Ordinary Least Squares Estimation

How to Draw the Best Line Through a Scatter Plot

Jake Anderson

UCLA Economics

Econ 103 – Lecture 3

Outline

The Food Expenditure Data

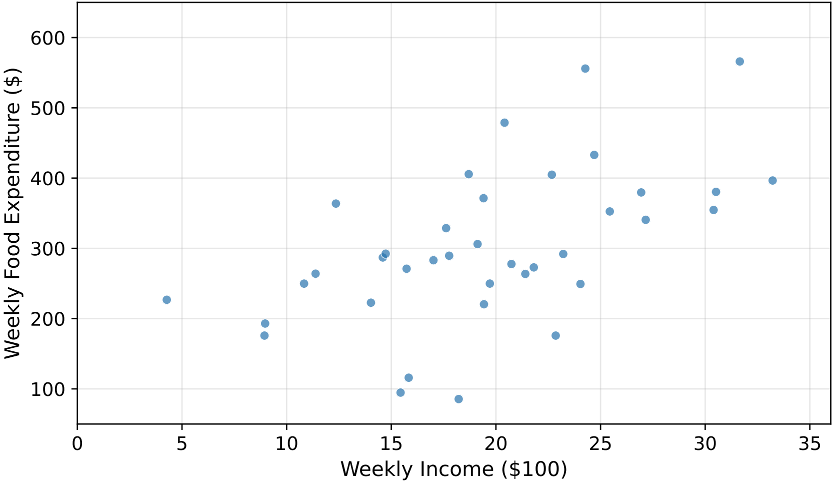
$N = 40$ three-person households from southern Australia.

- y_i: weekly food expenditure per person ($)
- x_i: weekly household income (in $100 units)

	Food Expenditure ($)	**Weekly Income ($100)**
Mean	283.57	19.60
Std. dev.	112.68	6.85
Min	109.71	3.69
Max	587.66	33.40

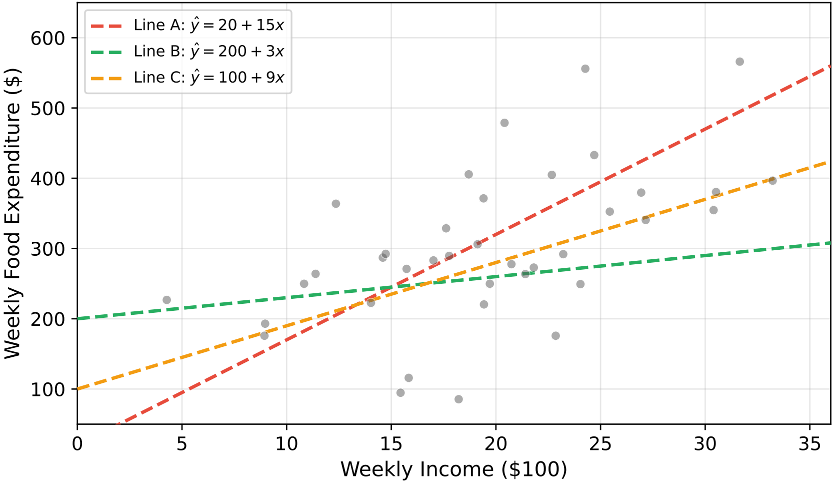
Do richer households spend more on food? Let's look at the scatter plot.

The Scatter Plot



Many lines could be drawn through these points. Which one is “best”?

Eyeballing Fails



Three different people could draw three different “best” lines. We need an **objective criterion** for choosing the line.

Recap: The Simple Linear Regression Model

From Topic 5, we have the simple linear regression model:

$$y_i = \beta_1 + \beta_2 x_i + e_i, \quad i = 1, \ldots, N$$

Under our model assumptions, the regression function is:

$$\mathrm{E}(y_i \mid x_i) = \beta_1 + \beta_2 x_i$$

β_1 and β_2 are **unknown population parameters**. We have a sample of N data pairs (y_i, x_i).
Model: $\text{food}_i = \beta_1 + \beta_2 \, \text{income}_i + e_i$
We need estimates b_1 and b_2 to draw a fitted line through the scatter plot.

Residuals: Measuring the Misfit

For any candidate line $\hat{y}_i = b_1 + b_2 x_i$, the **residual** for observation i is:

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

- $\hat{e}_i > 0$: the point is **above** the line (underprediction)
- $\hat{e}_i < 0$: the point is **below** the line (overprediction)
- $\hat{e}_i = 0$: the point falls **exactly on** the line

Do not confuse e_i (the true error, which we never observe) with $\hat{e}_i$ (the residual, which we compute from our fitted line).

A good line should make these residuals **small overall**.

Why not minimize $\sum \hat{e}_i$? Because positive and negative residuals cancel out. A terrible line through the middle could have $\sum \hat{e}_i = 0$.

