduction to Regression and Simple Linear Regression

Written by /u/econpanda

Problems with a * are not necessary but may provide additional insight. The readings for this problem set are

- Chapter 1
- 2.1, 2.2, 2.4, 2.6

Pay attention to the following key topics

- Meaning of **ceteris paribus**
- Examples 1.3, 1.4, 1.5, 1.6
- Problems with nonrandom assignment (pages 14-15)
- What the regression error term $u$ captures
- Assumptions about relation between $x$ and $u$
- Estimation of regression coefficients
- Interpretation of regression coefficients
- How $\log(\cdot)$ changes interpretation of regression coefficients (Table 2.3)
- Regression assumptions (section 2.3) and properties of OLS estimators (section 2.5) will be covered next week

1. (Wooldridge 1.1) Suppose that you are asked to conduct a study to determine whether smaller class sizes lead to improved performance of fourth graders.[1]

   (a) If you could conduct any experiment you want, what would you do?

   **Solution:** Ideally, we could randomly assign students to classes of different sizes. That is, each student is assigned a different class size without regard to any student characteristics sich as ability and family background.

   (b) More realistically, suppose you can collect observational data on several thousand fourth graders in a given state. You can obtain the size of their fourth-grade class and a standardized test score taken at the end of fourth grade. Why might you expect a negative correlation between class size and test score?

[1] For a good answer to this question see the Tennessee STAR experiment and Krueger (1999)

**Solution:** A negative correlation means that larger class size is associated with lower performance. We might find a negative correlation because larger class size actually hurts performance. However, with observational data, there are other reasons we might find a negative relationship. For example, children from more affluent families might be more likely to attend schools with smaller class sizes, and affluent children generally score better on standardized tests. Another possibility is that, within a school, a principal might assign the better students to smaller classes. Or, some parents might insist their children are in the smaller classes,

(c) Would a negative correlation necessarily show that smaller class sizes cause better performance?

**Solution:** Given the potential for confounding factors - some of which are listed in (b) - finding a negative correlation would not be strong evidence that smaller class sizes actually lead to better performance. Some way of controlling for the confounding factors is needed, and this is the subject of multiple regression analysis.

2. (Wooldridge 2.1) Let *kids* denote the number of children born to a woman, and let *educ* denote years of education for the woman. A simple model relating education to fertility to years of education is

$$kids = \beta_0 + \beta_1 educ + u$$

(a) List 5 specific variables that are in $u$

**Solution:** Income, age, and family background (such as number of siblings) are just a few possibilities.

(b) Are any of these things likely to be correlated with *educ*?

**Solution:** It seems that each of these could be correlated with years of education. (Income and education are probably positively correlated; age and education may be negatively correlated because women in more recent cohorts have, on average, more education; and number of siblings and education are probably negatively correlated.)

(c) Would this simple regression uncover the ceteris paribus effect of education on fertility (Is $E(u|x)$ likely to hold)?

**Solution:** Not if the factors we listed in part (i) are correlated with educ. Because we would like to hold these factors fixed, they are part of the error term.

3. You are interested in finding the relation between time allowed for college students to take an exam and their performance on the exam. You notice that at your university a class is offered on MWF (for 50 minutes) and on TTH (for 75 minutes) and it is taught by the same professor that uses the same exam.

(a) Is a simple regression of time on test score likely to uncover a ceteris paribus effect of time on test score (Hint: Are students randomly assigned between classes? Is $E(u|x)$ likely to hold)?

> **Solution:** It is unlikely that $E(u|x)$ holds if students are able to select which class they enroll it. That is, people with a higher ability may self select into one section.

(b) Alternatively you can convince the professor to pool the sections for exams and flip a coin for each student to determine their time allotment, heads means they get 75 minutes and tails means they get 50 minutes. Would this approach uncover a ceteris paribus effect?

