Math Camp for Economists:
Simple Regression Model

Justin C. Wiltshire

Department of Economics
University of Victoria
Summer 2023

Motivation

For most economists, a fundamental part of every research project involves asking how two variables in a population, x and y, relate to each other

→ Here we deviate from the standard practice in statistics by using lower case letters to refer to r.v.s, which is the norm in economics and econometrics

- Assume we can collect a random sample from a population of interest
- "Causal inference" is a process which (may) allow researchers to use observed data to draw causal conclusions about the independent, true effect of a change in x on y, especially when y has a complex "data generating process" (DGP)
  → A DGP is the true, underlying process which produces some data
- Credible causal inference depends on the underlying assumptions, the research design, and the estimating strategy
- Ultimately, we are trying to use statistics (functions of sample data) to infer unknown population parameters

Exercise: What are a few examples that are interesting to you?

Finding the causal effect of x on y can be hard!

Before researchers touch any data, or before analysts read the methods and results of any study, several questions should be asked. E.g.

- What if y is affected by factors other than x? How should that be dealt with?
- What is the functional relationship between x and y?
- How can a causal effect of x on y be distinguished from mere correlation?

Exercise: What are some examples of factors other than x which may impact y?

Population model

Assume a specific reduced form model of the DGP for y holds in the population

$$y = \beta_0 + \beta_1 x + u \tag{1}$$

- y is "the dependent variable" or "regressand"
- x is "the independent variable" or "regressor"
- β_0 is the "intercept" parameter (often called the "constant" term)
- β_1 is the "slope" parameter
- u is the "error term", which represents 'everything else'
  - $\rightarrow$ u represents all factors <u>other than x</u> that affect y
  - $\rightarrow$ This model treats such factors as "unobserved", so think of u this way
- This is a population model. Because we rarely observe an entire population, β_0 and β_1 are generally unobserved and need to be "estimated" using statistics and assumptions

Exercise: Come up with some examples. What might be in u in these examples?

Functional relationship between x and y

$$y = \beta_0 + \beta_1 x + u \tag{1}$$

This defines the "bivariate linear regression model"

- It assumes a linear relationship between x and y
- To see this, consider when x and y change but <u>everything else is held fixed ($\Delta u = 0$)</u>

$$\Delta y = \beta_1 \Delta x \text{ if } \Delta u = 0 \tag{2}$$

$\implies \quad \beta_1 = \frac{\Delta y}{\Delta x}$ is the change in y given the change in x

$\rightarrow$ A one unit change in x has a constant effect on y, β_1, regardless of x's initial value

Exercise: Give an example of when this would be unrealistic

The error term, u, contains everything else

We imposed in (2) that u remained fixed when x and y changed. In general, though we need to explicitly make more rigorous assumptions about how u and x are related

First, though, we can make a more general assumption about u if a constant term is included in (1)

$$E[u] = 0 \tag{3}$$

→ We can normalize the average of all unobserved factors to be 0 in the population

→ This is a trivial assumption when we include a constant term

Exercise: Suppose that $E[u] = \alpha \neq 0$

(1) Show that, as long as the constant is allowed to adjust, this does not affect the slope parameter

Relation of x and u

$E[u] = 0$ is a statement about all unobserved factors in the population model. It says nothing about how x and u are related

- If x and u are "uncorrelated" ($\rho_{xu} = 0$), then they are not "linearly related"
  - $\rightarrow$ Uncorrelatedness only measures "linear dependence"
  - $\rightarrow$ A r.v. w may be uncorrelated with a r.v. z but correlated with $f(z)$

Mean independence

A stronger assumption involves the distribution of u given any value of x

$$E[u|x] = E[u] \tag{4}$$

- $\rightarrow$ The mean value of u does *not* depend on the value of x
- $\rightarrow$ $E[u|x]$ is the same across all slices of the population determined by values of x
- $\rightarrow$ Then we say that u is "mean independent" of x
  - $\Rightarrow$ $\rho_{xu} = 0$
  - $\not\Rightarrow$ independence of x and u
- $\rightarrow$ This is *not* a trivial assumption!

Exercise: Let y be earnings and x be years of education, and let u be everything else, including unobserved ability A, such that $u = v + \gamma A$, $\gamma \neq 0$, with $E[v] = 0$ and $E[v|x] = E[v|A] = E[v]$

(1) How would you describe v?

