Econ 103: Introduction to Econometrics

Lecture 6 — OLS Estimation

Ryan Longmuir

UCLA

Summer Session C, 2026

Reading: Hill, Griffiths & Lim (5th ed.), §2.3; Stock & Watson (4th ed.), §4.2.

Where we are

Last lecture we wrote down the model and its assumptions but *never fit the line*:

$$y_i = \beta_1 + \beta_2 x_i + e_i, \qquad \mathbb{E}(y \mid x) = \beta_1 + \beta_2 x.$$

The parameters β_1, β_2 are fixed but unknown. All we have is a sample of N points (x_i, y_i).

Today we turn the sample into numbers. We will

- state the **least squares principle** — the rule for choosing a line,
- derive the estimators $b_2 = \widehat{\mathrm{Cov}}(x, y)/\widehat{\mathrm{Var}}(x)$ and $b_1 = \bar{y} - b_2\bar{x}$,
- and compute them for the food-expenditure data, by hand and in R.

The one-sentence preview

OLS picks the line that makes the residuals as small as possible — and the answer is just a ratio of sample moments you already met in Lecture 3.

Today's plan

1. The least squares principle
2. Deriving the OLS estimators
3. The food-expenditure example
4. OLS in R

Part 1

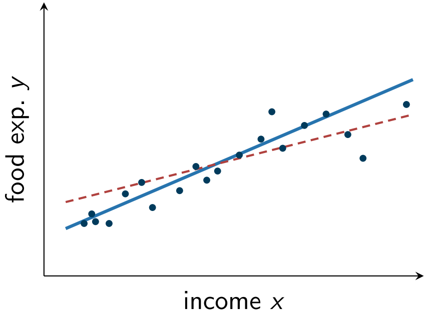
The least squares principle

The problem: which line?

We want to locate the population mean line $\mathbb{E}(y \mid x) = \beta_1 + \beta_2 x$ somewhere in the middle of the cloud. Two bad ideas:

- Freehand. Draw it by eye — but everyone draws a different line, and there is no rule to assess.
- Two endpoints. Connect the lowest- and highest-income points — a rule, but it throws away the other 38 observations.

We want a rule that uses every point and gives *one* answer.



Many lines pass through the cloud. Which is "best"?

Residuals: the vertical misses

Fix *any* candidate line with intercept b_1 and slope b_2. Its **fitted value** at x_i is

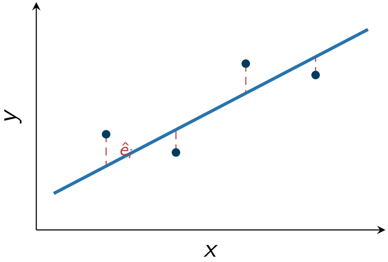
$$\hat{y}_i = b_1 + b_2 x_i,$$

and the **least squares residual** is the vertical gap from the point to the line:

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

- $\hat{e}_i > 0$: point above the line (we under-predicted).
- $\hat{e}_i < 0$: point below the line.
- A good line should make these misses small *overall*.

Recall from Lecture 5: the residual $\hat{e}_i$ is the observable *stand-in* for the unobservable error e_i.



Residuals = dashed vertical segments.

The least squares criterion

The least squares principle

Choose the line that makes the sum of squared residuals as small as possible:

$$\min_{b_1, b_2} S(b_1, b_2) = \sum_{i=1}^{N} \hat{e}_i^2 = \sum_{i=1}^{N} \left(y_i - b_1 - b_2 x_i\right)^2.$$

Why *squared* distances?

- Squaring makes every miss positive, so large positive and large negative misses cannot cancel (which is why we don't just minimize $\sum \hat{e}_i$).
- It penalizes big misses much more than small ones — the line is pulled toward the bulk of the data.
- It makes the minimization a clean calculus problem with a unique closed-form answer (next section).

What "least squares" buys us

Call the minimizing values b_1, b_2 and write the sum of squared residuals they achieve as

$$\text{SSE} = \sum_{i=1}^{N} \hat{e}_i^2, \qquad \hat{e}_i = y_i - b_1 - b_2 x_i.$$

For *any* other line $\hat{y}_i^* = b_1^* + b_2^* x_i$ with squared-residual total SSE^*,

$$\boxed{\text{SSE} \leq \text{SSE}^*} \qquad \text{(strict unless the lines coincide).}$$

No matter how cleverly you draw an alternative, you cannot beat the least squares line on this criterion. The intercept and slope that achieve the minimum are the **ordinary least squares** (OLS) estimates.

