Econ 103: Introduction to Econometrics
Week 2

Francisco Díaz-Valdés*

Fall 2025

1 Introduction to Linear Regression

As economists, we are frequently interested in the relationship between two economic variables X and Y, say $X =$ years of schooling and $Y =$ wages. First, we should investigate the summary statistics we reviewed in the first week, for example, mean, variance, and correlation. However, focusing on the summary statistics won't get us far in questions regarding prediction and causality. That's when we need an economic model. As a starting point, we will focus on a **linear model**; we model the relationship between X and Y as linear. Suppose we are given an sample $\{X_i, Y_i\}$ for $i = 1, \ldots, n$ of n observations, and we have the model

$$\underbrace{Y_i}_{\text{dependent variable}} = \underbrace{\beta_1}_{\text{intercept}} + \underbrace{\beta_2}_{\text{slope}} \times \underbrace{X_i}_{\text{independent variable}} + \underbrace{e_i}_{\text{error/residual term}}$$

Where the e_i is the error term. Here, we assume that this model is the **true** relationship between X and Y, and we are interested in the **parameters** β_1 and β_2. These parameters are **unknown**.

- Why do we need this model?
  - Econometric methodology examines and analyzes a **sample of data** from the **population**. After analyzing the data, we make statistical inferences. These are conclusions about a population based on the data analysis. Great care must be taken when drawing inferences. The inferences are conclusions about the population from which the data were collected.
  - **Prediction:** Predicting the value of one variable given the value of another, or others, is one of the primary uses of regression analysis.
  - **Causality:** A second primary use of regression analysis is to attribute or relate changes in one variable to changes in another variable.

1.1 Assumptions of the Simple Linear Regression Model

- **SR1 Econometric Model:** All data pairs (y_i, x_i) collected from a population satisfy the relationship

$$Y_i = \beta_1 + \beta_2 X_i + e_i, \ i = 1, ..., N$$

- **SR2 Zero mean error :** $\mathbb{E}[e_i] = 0$ for any i. This assumption is just a normalization.
- **SR3 Constant X :** $\mathbb{P}(X_i = x_i) = 1$ for any i. It helps when learning regression for the first time, but it can be substituted by the exogeneity assumption $\mathbb{E}[e_i \mid X_i] = 0$ for any i.

*Many thanks to all previous TAs for providing the notes. All mistakes are my own. Please contact me at fdiazvaldes@g.ucla.edu if you spot any typos or mistakes. Some parts of this note rely on https://scpoecon.github.io/ScPoEconometrics.

- **SR4 Homoscedasticity:** $\mathbb{V}(e_i) = \sigma^2$ for any i. If X is not constant, this assumption becomes $\mathbb{V}(e_i \mid X = x) = \sigma^2$ for any x. It is only important for the BLUE result and to simplify the exposition of the material. In practice, no researcher assumes that this assumption holds
- **SR5 Uncorrelated errors:** $\mathbb{C}(e_i, e_j) = 0$ for any $i \neq j$. Unobserved variables are uncorrelated.
- **SR6 Normal errors:** $e_i \sim \mathcal{N}(0, \sigma^2)$ for any i. Only to simplify the exposition of the material. Versions of the CLT will guarantee that this is true asymptotically (that is when the sample size tends to infinity).

1.2 Interpretation of Parameters in Linear Models

How do we interpret β_2 in the model $Y_i = \beta_1 + \beta_2 X_i + e_i$? Note that under assumptions

$$E[Y|X = x] = \beta_1 + \beta_2 x$$

which implies

$$\beta_2 = \frac{\partial E[Y|X = x]}{\partial x}$$

On average, one unit increase in X is associated with β_2 units change in Y.

1.3 Estimating the Model with OLS

One common method for estimating β_1 and β_2 is called Ordinary Least Squares, or **OLS**. This method aims to **minimize the sum of squared residuals**.
Let b_1 and b_2 be arbitrary constants used as estimates for β_1 and β_2, respectively. We can rearrange the model given above to calculate the residual (or error term) for each observation, (Y_i, X_i).[1]

$$e_i = Y_i - (b_1 + b_2 X_i)$$

Summing up the squared error for each observation gives us the **sum of squared residuals/errors (SSE)**, conditional on our arbitrary choice of b_1 and b_2:

$$SSE(b_1, b_2) = \sum_{i=1}^{n} (Y_i - (b_1 + b_2 X_i))^2$$

Our job as econometricians is to find the best estimators of β_1 and β_2. Under the OLS methodology, this means we need b_1 and b_2 such that the SSE is **minimized**. We denote the estimator with the "hat" symbol as a mathematical convention.