> **Solution:** This approach would make it more likely for $E(u|x)$ to hold. Since time allowed for the exam is randomly assigned it should be uncorrelated with other factors that impact a students exam score (ability, time spent studying, etc. )

4. * Wooldridge derives OLS through the method of moments estimator, an alternative way to estimate $\hat{\beta}_0$ and $\hat{\beta}_1$ is through minimizing the sum of squared residuals. Define the residuals as $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1}$, the objective function is:

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{n} \hat{u}^2 = \min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1})^2$$

(a) Show that the derivatives of this function with respect to $\beta_0$ and $\beta_1$ are[2]

$$-2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1}) \tag{1}$$

$$-2\sum_{i=1}^{n}x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1}) \tag{2}$$

> **Solution:** This is a straight forward application of the chain rule and the power rule, for completeness I will state it here. The chain rule thats that if we have two functions $f(x)$ and $g(x)$ then the derivative of $f(g(x))$ with respect to $x$ is $f'(g(x)) \times g'(x)$. The power rule states that the derivitave of $x^n$ with respect to $x$ is $n \times x^{n-1}$. Combining these two yields the desired result.

(b) Set these equations equal to 0 and solve for $\hat{\beta}_0$ and $\hat{\beta}_1$ and show that they are equivalent to equations 2.17 and 2.19 from Wooldridge. You will need the following properties

$$\sum_{i=1}^{n} x_i(x_i - \bar{x}) = \sum_{i=1}^{n}(x_i - \bar{x})^2 \text{ and } \sum_{i=1}^{n} x_i(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

[2]If you need a review of calculus and summation operators see Appendix A in Wooldridge

**Solution:** First from equation (1) we can drop the $-2$ and "distribute" the summation operator to get

$$\begin{aligned}
&\sum y_i - \hat{\beta}_0 \sum 1 - \hat{\beta}_1 \sum x_i = 0 \\
&n\bar{y} - n\hat{\beta}_0 - \hat{\beta}_1 n\bar{x} = 0 \\
&\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}
\end{aligned}$$

where $\sum y_i = n\bar{y}$ is true by definition of the sample mean. From equation (2) we have

$$\begin{aligned}
&\sum x_i y_i - \sum \hat{\beta}_0 x_i - \hat{\beta}_1 \sum x_i^2 = 0 \\
&\sum x_i y_i - \sum \bar{y} x_i + \hat{\beta}_1 \sum \bar{x} x_i - \hat{\beta}_1 \sum x_i^2 = 0 \\
&\sum x_i y_i - \sum \bar{y} x_i = \hat{\beta}_1 \left( \sum x_i^2 - \sum \bar{x} x_i \right) \\
&\sum x_i (y_i - \bar{y}) = \hat{\beta}_1 \sum \left( x_i (x_i - \bar{x}) \right) \\
&\sum (x_i - \bar{x})(y_i - \bar{y}) = \hat{\beta}_1 \sum (x_i - \bar{x})^2 \\
&\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}
\end{aligned}$$

5. (Wooldridge 2.2) In the simple linear regression model $y = \beta_0 + \beta_1 x_1 + u$, suppose that $E(u) = \alpha_0 \neq 0$. Show that the model can always be written with the same slope, but a new intercept and new error, where the new error has zero mean.

**Solution:** Add and subtract $\alpha_0$ from the right hand side to get

$$\begin{aligned}
y &= \beta_0 + \beta_1 x_1 + u + \alpha_0 - \alpha_0 \\
&= (\beta_0 + \alpha_0) + \beta_1 x_1 + (u - \alpha_0) \\
&= \beta_0^* + \beta_1 x_1 + u^*
\end{aligned}$$

Where $\beta_0^* = \beta_0 + \alpha_0$ and $u^* = u - \alpha_0$. So $E(u^*) = E(u - \alpha_0) = E(u) - \alpha_0 = \alpha_0 - \alpha_0 = 0$

6. For each of the following regressions on the relation between a persons high school GPA and the ACT score provide a general interpretation of $\beta_1$

   (a) $GPA = \beta_0 + \beta_1 ACT + u$

   **Solution:** If a persons ACT score increases by 1 point we expect their GPA to increase by $\beta_1$ points

   (b) $\log(GPA) = \beta_0 + \beta_1 ACT + u$

**Solution:** If a persons ACT score increase by 1 point we expect their GPA to increase by $100 \times \beta_1$ percent

(c) $GPA = \beta_0 + \beta_1 \log(ACT) + u$

**Solution:** If a persons ACT score increases by 1 percent we expect their GPA to increase by $\beta_1/100$ points

(d) $\log(GPA) = \beta_0 + \beta_1 \log(ACT) + u$

**Solution:** If a persons ACT score increases by 1 percent we expect their GPA to increase by $\beta_1$ percent

# References

Krueger, A. B. (1999). Experimental estimates of education production functions*. *The Quarterly journal of Economics 114*(2), 497–532.