Econ 103: Introduction to Econometrics

Lecture 5 — The Simple Linear Regression Model

Ryan Longmuir

UCLA

Summer Session C, 2026

Reading: Hill, Griffiths & Lim (5th ed.), §2.1–2.2; Stock & Watson (4th ed.), §4.1, 4.4.

Where we are

The last four lectures built a probability toolkit. The very last idea — from the bivariate Normal — was that a conditional mean can be a straight line:

$$\mathbb{E}(Y \mid X = x) = \alpha + \beta x, \qquad \beta = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}.$$

Today that line becomes the object of the whole course: the **simple linear regression model**. We will

- write down the model $y = \beta_1 + \beta_2 x + e$ and interpret its pieces,
- separate parameters, estimators, and estimates,
- and state the assumptions (SR1–SR6) that make it all work.

Recall Lecture 1

We scatter-plotted weekly food expenditure against income and eyeballed an upward cloud. Today we write the model behind that cloud; next lecture we fit the line.

Today's plan

1. From an economic idea to a model
2. The simple linear regression model
3. Parameters, estimators, estimates
4. The assumptions: SR1–SR6

Part 1

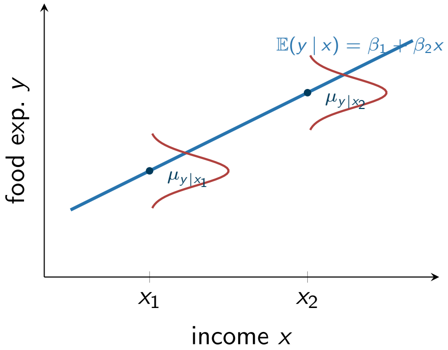
From an economic idea to a model

The economic question

Running example (Hill, Griffiths & Lim). How does a household's weekly *food expenditure* y depend on its weekly *income* x?

Even among households with the *same* income, food spending varies (tastes, household size, restaurants, impulse buys). So at each income x, y has a conditional distribution $f(y \mid x)$.

Economic theory speaks to the center of that distribution — the **conditional mean** $\mathbb{E}(y \mid x)$ — which we expect to rise with income.



At each income, a spread of outcomes centered on the line.

From a rule to a model

A made-up deterministic rule. Suppose a household spends \$80 plus 10 cents of each dollar of income on food:

$$y = 80 + 0.10\,x.$$

A \$100 rise in income x raises spending by exactly \$10 — the marginal propensity to spend on food, 0.10, is the slope.

Reality is not deterministic. Countless other factors move food spending. Collect them into a **random error** e, and replace the fixed numbers $80, 0.10$ by unknown parameters β_1, β_2:

$$\boxed{y = \beta_1 + \beta_2 x + e}.$$

This is the "systematic part + random error" template from Lecture 1 — now with one explanatory variable and named parameters.

Part 2

The simple linear regression model

The model and its vocabulary

The simple linear regression model

For each observation $i = 1, \ldots, N$,

$$y_i = \beta_1 + \beta_2 x_i + e_i.$$

- y_i — **dependent** variable (regressand, "left-hand side").
- x_i — **independent** / explanatory variable (regressor).
- e_i — **random error**: everything else affecting y.
- β_1 — **intercept** parameter.
- β_2 — **slope** parameter.
- β_1, β_2 are fixed, unknown population parameters.

"Simple" means *one* regressor — not that it is easy. Everything generalizes to many regressors in Lecture 13.

The regression function and the systematic/random split

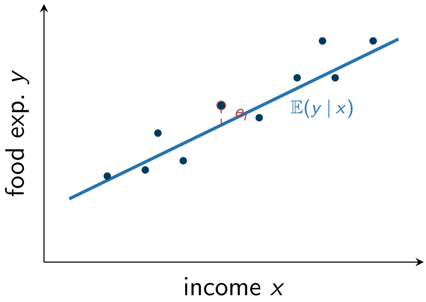
If (as we will assume) the errors average to zero at each x, then taking the conditional mean of $y_i = \beta_1 + \beta_2 x_i + e_i$ gives the **population regression function**

$$\mathbb{E}(y \mid x) = \beta_1 + \beta_2 x.$$

So every observation splits into

$$y_i = \underbrace{\mathbb{E}(y_i \mid x_i)}_{\text{systematic}} + \underbrace{e_i}_{\text{random}} .$$

The line is the *average* behavior; e_i is the ith household's departure from it.



e_i = vertical gap from the point to the line.