$$(\hat{b}_1, \hat{b}_2) = \arg\min_{b_1, b_2} \sum_{i=1}^{n} (Y_i - b_1 - b_2 X_i)^2$$

This line should be read as "$\hat{b}_1$ and $\hat{b}_2$ are the estimators that minimize the SSE among all possible values b_1 and b_2". (arg min denotes the values/arguments that minimize the object.) So we know that

$$SSE(\hat{b}_1, \hat{b}_2) \leq SSE(b_1, b_2) \quad \text{for all } b_1, b_2$$

1.4 Some Terminology

- **Estimators** are formulas/functions of random variables, so they are random. In the section above, we solved for $\hat{b}_1$ and $\hat{b}_2$, which are estimators for β_1 and β_2.

[1] Note that there will always be some error unless the data points form an exact straight line. If that is the case, we can solve for β_1 and β_2 using any two points and do not need more complicated econometrics.

- **Estimates** are **realizations** of the estimator that depend on the data a researcher uses. For example, once you obtain some data $\{X_i = x_i, Y_i = y_i\}_{i=1}^n$, you can plug them into the formulas for $\hat{b}_1$ and $\hat{b}_2$ to get estimates - these values, unlike the estimator, are not random.
- The **fitted line** is given by $y = \hat{b}_1 + \hat{b}_2 x$.
- The **fitted value**, $\hat{Y}_i$ is given by $\hat{Y}_i = \hat{b}_1 + \hat{b}_2 X_i$.
- The difference between the true value, Y_i, and the fitted value, $\hat{Y}_i$, is the **fitted residual**

$$\hat{e}_i = Y_i - \hat{Y}_i = Y_i - \underbrace{(\hat{b}_1 + \hat{b}_2 X_i)}_{\hat{Y}_i}$$

1.5 Estimators are BLUE

Gauss-Markov Theorem If the assumptions SR1 to SR5 are true, then estimators $\hat{b}_1$ and $\hat{b}_2$ are the best linear unbiased estimators. Best means that the variance of these estimators is smaller than any other linear and unbiased estimators. Linear comes from the fact that one can rewrite $\hat{b}_1$ and $\hat{b}_2$ as a linear function of $(y_1, \ldots, y_n)$.

2 Practice Problems

2.1 Problem 1: demeaned variables and estimation without the regressor

Suppose you have the following sample $\{Y_i, X_i\}_{i=1}^n$ and you are interested in the following linear model:

$$Y_i = \beta_1 + \beta_2 X_i + e_i \tag{1}$$

Let's denote $\overline{Y} = \frac{1}{n}\sum_i Y_i$ and $\overline{X} = \frac{1}{n}\sum_i X_i$ the average of the dependent variable and independent variable, respectively.

1. Denote by $\tilde{Y}_i = Y_i - \overline{Y}$ and $\tilde{X}_i = X_i - \overline{X}$ the demeaned variables. Starting from the functional form (1), propose a model that employs $\{\tilde{Y}_i, \tilde{X}_i\}_{i=1}^n$ to estimate β_2.

   Solution:
   We know that:

   $$Y_i = \beta_1 + \beta_2 X_i + e_i$$

   Then, sum over i and divide by n.

   $$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + e_i \\ \Rightarrow \underbrace{\frac{1}{n}\sum_i Y_i}_{=\overline{Y}} &= \beta_1 + \beta_2 \underbrace{\frac{1}{n}\sum_i X_i}_{=\overline{X}} + \underbrace{\frac{1}{n}\sum_i e_i}_{=\overline{e}} \end{aligned}$$

   So we have:

   $$\overline{Y} = \beta_1 + \beta_2 \overline{X} + \overline{e} \tag{2}$$

   Subtracting the equation (1) with (2) we get:

   $$\underbrace{Y_i - \overline{Y}}_{=\tilde{Y}} = \underbrace{(\beta_1 - \beta_1)}_{=0} + \beta_2 \underbrace{(\overline{X} - X_i)}_{=\tilde{X}} + \underbrace{e_i - \overline{e}}_{=\tilde{e}}$$