(2) Are x and v mean independent? Are they independent? What about A and v?

(3) What would have to be true for x and u to be mean independent? Do you believe this?

(4) Suppose $\sigma_x^2 > 0$ and $\sigma_u^2 > 0$. Show that $E[u|x] = E[u] \Rightarrow \rho_{xu} = 0$
  - $\rightarrow$ Recall $E[vx|x] = xE[v|x]$ and $E[E[vx|x]] = E[vx]$

Zero conditional mean and the population regression function

If we assume (3) and (4), then we can combine them and say

$$E[u|x] = 0 \tag{5}$$

→ This is called the "zero conditional mean assumption"

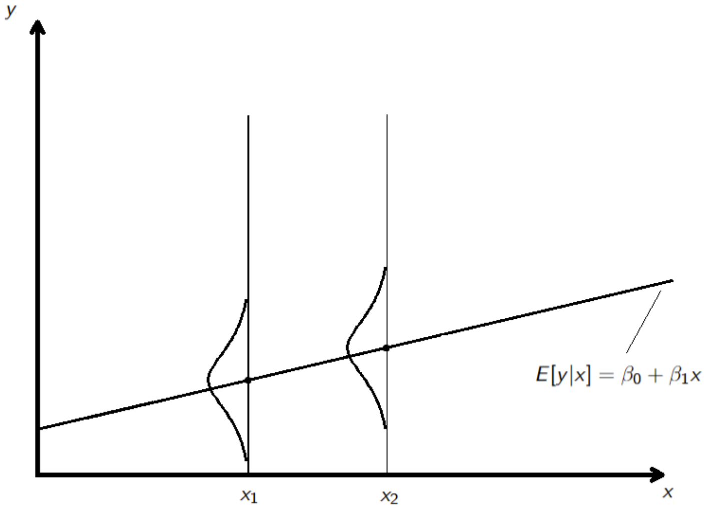
→ This is an incredibly important identifying assumption in regression models!

The population model and the zero conditional mean assumption together imply

$$E[y|x] = \beta_0 + \beta_1 x \tag{6}$$

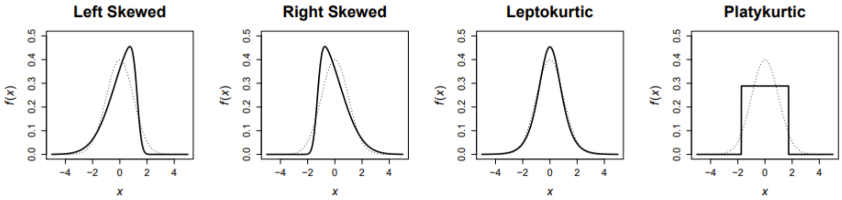
→ The "population regression function" (or "conditional expectation function")

$E[y|x]$ is the average value of y given x



Population moments

- The k-th "moment" of a r.v. z is the expected value of z^k
- The k-th "central moment" of a r.v. z is the expected value of $(z - E[z])^k$
- The k-th "standardized moment" of a r.v. z is the k-th central moment divided by $\frac{k}{2}$-th power of the second central moment
  - $\rightarrow$ The mean, μ, is the first moment
  - $\rightarrow$ The variance, σ^2 is the second central moment
  - $\rightarrow$ Skewness, $\frac{E[(x-\mu)^3]}{\sigma^3}$, the third standardized moment, measures asymmetry of the prob dist'n
  - $\rightarrow$ Kurtosis, $\frac{E[(x-\mu)^4]}{\sigma^4}$, the fourth standardized moment, measures peakedness of the prob dist'n



Moment conditions and restrictions

We have two "moment conditions" from our assumptions:

- $E[u] = 0$
- $E[ux] = 0$ (implied by $E[u] = 0$ and $E[u|x] = E[u]$)

Rearranging (1) yields $u = y - \beta_0 - \beta_1 x$, which we can plug into the moment conditions to yield two "moment restrictions" on the joint probability distribution of (x, y) in the population:

- $E[y - \beta_0 - \beta_1 x] = 0$
- $E[(y - \beta_0 - \beta_1 x)x] = 0$

Method of moments

The "method of moments" involves equating population moment conditions to their sample analogues, then solving the system of equations for the parameter estimates

So we need sample data:

- Draw pairs $\{(x_i, y_i): i = 1, ..., n\}$ of random samples of size n from the population
- Plug them into the population equation: $y_i = \beta_0 + \beta_1 x_i + u_i$

Then, with our sample data and parameter estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$, we set:

$$E[u] = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{7}$$

and

$$E[ux] = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i = 0 \tag{8}$$

Solving...