The Sum of Squared Residuals

You might think: minimize $\sum |\hat{e}_i|$ instead. That works (it gives **median regression**) but has no closed-form solution. Squaring gives us clean calculus.

Define the **sum of squared residuals (SSR)**:

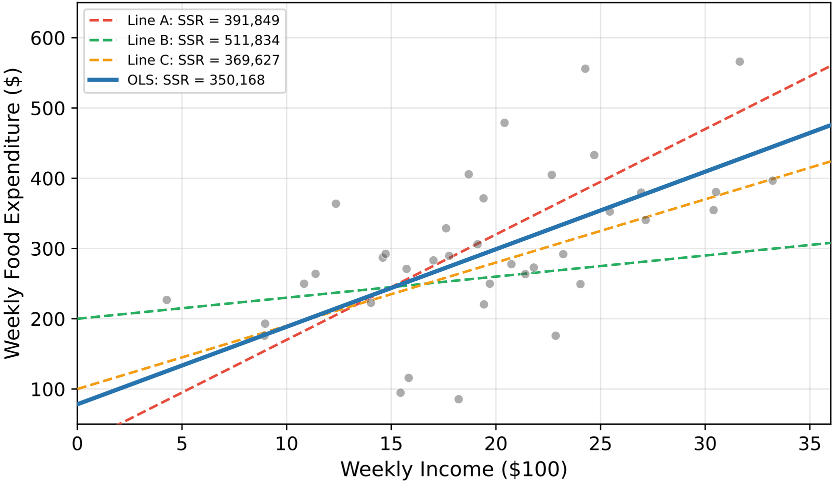
$$S(b_1, b_2) = \sum_{i=1}^{N} \hat{e}_i^2 = \sum_{i=1}^{N} (y_i - b_1 - b_2 x_i)^2$$

- Squaring ensures every residual contributes **positively**
- Large residuals are penalized more heavily than small ones
- $S(b_1, b_2) \geq 0$ always, with $S = 0$ only if the line passes through every point

The Least Squares Principle: choose b_1 and b_2 to minimize $S(b_1, b_2)$.

The values of b_1 and b_2 that achieve this minimum are the **Ordinary Least Squares (OLS) estimators**.

Comparing Lines by Their SSR



The OLS line has the smallest possible SSR. No other line can do better.

The Minimization Problem

We want to find the b_1 and b_2 that minimize:

$$S(b_1, b_2) = \sum_{i=1}^{N} (y_i - b_1 - b_2 x_i)^2$$

This is a function of **two variables** (b_1 and b_2). The data (y_i, x_i) are fixed numbers from our sample.

The function S is a "bowl-shaped" surface (a paraboloid opening upward). The minimum is at the bottom of the bowl.

Strategy: take partial derivatives with respect to b_1 and b_2, set them equal to zero, and solve.

First-Order Conditions

Taking the partial derivative with respect to b_1:

$$\frac{\partial S}{\partial b_1} = -2\sum_{i=1}^{N}(y_i - b_1 - b_2 x_i) = 0$$

Taking the partial derivative with respect to b_2:

$$\frac{\partial S}{\partial b_2} = -2\sum_{i=1}^{N}x_i(y_i - b_1 - b_2 x_i) = 0$$

These are two equations in two unknowns (b_1, b_2).

Dividing by -2 and expanding the first equation:

$$\sum y_i - Nb_1 - b_2 \sum x_i = 0$$

Rearranging both, we get the **normal equations**:

$$Nb_1 + \left(\sum x_i\right) b_2 = \sum y_i$$

$$\left(\sum x_i\right) b_1 + \left(\sum x_i^2\right) b_2 = \sum x_i y_i$$

Solving: The Intercept

From the first normal equation:

$$Nb_1 + \left(\sum x_i\right) b_2 = \sum y_i$$

Divide both sides by N:

$$b_1 + \bar{x}\, b_2 = \bar{y}$$

Solve for b_1:

$$\boxed{b_1 = \bar{y} - b_2 \bar{x}}$$

$\Longrightarrow$ The fitted line always passes through the point $(\bar{x}, \bar{y})$.
Once we find b_2, we get b_1 for free.

Solving: The Slope (Raw Form)

Substitute $b_1 = \bar{y} - b_2\bar{x}$ into the second normal equation and simplify:

$$b_2 = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{N\sum x_i^2 - (\sum x_i)^2}$$

This is the "raw sums" formula. It works but is hard to interpret.

Using the summation identities (you can verify these by expanding $(x_i - \bar{x})(y_i - \bar{y})$ and distributing the sum):

$$N\sum x_i y_i - \sum x_i \sum y_i = N\sum (x_i - \bar{x})(y_i - \bar{y})$$
$$N\sum x_i^2 - (\sum x_i)^2 = N\sum (x_i - \bar{x})^2$$

we can rewrite b_2 in a more revealing form.

