Estimating the Error Variance and Prediction

How Precise Are Our Estimates, and What Can We Predict?

Jake Anderson

March 21, 2026

The Precision Question

We estimated that each \$100 of income raises food spending by \$10.21.

But is that estimate **reliable**? Could it easily have been \$5 or \$15?

Right now, we have no way to answer this. We need a measure of **precision** for our estimates.

$\Longrightarrow$ Today: how to quantify how much b_2 would bounce around across repeated samples.

Outline

Outline

What We Have So Far

From the food expenditure data ($N = 40$ households):

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

We showed that b_2 is:

- **Unbiased**: $\mathrm{E}(b_2) = \beta_2$ (on average, we hit the truth)
- **BLUE**: smallest variance among all linear unbiased estimators

And we derived its variance:

$$\mathrm{Var}(b_2) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

Problem: σ^2 is unknown. We cannot compute $\mathrm{Var}(b_2)$ without it.

$\Longrightarrow$ We need to **estimate** σ^2 from the data.

Outline

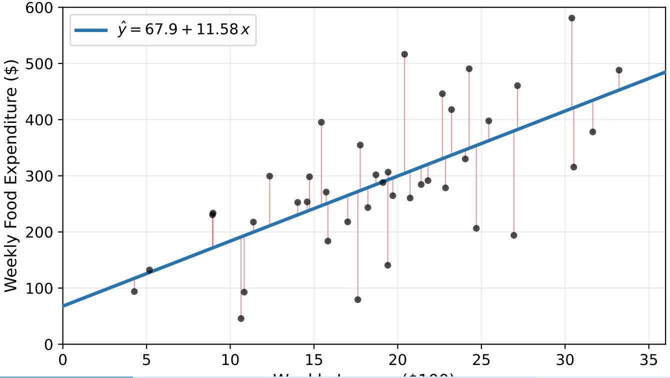
The Idea: Use the Residuals

Recall: $\sigma^2 = \text{Var}(e_i) = \text{E}(e_i^2)$ (since $\text{E}(e_i) = 0$ by SR2).

A natural estimator would be $\sum e_i^2 / N$, but we never observe the true errors e_i.

We *do* observe the **residuals**:

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



Why Not Divide by N?

Why not just use $\sum \hat{e}_i^2 / N$?

Because the residuals $\hat{e}_i$ are **not** the true errors e_i. OLS *chooses* b_1 and b_2 to minimize $\sum \hat{e}_i^2$. Any other coefficients, including the true β_1 and β_2, would produce a larger sum.

$\Longrightarrow$ The sum of squared residuals is smaller than $\sum e_i^2$: the squared residuals underestimate the true squared errors on average.

$$\sum \hat{e}_i^2 \leq \sum e_i^2$$

Dividing by N gives an estimator that is biased **downward**.

Correction: divide by $N - 2$ instead of N.

Degrees of Freedom: Why $N - 2$?

We estimated **two** parameters (b_1 and b_2) from the data.

This "uses up" 2 degrees of freedom:

- OLS forces $\sum \hat{e}_i = 0$ (one constraint from estimating b_1)
- OLS forces $\sum x_i \hat{e}_i = 0$ (one constraint from estimating b_2)

$\Longrightarrow$ Only $N - 2$ residuals are "free" to vary. The other 2 are pinned down by these constraints.

Analogy: sample variance uses $N - 1$ because estimating $\bar{x}$ uses 1 df. Here, estimating a line uses 2 df.

The Estimator $\hat{\sigma}^2$

The unbiased estimator of the error variance is:

$$\boxed{\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{e}_i^2}{N-2}}$$

- $E(\hat{\sigma}^2) = \sigma^2$ (unbiased)
- $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ is the **standard error of the regression**