Therefore, we will estimate β_2 using the following equation:

$$\tilde{Y}_i = \beta_2 \tilde{X}_i + \tilde{e}_i$$

So, we need to solve:

$$\min_{b_2} \sum_{i=1}^{n} (\tilde{Y}_i - b_2 \tilde{X}_i)^2$$

The first-order condition is:

$$\begin{aligned} & \frac{\partial}{\partial b_2} \left[\sum_{i=1}^{n} (\tilde{Y}_i - b_2 \tilde{X}_i)^2 \right] = 0 \\ \Leftrightarrow \quad & 2 \sum_{i=1}^{n} (\tilde{Y}_i - \hat{b}_2 \tilde{X}_i)(-\tilde{X}_i) = 0 \\ \Leftrightarrow \quad & \hat{b}_2 = \frac{\sum_{i=1}^{n} \tilde{Y}_i \tilde{X}_i}{\sum_{i=1}^{n} \tilde{X}_i^2} \end{aligned}$$

2. Show that the coefficient $\hat{b}_2$ can be expressed as:

$$\hat{b}_2 = \frac{\mathbb{C}(X_i, Y_i)}{\mathbb{V}(X_i)}$$

Solution:
Using our estimated coefficient for b_2:

$$\begin{aligned} & \hat{b}_2 = \frac{\sum_{i=1}^{n} \tilde{Y}_i \tilde{X}_i}{\sum_{i=1}^{n} \tilde{X}_i^2} \\ \Leftrightarrow \quad & \hat{b}_2 = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})(X_i - \overline{X})}{\sum_{i=1} (X - \overline{X})^2} \\ \Leftrightarrow \quad & \hat{b}_2 = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})(X_i - \overline{X})}{\sum_{i=1} (X - \overline{X})^2} \frac{(1/n)}{(1/n)} \\ \Leftrightarrow \quad & \hat{b}_2 = \frac{\mathbb{C}(X_i, Y_i)}{\mathbb{V}(X_i)} \end{aligned}$$

3. Using this framework, how do you recover the estimated parameter for b_1 without minimizing again? You can assume that the sample is large enough so averages approximate expectations quite well.

Solution:
We know that:

$$\overline{Y} = b_1 + b_2 \overline{X} + \underbrace{\overline{e}}_{\mathbb{E}[e_i]=0}$$

Therefore,

$$\begin{aligned} & \overline{Y} = \hat{b}_1 + \hat{b}_2 \overline{X} \\ \Leftrightarrow \quad & \hat{b}_1 = \overline{Y} - \hat{b}_2 \overline{X} \end{aligned}$$

4. Now, suppose that you face the following linear model without an independent variable, just the constant:

$$Y_i = b_1 + e_i$$

Find the expression of the estimator for b_1. Then, give an intuitive explanation of your result. Hint: A graphical illustration can help visualize intuition.

Solution:
The error can be written as $e_i = Y_i - b_1$. Hence, the minimization problem is:

$$\min_{b_1} \sum_{i=1}^{n} (Y_i - b_1)^2$$

The FOC is:

$$\begin{aligned} & \frac{\partial}{\partial b_1} \left[\sum_{i=1} (Y_i - b_1)^2 \right] = 0 \\ \Rightarrow \quad & 2 \sum_{i=1}^{n} (Y_i - \hat{b}_1)(-1) = 0 \\ \Leftrightarrow \quad & \hat{b}_1 = \frac{1}{n} \sum_{i=1}^{n} Y_i = \overline{Y} \end{aligned}$$

2.2 Problem 2: wage rate and years of schooling

[2] Consider the following linear model based on the Cubic function:

$$w_i = \beta_1 + \beta_2 e_i^3 + \varepsilon_i,$$

where

1. w_i means daily wage measured in dollars of individual i. Let's label this variable by wage.
2. e_i means years of education, measured in years of schooling of individual i. Let's label e_i^3 by $educ3$.
3. ε_i is the error.

The `STATA` Output for this linear regression is given below:

```
      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  1,  98  ) =  100.00
       Model |   10000.00     1     10000.00           Prob > F      =  0.0000
    Residual |   10000.00    98       100.00           R-squared     =  0.5000
-------------+------------------------------           Adj R-squared =  0.4950
       Total |   20000.00    99       200.00           Root MSE      =   10.00

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       educ3|   0.05         0.0050      10.00   0.000     0.0400     0.0600
       _cons |   100.00      10.0000      10.00   0.000     80.000     120.000
------------------------------------------------------------------------------
```

Answer the following questions based on the regression output:
Question 1. What is the expected wage for a person with ten years of education?