Solving: The Slope (Deviation from Mean Form)

$$\boxed{b_2 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}}$$

This formula has a natural interpretation:

- **Numerator:** measures how x and y **co-vary** around their means
- **Denominator:** measures how much x **varies** around its mean

In other words:

$$b_2 = \frac{\text{sample covariance of } x \text{ and } y}{\text{sample variance of } x}$$

(the $N - 1$ denominators cancel exactly)

$\Longrightarrow$ The slope estimate captures how y moves with x, scaled by how much x moves on its own.

The OLS Estimators: Summary

Ordinary Least Squares (OLS) Estimators

Slope:

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

Intercept:

$$b_1 = \bar{y} - b_2\bar{x}$$

These formulas are **perfectly general**: they work for any sample of (y_i, x_i) data, provided x_i takes at least two distinct values (SR5).

The **fitted line** is: $\hat{y}_i = b_1 + b_2 x_i$

From Formulas to Numbers

We started with a scatter plot and no idea which line to draw.

The least squares principle gave us a criterion: minimize the SSR.

Calculus gave us formulas for b_1 and b_2.

Now let's use them on the food expenditure data.

Computing b_2: The Slope

From the $N = 40$ food expenditure observations:

$$\bar{x} = 19.6048, \quad \bar{y} = 283.5735$$

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = 18{,}671.27$$

$$\sum (x_i - \bar{x})^2 = 1{,}828.79$$

$$b_2 = \frac{18{,}671.27}{1{,}828.79} = 10.2096$$

Interpretation: A $100 increase in weekly income is associated with a $10.21 increase in expected weekly food expenditure.

Computing b_1: The Intercept

Using full-precision values:

$$b_1 = \bar{y} - b_2\bar{x} = 283.5735 - (10.2096)(19.6048) \approx 83.42$$

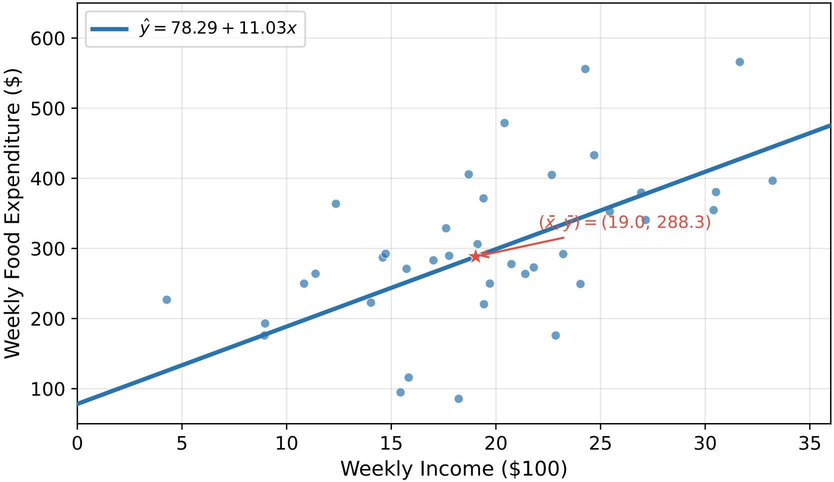
Interpretation: The estimated expected food expenditure when income is zero is $83.42.

Caution: no households in the sample have income near zero. This is an **extrapolation** beyond the data range. The intercept completes the line but may not have a meaningful economic interpretation here.

The fitted regression line:

$$\hat{y}_i = 83.42 + 10.21\, x_i$$

The OLS Fitted Line



The line passes through $(\bar{x}, \bar{y}) = (19.60, 283.57)$, as guaranteed by the formula $b_1 = \bar{y} - b_2\bar{x}$.

Fitted Values and Residuals

For each observation *i*, OLS produces:

Fitted value (the model's prediction for household *i*):

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

Residual (the prediction error for household *i*):

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

$\Longrightarrow$ Every data point decomposes as:

$$y_i = \underbrace{\hat{y}_i}_{\text{explained}} + \underbrace{\hat{e}_i}_{\text{unexplained}}$$

Example: A Specific Household

Consider a household with income $x_i = 20$ (i.e., \$2,000/week) and actual food expenditure $y_i = 350$.