First solve (7) for $\hat{\beta}_0$:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{9}$$

Then plug this into (8) and solve for $\hat{\beta}_1$, recalling that one basic property of summation operators is $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})y_i$:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\ &= \frac{\widehat{Cov}(x,y)}{\widehat{V}(x)} \end{aligned} \tag{10}$$

Then plug this into (9) and solve for $\hat{\beta}_0$

Exercise: Solve for $\hat{\beta}_0$ and $\hat{\beta}_1$

Fitted values and residuals

$\hat{\beta}_0$ and $\hat{\beta}_1$ are called the "ordinary least squares" (OLS) estimates of β_0 and β_1. Why?

- Define the "fitted" (or "predicted") value of y_i given x_i (both from our sample) as

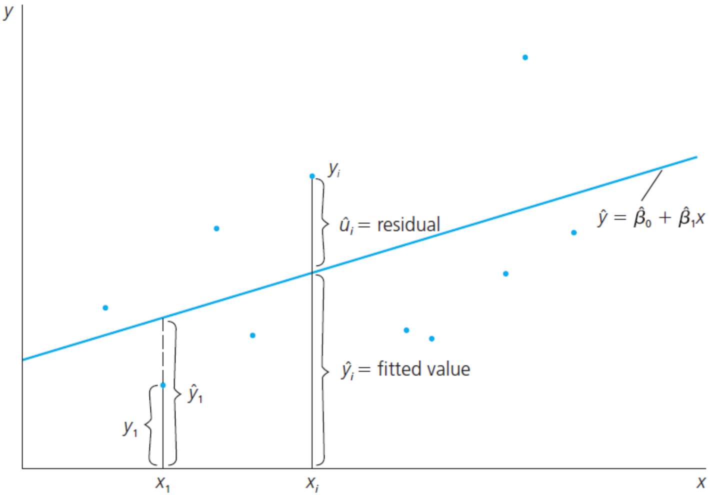
$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{11}$$

$\rightarrow$ Note that, in general, $y_i \neq \hat{y}_i$, so we have "prediction error"

- Define the "residual" (or prediction error) for observation i as

$$\begin{aligned} \hat{u}_i &= y_i - \hat{y}_i \\ &= y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \end{aligned} \tag{12}$$

Visualize the residuals. Some are positive, some are negative



Source: Wooldridge (2016)

Sum of squared residuals

It's reasonable to want estimates for β_0 and β_1 that account for the prediction errors (residuals)
But some are positive and some are negative, so summing them will cancel out a lot of information

If we square all $\hat{u}_i$, then we have measures of the n prediction errors that are all positive

- Sum the squared residuals, then choose $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimize that sum:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{13}$$

  → Note the resulting first order conditions are the same as (7) and (8) (without the $\frac{1}{n}$)

- The resulting OLS estimates, along with the fitted values, admit the "OLS regression line"

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{14}$$

  → Also called the "sample regression function"

- Given $\hat{\beta}_0$ is constant, we can difference this and isolate $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\Delta \hat{y}}{\Delta x} \tag{15}$$

  → So $\hat{\beta}_1$ is the amount that $\hat{y}$ (*not* y!) changes when x changes by one unit

Algebraic properties of OLS statistics

By construction:

(1) The residuals always sum to zero: $\sum_{i=1}^{n} \hat{u}_i = 0$

(2) The sample covariance between x and $\hat{u}$ is always zero: $\sum_{i=1}^{n} x_i \hat{u}_i = 0$

(3) The point $(\bar{x}, \bar{y})$ is always on the OLS regression line: $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$

Conditions for unbiasedness of OLS

Other things equal, unbiasedness is desirable in an estimator. Recall that bias is the difference between the expected value of an estimator and the true value of the population parameter being estimated

Four assumptions are required to establish the unbiasedness of OLS:

(1) **Linear in parameters** $\rightarrow$ In the population model, y is related to x and u as: $y = \beta_0 + \beta_1 x + u$