Interpreting the slope: the "how much" question

The slope is the marginal effect of x on the average of y:

$$\beta_2 = \frac{\Delta\, \mathbb{E}(y \mid x)}{\Delta x} = \frac{d\, \mathbb{E}(y \mid x)}{dx}.$$

- Hold "everything else" fixed ($\Delta e = 0$). A change Δx moves average spending by $\beta_2\, \Delta x$ — *ceteris paribus*.
- In the food example, if income rises by \$100 then average food expenditure rises by $\beta_2 \times \$100$. This is the number a decision-maker wants.

The intercept $\beta_1 = \mathbb{E}(y \mid x = 0)$

Sometimes meaningful, often not. With test scores on class size, β_1 would be the predicted score for a class of *zero* students — nonsense. Then treat β_1 as just the height that pins the line.

Part 3

Parameters, estimators, estimates

Three things people sloppily all call "beta"

Keeping these straight is the key conceptual hurdle of the course.

Parameter

β_1, β_2 Fixed, unknown features of the population. There is one true value. *Not* random.

Estimator

b_1, b_2 A formula applied to a sample. Since the sample is random, the estimator is a random variable — it has a sampling distribution.

Estimate

$b_1 = 83.4, \ldots$ The number the estimator produces in *one* particular sample. Just a number — not random.

The Lecture-4 connection

The estimator b_2 is to β_2 exactly as the sample mean $\bar{Y}$ is to μ: a random variable that varies sample to sample, with a center and a spread we can study. *That* is how we will judge it (Lectures 7–10).

Error vs. residual

A distinction that trips everyone up — because it hinges on parameter vs. estimate.

Random error e_i (population)

$$e_i = y_i - (\beta_1 + \beta_2 x_i) = y_i - \mathbb{E}(y_i \,|\, x_i).$$

Uses the true parameters. Unobservable — we never know β_1, β_2.

Residual $\hat{e}_i$ (sample)

$$\hat{e}_i = y_i - (b_1 + b_2 x_i) = y_i - \hat{y}_i.$$

Uses the estimated line. Observable — we compute it once we fit (next lecture).

The parallel

e_i is to β as $\hat{e}_i$ is to b. The residual is our visible *stand-in* for the invisible error — and minimizing the residuals is exactly how OLS picks the line in Lecture 6.

Part 4

The assumptions: SR1–SR6

Why assumptions?

A model is only as trustworthy as the conditions behind it. The simple regression assumptions **SR1–SR6** are the conditions under which

- the slope β_2 measures a causal marginal effect, and
- our estimators b_1, b_2 are well behaved (unbiased, with a known sampling distribution for inference).

Much of econometrics is what to do *when* an assumption fails. So it pays to know exactly what we are assuming — and which assumption each later technique is rescuing.

"SR" = simple regression. We meet them one at a time, then collect them.

SR1–SR2: the model and strict exogeneity

SR1 — the model holds in the population

$y_i = \beta_1 + \beta_2 x_i + e_i$ for all $i = 1, \ldots, N$.

SR2 — strict exogeneity (the crucial one)

The error has conditional mean zero given the regressor(s):

$$\mathbb{E}(e_i \mid x) = 0.$$

What SR2 says: knowing x tells you nothing about the average error — the omitted factors balance out to zero at every value of x. It delivers two things:

$$\mathbb{E}(e_i \mid x) = 0 \implies \mathbb{E}(e_i) = 0 \text{ and } \mathrm{Cov}(e_i, x_i) = 0,$$

and hence the regression function $\mathbb{E}(y_i \mid x) = \beta_1 + \beta_2 x_i$.

SR2, continued: exogenous vs. endogenous

- If $\mathrm{Cov}(e, x) = 0$, x is **exogenous** — regression can recover β_1, β_2, and β_2 is the causal marginal effect.
- If $\mathrm{Cov}(e, x) \neq 0$, x is **endogenous** — and β_2 is not causal. This is the formal version of "correlation $\neq$ causation" from Lecture 1.