Fitted value:

$$\hat{y}_i = 83.42 + 10.21 \times 20 = 287.62$$

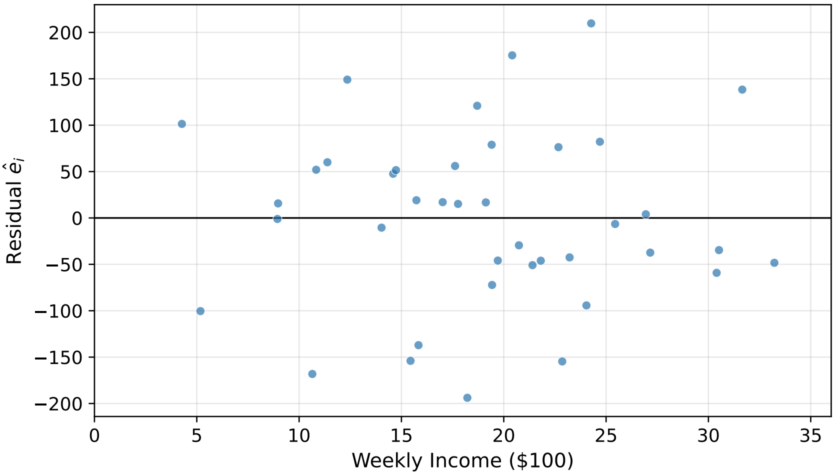
Residual:

$$\hat{e}_i = 350 - 287.62 = 62.38$$

This household spends \$62.38 **more** on food than the model predicts for their income level.

$\Longrightarrow$ The residual captures household-specific factors (tastes, family composition, etc.) that the model does not explain.

Visualizing the Residuals



The residuals scatter around zero with no obvious pattern. If you see a pattern (e.g., a fan shape, a curve), that suggests the model is missing something.

Estimators vs. Estimates

Estimator: the formula $b_2 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$ (a **random variable**).

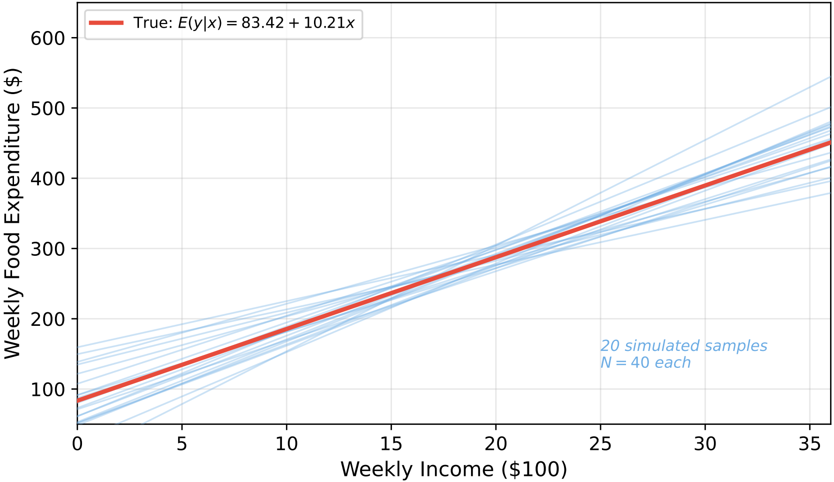
Estimate: the number $b_2 = 10.21$ from our particular sample (a **fixed number**).

Why is b_2 a random variable?

- The formula depends on the sample values $y_1, \ldots, y_N$
- A different random sample of 40 households would give different y_i's
- $\implies$ A different sample gives a **different estimate** b_2

The estimator has a probability distribution, a mean, and a variance. Understanding these properties is the subject of the next few topics.

Sampling Variation: Different Samples, Different Estimates



Each sample of $N = 40$ households from the same population produces a different fitted line. The slopes cluster around the true β, but no single sample hits it exactly.

Ten Hypothetical Samples

The true parameter values are $\beta_1 = 83.42$ and $\beta_2 = 10.21$. If we could repeatedly draw samples of $N = 40$ from the same population:

Sample	b_1	b_2
1	93.64	8.24
2	91.62	8.90
3	126.76	6.59
4	55.98	11.23
5	87.26	9.14
6	122.55	6.80
7	91.95	9.84
8	72.48	10.50
9	90.34	8.75
10	128.55	6.99
Average	96.11	8.70

The estimates bounce around. Ten samples is too few for the average to converge to β_1 and β_2, but

What We Have So Far

We started with a scatter plot and a question: which line?

The least squares principle gave us a criterion: minimize the sum of squared residuals.

The OLS formulas gave us the answer: b_1 and b_2.

But different samples give different answers.

$\Longrightarrow$ So next we ask: how reliable is OLS?

What Comes Next

We now have the OLS formulas and can compute estimates from any sample. But several questions remain:

1. Is b_2 **unbiased**? Does $\mathrm{E}(b_2) = \beta_2$?
2. How **precise** is b_2? What is $\mathrm{Var}(b_2)$?
3. Is OLS the **best** we can do, or is there a better estimator?
4. How do we quantify the **uncertainty** in our estimates?

$\Longrightarrow$ Topics 7–8 address these questions using assumptions SR1–SR5 and the Gauss–Markov theorem (normality, SR6, is not needed here).

Thank you!
jakeanderson@g.ucla.edu