"Ordinary" distinguishes OLS from variants (generalized, weighted, two-stage least squares) you may meet later. There is nothing ordinary about how often it is used.

Part 2

Deriving the OLS estimators

Minimizing the sum of squares

$S(b_1, b_2) = \sum (y_i - b_1 - b_2 x_i)^2$ is a smooth, bowl-shaped (convex) function of two unknowns. Its minimum is where both partial derivatives vanish:

$$\frac{\partial S}{\partial b_1} = -2 \sum \big(y_i - b_1 - b_2 x_i\big) = 0,$$
$$\frac{\partial S}{\partial b_2} = -2 \sum x_i \big(y_i - b_1 - b_2 x_i\big) = 0.$$

Dropping the -2 and rearranging gives the two **normal equations**:

$$\sum y_i = N b_1 + b_2 \sum x_i, \qquad \sum x_i y_i = b_1 \sum x_i + b_2 \sum x_i^2.$$

Two linear equations, two unknowns (b_1, b_2). Solve them.

Notice each first-order condition is a statement about residuals: $\sum \hat{e}_i = 0$ and $\sum x_i \hat{e}_i = 0$. The least squares residuals sum to zero and are uncorrelated with x *by construction*.

Solving for the intercept

Take the first normal equation $\sum y_i = Nb_1 + b_2 \sum x_i$ and divide by N:

$$\bar{y} = b_1 + b_2\bar{x} \quad \Longrightarrow \quad \boxed{b_1 = \bar{y} - b_2\bar{x}}.$$

The fitted line passes through the point of the means

Rearranged, this says $\bar{y} = b_1 + b_2\bar{x}$: the OLS line always goes through $(\bar{x}, \bar{y})$. The "point of the means" is a pivot — the line is anchored there and tilts to the best slope.

So once we know the slope b_2, the intercept is immediate. The real work is the slope.

Solving for the slope

Substitute $b_1 = \bar{y} - b_2\bar{x}$ into the second normal equation and collect terms (algebra in HGL Appendix 2A). The result, in **deviation-from-means** form:

$$b_2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

- Numerator: how x and y co-move about their means.
- Denominator: how much x varies about its mean.
- Needs $\sum(x_i - \bar{x})^2 \neq 0$ — exactly assumption SR5 (x must take ≥ 2 values), or the slope is $0/0$.

Sign of the slope

b_2 has the same sign as the sample covariance of x and y: positive co-movement $\Rightarrow$ upward-sloping fit.

The slope is a ratio of sample moments

Divide top and bottom by $N-1$. The numerator becomes the sample covariance and the denominator the sample variance of x:

$$b_2 = \frac{\frac{1}{N-1}\sum(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{N-1}\sum(x_i-\bar{x})^2} = \frac{\widehat{\mathrm{Cov}}(x,y)}{\widehat{\mathrm{Var}}(x)}.$$

The Lecture-3 echo

In Lecture 3 we found the bivariate-Normal regression slope

$$\beta_2 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}.$$

OLS is the sample analog: replace population moments with their sample counterparts. The estimator mirrors the parameter, moment for moment — this is the **analogy principle** at work.

Estimator vs. estimate, one more time

The formulas $b_2 = \dfrac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$ and $b_1 = \bar{y} - b_2\bar{x}$ are perfectly general — they work for whatever data turn up.

As formulas ⇒ estimators

Viewed as rules to be applied to a *random* sample, b_1, b_2 are random variables — they have a sampling distribution. This is what we study in Lecture 7.

Plugged-in numbers ⇒ estimates

Applied to *one* observed sample, they spit out numbers ($b_2 = 10.21, \dots$). Just numbers — not random.

Same symbol, two meanings. Keeping them apart is the through-line of the whole course.

Part 3

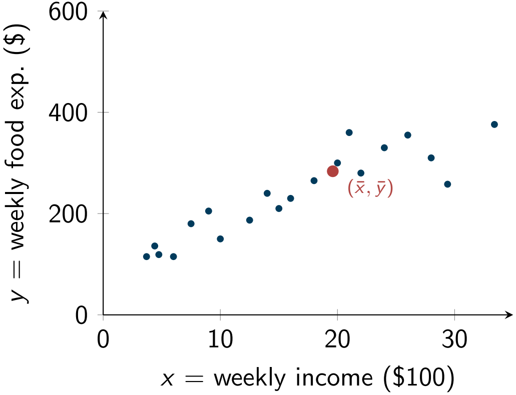
The food-expenditure example

The data

$N = 40$ three-person households (HGL data file `food`). For each:

- y_i — weekly food expenditure ($),
- x_i — weekly income ($100 units).

household	y_i	x_i
1	115.22	3.69
2	135.98	4.39
⋮	⋮	⋮
40	375.73	33.40
mean	283.57	19.60



Representative subset; red dot is the point of the means.