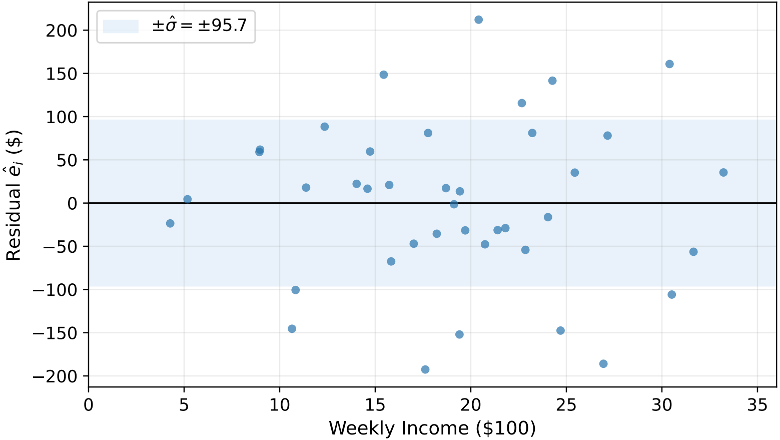
Food expenditure example (Example 2.5):

$$\hat{\sigma}^2 = \frac{304{,}505.2}{40-2} = \frac{304{,}505.2}{38} = 8{,}013.29$$

$$\hat{\sigma} = \sqrt{8{,}013.29} = 89.52$$

$\Longrightarrow$ Typical deviation of actual food spending from the regression line is about $89.52.

Visualizing the Residual Spread



$\hat{\sigma}$ estimates the typical spread of y around the fitted line. Under homoskedasticity (SR3), this spread is the same at every income level.

Outline

From $\hat{\sigma}^2$ to the Standard Error of b_2

Now we can estimate the variance of b_2. Replace σ^2 with $\hat{\sigma}^2$:

$$\widehat{\mathrm{Var}}(b_2) = \frac{\hat{\sigma}^2}{\sum(x_i - \bar{x})^2}$$

The hat on Var means **estimated**: we plugged in $\hat{\sigma}^2$ for the unknown σ^2.

The **standard error** is the square root:

$$\mathrm{se}(b_2) = \sqrt{\widehat{\mathrm{Var}}(b_2)} = \frac{\hat{\sigma}}{\sqrt{\sum(x_i - \bar{x})^2}}$$

(The intercept variance $\widehat{\mathrm{Var}}(b_1)$ has a similar formula; software computes it automatically.)

Anatomy of se(b_2)

$$\text{se}(b_2) = \frac{\hat{\sigma}}{\sqrt{\sum (x_i - \bar{x})^2}}$$

Two pieces, each with a clear interpretation:

Numerator: $\hat{\sigma}$ (noisiness of the data)

- More noise in y around the regression line $\implies$ harder to pin down the slope
- $\hat{\sigma}$ large $\implies$ se(b_2) large $\implies$ less precise estimate

Denominator: $\sqrt{\sum (x_i - \bar{x})^2}$ (spread of the x-values)

- More variation in income $\implies$ easier to see how y responds to x
- Larger $\sum (x_i - \bar{x})^2 \implies$ smaller se(b_2) $\implies$ more precise

Also: larger N increases $\sum (x_i - \bar{x})^2$, so more data improves precision.

Food Expenditure: Computing the Standard Errors

From Example 2.5:

- $\hat{\sigma}^2 = 8{,}013.29$
- $\sum(x_i - \bar{x})^2 = 1{,}828.79$

$$\widehat{\text{Var}}(b_2) = \frac{8{,}013.29}{1{,}828.79} = 4.382$$

$$\text{se}(b_2) = \sqrt{4.382} = 2.093$$

Reported together (with $\text{se}(b_1) = 43.41$ from software):

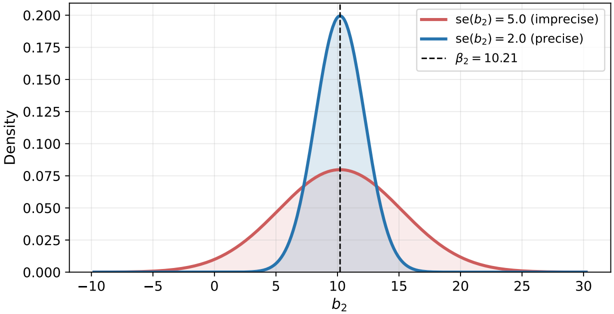
$$\hat{y} = \underset{(43.41)}{83.42} + \underset{(2.093)}{10.21}\, x$$

The numbers in parentheses below the coefficients are standard errors. This is standard notation in econometrics.