(a) 15

(b) 50

(c) 100

[2] These questions were taken from the practice midterm provided by Professor Pinto during Spring 2024

(d) 150

(e) 500

Solution: The estimated coefficients from the `STATA` output are $\hat{\beta}_1 = 100$ and $\hat{\beta}_2 = 0.05$. So, the expected wage for a person with ten years of education is given by

$$\begin{aligned}\mathbb{E}[Y|x_0 = 10] =& \hat{b}_1 + \hat{b}_2 \cdot x_0^3 \\ =& \hat{b}_1 + \hat{b}_2 \cdot 10^3 \\ =& 100 + 0.05 \cdot 10^3 \\ =& 150\end{aligned}$$

Question 2. What is the effect of an additional year of education for a person with ten years of education?

(a) 15.66

(b) 16

(c) 16.55

(d) 17

(e) 18.66

Solution: Note that the question asks for one year more, not for the *marginal effect*. Hence, the answer does not rely on derivatives. Instead, let's take the difference between the expected wage for an individual with eleven years of education and an individual with ten years of education.

$$\begin{aligned}\mathbb{E}[Y|e = 11] - \mathbb{E}[Y|e = 10] =& (\hat{b}_1 + \hat{b}_2 \cdot 11^3) - (\hat{b}_1 + \hat{b}_2 \cdot 10^3) \\ =& \hat{b}_2 \cdot (11^3 - 10^3) \\ =& 0.05 \cdot 331 \\ =& 16.55\end{aligned}$$

Question 3. What is the *marginal effect* of education for a person with ten years of education?

(a) 13

(b) 14

(c) 15

(d) 16

(e) 17

Solution: Here, we are asked about the *marginal effect* we need to take a derivative with respect to years of education:

$$\frac{\partial \mathbb{E}[Y|e]}{\partial e} = 3 \cdot \hat{b}_2 \cdot e^2$$

Now, we evaluate the expression using $e = 10$

$$\begin{aligned}=& 3 \cdot 0.05 \cdot 10^2 \\ =& 15\end{aligned}$$

Problem 3: Ice creams and regression

[3] An ice cream vendor at UCLA football games wants to know how many ice creams to stock before each game so she doesn't run out and isn't left with too many unsold ice creams. She determines that weather is likely a large factor in the variation of ice creams sold and estimates the following relationship between ice cream sales and the temperature based on 32 home games

$$\hat{y} = -240 + 8x,$$

where $\hat{y}$ is the predicted number of ice creams sold, and x is the temperature in degrees Fahrenheit.

(a) Interpret the estimated slope and intercept. Do the estimates make sense? Why, or why not?

Solution: -240 is the intercept; it is the estimate of ice creams sold when the temperature is zero degrees Fahrenheit. This doesn't make much sense since we probably won't operate the ice cream stand when the temperature is zero degrees! (Hint for me: 0 degrees Fahrenheit equals -17 degrees Celsius)

8 is the slope term. In this context, on average, an increase of 1 degree (change in X) is associated with an increase of 8 in sales (change in Y).

(b) When the temperature on game day is predicted to be 80 degrees Fahrenheit, how many ice creams is the vendor expected to sell?

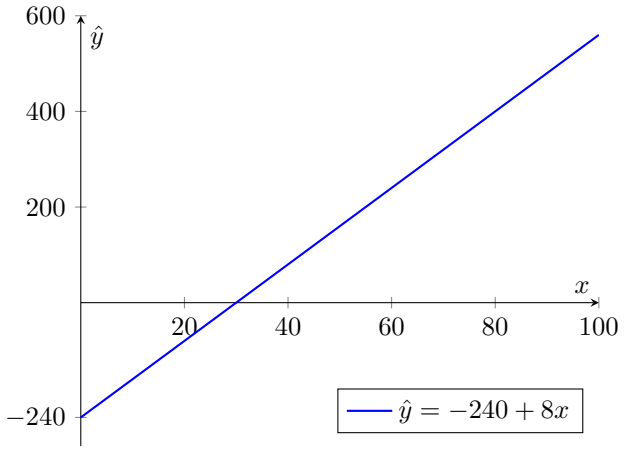
Solution: Plug in $x = 80$ to find that the vendor expects to sell $-240 + 8 \cdot 80 = 400$ ice creams.

(c) Is there a temperature at which she is not expected to sell ice cream? If so, what is it?

Solution: Here, we want to solve for the temperature x so that we expect zero sales y. So $0 = -240 + 8x$ yields $x = 30$. So, if it is thirty degrees outside, the vendor expects to make no sales.

(d) Sketch a graph of the estimated regression line.

Solution:



[3] This problem is taken, or modified, from the fourth edition of "Principles of Econometrics", by Hill, Griffiths, and Lim.