Turning the crank

Plug the sample sums into the formulas (HGL Example 2.4):

$$b_2 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{18671.2684}{1828.7876} = 10.2096,$$

$$b_1 = \bar{y} - b_2\bar{x} = 283.5735 - (10.2096)(19.6048) = 83.4160.$$

The fitted regression line

$$\widehat{\text{FOOD\_EXP}}_i = 83.42 + 10.21\,\text{INCOME}_i$$

This is *the* line: of all possible lines, it minimizes $\sum \hat{e}_i^2$, and it passes through $(\bar{x}, \bar{y}) = (19.60, 283.57)$.

Interpreting the estimates

Slope $b_2 = 10.21$

Income is in \$100 units, so: a \$100 rise in weekly income is associated with about \$10.21 more weekly food spending, on average, holding all else fixed. This is the "how much" number a planner wants.

Intercept $b_1 = 83.42$

Literally, predicted food spending at zero income. But we have no data near $x = 0$ (poorest household earns \$369/wk), so don't take it literally — read it as the height that pins the line.

Example (Point prediction)

For a household with \$2,000 weekly income ($x_0 = 20$):

$$\hat{y}_0 = 83.42 + 10.21(20) = 287.61.$$

We predict \$287.61 of weekly food spending. (How sure are we? That is a prediction interval — Lecture 8 and Lecture 11.)

Elasticity: a unit-free reading

A slope depends on units. An **elasticity** — the % change in y per % change in x — does not. On a line,

$$\hat{\varepsilon} = b_2 \cdot \frac{x}{\hat{y}},$$

which changes along the line, so we report it at the representative point of the means:

$$\hat{\varepsilon} = 10.21 \times \frac{19.60}{283.57} = 0.71.$$

Reading it

A 1% rise in income is associated with about a 0.71% rise in food spending. Since $0.71 < 1$, food is a **necessity** (demand grows less than proportionately with income) — exactly what economic theory predicts.

Part 4

OLS in R

Doing it in R: `lm()`

You will almost never compute b_1, b_2 by hand again. In R the workhorse is `lm()` ("linear model"):

```
library(POE5Rdata)      # course data package
data(food)              # load the food data frame

fit <- lm(food_exp ~ income, data = food)
coef(fit)
#> (Intercept)      income
#>    83.41600    10.20964
```

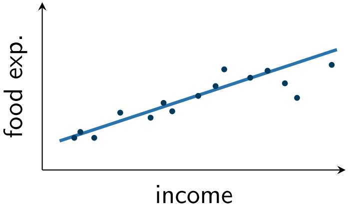
Read the formula `food_exp ~ income` as "regress `food_exp` on `income`." R minimizes $\sum \hat{e}_i^2$ for you and returns the same $b_1 = 83.42$, $b_2 = 10.21$ we found by hand.

Reading the output and plotting the fit

```
summary(fit)
# Coefficients:
#              Estimate Std. Error ...
# (Intercept)   83.4160    43.4102 ...
# income        10.2096     2.0933 ...
```

- `Estimate` — the *b*'s, our 83.42 and 10.21.
- `Std. Error` — how much they'd wobble across samples (Lecture 7).
- The rest (t, p, R^2) — Lectures 8–10.

```
plot(food$income,
     food$food_exp)
abline(fit, col = "blue",
       lwd = 2)
```



Recap

The principle

- residual $\hat{e}_i = y_i - b_1 - b_2 x_i$
- OLS minimizes $\sum \hat{e}_i^2$ (squares so misses can't cancel)
- beats every other line: $\mathrm{SSE} \leq \mathrm{SSE}^*$

The estimators

- $b_2 = \dfrac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \dfrac{\widehat{\mathrm{Cov}}(x,y)}{\widehat{\mathrm{Var}}(x)}$
- $b_1 = \bar{y} - b_2\bar{x}$; line hits $(\bar{x}, \bar{y})$

The example

- $\widehat{\text{FOOD\_EXP}} = 83.42 + 10.21\,\text{INCOME}$
- slope: \$10.21 more food per \$100 income
- elasticity at means $= 0.71$ (food a necessity)
- in R: `lm(food_exp ~ income)`

Next time (Lecture 7): is OLS any good?

We have *a* line. But b_1, b_2 are random variables — so is the procedure unbiased? How precise? We will show OLS is unbiased and, under SR1–SR5, the **best linear unbiased estimator** (Gauss–Markov).

Questions?