(2) **Random Sampling** $\rightarrow$ We have a random sample of size n, $\{(x_i, y_i) : i = 1, ..., n\}$ following the population model

(3) **Sample Variation in the Explanatory Variable** $\rightarrow$ $\{x_i : , i = 1, ..., n\}$ are *not* all the same value

(4) **Zero Conditional Mean** $\rightarrow$ $E[u|x] = 0$

Unbiasedness of OLS: Step 1

Step 1: Rewrite $\hat{\beta}_1$

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{SST_x}
\end{aligned}$$

where $SST_x = \sum_{i=1}^{n}(x_i - \bar{x})^2$

Unbiasedness of OLS: Step 2

Step 2: Replace y_i with $\beta_0 + \beta_1 x_i + u_i$, then rewrite the numerator

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{SST_x} \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})\beta_0 + \sum_{i=1}^n (x_i - \bar{x})\beta_1 x_i + \sum_{i=1}^n (x_i - \bar{x}) u_i}{SST_x} \\
&= \frac{\beta_0 \sum_{i=1}^n (x_i - \bar{x}) + \beta_1 \sum_{i=1}^n (x_i - \bar{x}) x_i + \sum_{i=1}^n (x_i - \bar{x}) u_i}{SST_x} \\
&= \frac{\beta_0 \times 0 + \beta_1 SST_x + \sum_{i=1}^n (x_i - \bar{x}) u_i}{SST_x} \\
&= \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{SST_x} \qquad (\beta_1 + \text{ slope coeff. from OLS regression of } u_i \text{ on } x_i) \\
&= \beta_1 + \sum_{i=1}^n w_i u_i \qquad \text{where } w_i = \frac{(x_i - \bar{x})}{SST_x}
\end{aligned}
$$

Unbiasedness of OLS: Step 3

Step 3: Find $E[\hat{\beta}_1]$

$$E[\hat{\beta}_1] = \beta_1 + E[\textstyle\sum_{i=1}^n w_i u_i]$$

Under random sampling and zero conditional mean

$$\begin{aligned} E[u_i|x_1,...,x_n] &= 0 \\ \Rightarrow \quad E[w_i u_i|x_1,...,x_n] &= w_i E[u_i|x_1,...,x_n] \\ &= 0 \end{aligned}$$

... because w_i is a function *only* of $\{x_1,...,x_n\}$, so it's nonrandom in the conditioning

Unbiasedness of OLS: Step 3 continued

So then, implicitly conditioning on the $\{x_1, ..., x_n\}$

$$\begin{aligned} E[\hat{\beta}_1] &= \beta_1 + E[\textstyle\sum_{i=1}^n w_i u_i] \\ &= \beta_1 + \textstyle\sum_{i=1}^n w_i E[u_i] \\ &= \beta_1 + 0 \\ &= \beta_1 \end{aligned}$$

Now average y_i across i to get $\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{u}$, and plug into $\hat{\beta}_0$:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \beta_0 + \beta_1 \bar{x} + \bar{u} \; - \hat{\beta}_1 \bar{x} = \beta_0 + (\beta_1 - \hat{\beta}_1)\bar{x} + \bar{u}$$

... and, finally, take the expected value conditioning on $\{x_1, ..., x_n\}$

$$E[\hat{\beta}_0] = \beta_0 + (\beta_1 - E[(\beta_1 - \hat{\beta}_1)\bar{x}] + E[\bar{u}] = \beta_0 + E[(\beta_1 - \hat{\beta}_1)]\bar{x} = \beta_0$$

... because $E[\hat{\beta}_1] = \beta_1 \Rightarrow E[(\beta_1 - \hat{\beta}_1)] = 0$

Exercise: OLS is unbiased under these assumptions! Give an example of when OLS would be biased

Another assumption is makes the variance nice (simple)

We also want to know how far, on average, we can expect $\hat{\beta}_j$ to be from β_j, $j \in \{0, 1\}$. Let's impose an additional assumption:

(5) **Homoskedasticity** $\rightarrow$ The variance of the error u is constant regardless of x: $V(u|x) = \sigma^2$