Example (Wages and education (HGL))

$WAGE_i = \beta_1 + \beta_2\, EDUC_i + e_i$. The error e holds ability, drive, intelligence — plausibly correlated with education. Then $\mathbb{E}(e \mid EDUC) \neq 0$, education is endogenous, and b_2 confounds the return to schooling with the effect of ability. (We tackle this much later.)

SR3–SR4: spread and dependence of the errors

SR3 — homoskedasticity

The error has constant conditional variance:

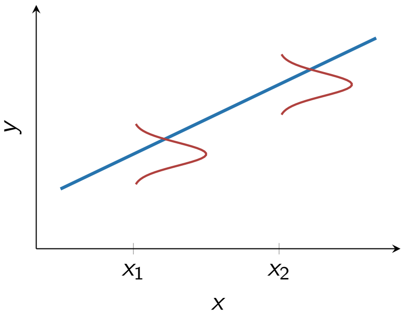
$$\mathrm{Var}(e_i \mid x) = \sigma^2.$$

Same spread of y about the line at *every* x. If it varies, errors are **heteroskedastic**.

SR4 — uncorrelated errors

$$\mathrm{Cov}(e_i, e_j \mid x) = 0, \quad i \neq j.$$

One observation's error carries no information about another's. (Fails with clustered or time-series data.)



SR3: the two bells have the *same* width.

SR5–SR6: variation in x, and (optional) normality

SR5 — the regressor must vary

In the sample, x_i takes at least two different values. "It takes two points to determine a line" — with no variation in x there is no slope to estimate.

SR6 — normality of errors (optional)

$$e_i \,|\, x \sim N(0, \sigma^2) \quad \iff \quad y_i \,|\, x \sim N(\beta_1 + \beta_2 x_i,\ \sigma^2).$$

SR6 is *not* needed for the estimators to work, but it makes small-sample inference exact (Lecture 9). And it is plausible: by the CLT (Lecture 4), an error that sums up many small factors tends toward normal.

SR1–SR6 at a glance

SR1	$y_i = \beta_1 + \beta_2 x_i + e_i$
SR2	$\mathbb{E}(e_i \mid x) = 0$ (strict exogeneity)
SR3	$\mathrm{Var}(e_i \mid x) = \sigma^2$ (homoskedastic)
SR4	$\mathrm{Cov}(e_i, e_j \mid x) = 0,\ i \neq j$
SR5	x_i takes ≥ 2 values
SR6	$e_i \mid x \sim N(0, \sigma^2)$ (optional)

Same idea in Stock & Watson

S&W write $Y_i = \beta_0 + \beta_1 X_i + u_i$ and list three:

1. $\mathbb{E}(u_i \mid X_i) = 0$ (= SR2)
2. (X_i, Y_i) i.i.d.
3. large outliers unlikely (finite 4th moments)

S&W drop homoskedasticity (they use robust errors) and add an outlier condition. We use HGL's SR1–SR6.

Recap

The model

- $y_i = \beta_1 + \beta_2 x_i + e_i$
- regression function $\mathbb{E}(y \mid x) = \beta_1 + \beta_2 x$
- y = systematic + random
- slope $\beta_2 = \Delta\mathbb{E}(y \mid x)/\Delta x$ (marginal effect)

Three "betas"

- parameter β (fixed) $\rightarrow$ estimator b (random) $\rightarrow$ estimate (a number)
- error e (unobserved) vs. residual $\hat{e}$ (computed)

Assumptions SR1–SR6

- SR1 model; SR2 $\mathbb{E}(e \mid x) = 0$ (exogeneity $\Rightarrow$ causal; else endogenous)
- SR3 homoskedastic; SR4 uncorrelated
- SR5 x varies; SR6 normal (optional)

Next time (Lecture 6): OLS estimation

We have the model and assumptions — but not the line. Next lecture: pick b_1, b_2 to minimize the residuals; the slope is $b_2 = \mathrm{Cov}(x, y)/\mathrm{Var}(x)$.

Questions?