$\rightarrow$ Not the same as $E[u|x] = 0$ (though both are implied if we assume independence)

$$\begin{aligned} V(u|x) &= E[u^2|x] - E[u|x]^2 \\ &= E[u^2|x] - 0 \quad \text{(because } E[u|x] = 0) \\ &= \sigma^2 \\ &= E[u^2] \\ &= V(u) \quad \text{(because } E[u] = 0) \end{aligned} \tag{16}$$

Useful restatements

Assumptions (4) and (5) let us write:

- $E[y|x] = \beta_0 + \beta_1 x$
- $V(y|x) = \sigma^2$

The expected value of y changes with x, but under homoskedasticity the variance of y does not

Also recall we defined:

- $w_i = \frac{(x_i - \bar{x})}{SST_x}$
- $SST_x = \sum_{i=1}^{n}(x_i - \bar{x})^2$

Also, recall we found:

- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
- $\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} w_i u_i$

Finally, note that $Cov(\bar{y}, \hat{\beta}_1) = 0$ because $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

Variances of the OLS estimates

Then we can find the sampling variances, given we condition on the x_i, and given the u_i are independent (by random sampling) means the variance of the sum is the sum of the variances

$$\begin{aligned}
V(\hat{\beta}_1) &= V(\beta_1 + \textstyle\sum_{i=1}^n w_i u_i) \\
&= V(\textstyle\sum_{i=1}^n w_i u_i) \\
&= (\textstyle\sum_{i=1}^n w_i)^2 V(u_i) \\
&= (\textstyle\sum_{i=1}^n w_i)^2 \sigma^2 \\
&= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^4} \sigma^2 \\
&= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}
\end{aligned}$$

Exercise: Given the same assumptions, find $V(\hat{\beta}_0)$

Lab exercise 1 (From Wooldridge 2016)

Suppose we have collected the following data on GPA and ACT score for 8 students (ACT is a common standardized test in the USA, and stands for "American College Testing")

Student	GPA	ACT Score
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

(1) Estimate the relationship between *GPA* and *ACT* using OLS; that is, obtain the intercept and slope estimates in the equation $\widehat{GPA} = \hat{\beta}_0 + \hat{\beta}_1 ACT$
Comment on the direction of the relationship. Does the intercept have a useful interpretation here? Explain. How much higher is the *GPA* predicted to be if the *ACT* score is increased by five points?

(2) Compute the fitted values and residuals for each observation, and verify that the residuals (approximately) sum to zero.

(3) What is the predicted value of *GPA* when $ACT = 20$?

Lab exercise 2 (repeats from the lesson)

Let $y = \beta_0 + \beta_1 x + u$ be earnings, where x is years of education and u is everything else, including unobserved ability A, such that $u = v + \gamma A$, $\gamma \neq 0$, with $E[v] = 0$ and $E[v|x] = E[v|A] = E[v]$

(1) How would you describe v?

(2) Are x and v mean independent? Are they independent? What about A and v?

(3) What would have to be true for x and u to be mean independent? Do you believe this?

(4) Suppose $\sigma_x^2 > 0$ and $\sigma_u^2 > 0$. Show that $E[u|x] = E[u] \Rightarrow \rho_{xu} = 0$

   $\rightarrow$ Recall $E[vx|x] = xE[v|x]$ and $E[E[vx|x]] = E[vx]$

Lab exercise 3 (repeats from the lesson)

Assume $y = \beta_0 + \beta_1 x + u$ holds in the population. Assume $E[u] = 0$ and $E[u|x] = E[u]$.

(1) What are the the two moment conditions we can define?

(2) What are the two associated moment restrictions?

(3) Use the method of moments to solve for the OLS estimators for β_0 and β_1

(4) Solve for these OLS estimators by minimizing the sum of squared residuals

Lab exercise 4 (repeats from the lesson)

Under the four assumptions required for the unbiasedness of OLS:

(1) Show that $\hat{\beta}_1$ is unbiased for β_1

(2) Show that $\hat{\beta}_0$ is unbiased for β_0

Lab exercise 5

Given the assumptions required for OLS to be unbiased, plus homoskedasticity:

(1) Find $V(\hat{\beta}_1)$

(2) Find $V(\hat{\beta}_0)$

References

This drew notes from:

- Introductory Econometrics: A Modern Approach (Wooldridge, 2016)
- *Causal Inference: The Mixtape* (Cunningham, 2021)
- Nathan Helwig's notes on Random Variables