Interpreting the Standard Error of b_2

We estimated $b_2 = 10.21$ with $\text{se}(b_2) = 2.093$.

The standard error tells us: if we repeated this study many times (different random samples of 40 households), the estimates of b_2 would have a standard deviation of approximately 2.093.



$\Longrightarrow$ Smaller $\text{se}(b_2)$ means the sampling distribution is more concentrated around β_2. Our estimates are more **reliable**.

What Makes se(b_2) Small?

$$\text{se}(b_2) = \frac{\hat{\sigma}}{\sqrt{\sum(x_i - \bar{x})^2}}$$

Three levers for more precise slope estimates:

Factor	**Change**	**Effect on se(b_2)**
σ (noise)	Lower noise in y	Smaller se
$\sum(x_i - \bar{x})^2$	More income variation	Smaller se
N (sample size)	More observations	Smaller se

In practice, we control N (collect more data) and $\sum(x_i - \bar{x})^2$ (sample households with diverse incomes). We rarely control σ.

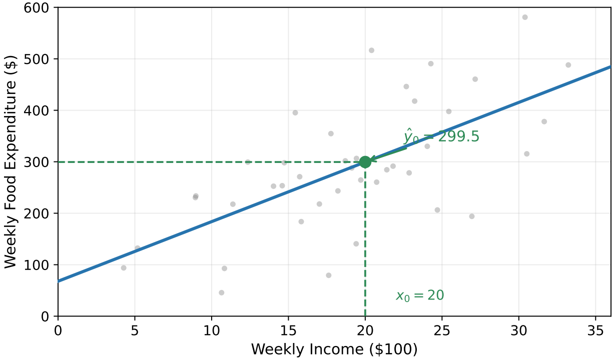
Outline

Using the Fitted Model for Prediction

A household earns \$2,000/week ($x_0 = 20$ in \$100 units). How much do we predict they spend on food?

Point prediction:

$$\hat{y}_0 = b_1 + b_2 x_0 = 83.42 + 10.21 \times 20 = 287.62$$



Two Sources of Prediction Uncertainty

Our prediction $\hat{y}_0 = 287.62$ is a single number. How confident should we be?

There are two distinct sources of uncertainty:

1. Estimation uncertainty (the line itself is uncertain)

- We estimated b_1 and b_2 from a sample of 40 households
- A different sample would give a different line
- Captured by $\text{se}(b_1)$ and $\text{se}(b_2)$

2. Inherent variability (individuals deviate from the line)

- Even if we knew the true line perfectly, any individual household's spending differs from $E(y|x)$ by the error e_i
- Captured by σ^2

$\Longrightarrow$ Prediction is always less precise than estimation. We formalize this in Chapter 4.

Outline

From Slope to Elasticity

Now that we have standard errors, we can say our slope is precise. But is it **economically meaningful**? Elasticity helps us answer that by converting to percentage terms.

The slope $b_2 = 10.21$ tells us: a \$100 increase in income raises food spending by \$10.21. But this is in **dollar terms**. Economists often prefer **percentage terms**.

From micro: elasticity $= (\partial y/\partial x) \cdot (x/y)$. Since $\partial y/\partial x = \beta_2$ in our linear model:

$$\varepsilon = \beta_2 \cdot \frac{x}{\mathrm{E}(y|x)}$$

In a linear model, elasticity **varies** along the regression line (it depends on where you evaluate it).

Convention: evaluate at the sample means $(\bar{x}, \bar{y})$.

Computing the Elasticity

Estimated elasticity at the means:

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

Using the food expenditure data:

$$\begin{aligned} \hat{\varepsilon} &= 10.21 \times \frac{19.60}{283.57} \\ &= 10.21 \times 0.0691 \\ &= 0.71 \end{aligned}$$

Interpretation: A 1% increase in income leads to approximately a 0.71% increase in food expenditure (at the mean income level).

Since $\hat{\varepsilon} < 1$, food is a **necessity**: spending grows, but slower than income.

Elasticity Is Not the Slope

A common mistake: reporting $b_2 = 10.21$ as "the elasticity."

Concept	Formula	Interpretation
Slope	$b_2 = 10.21$	\$100 more income $\implies$ \$10.21 more food
Elasticity	$\hat{\varepsilon} = 0.71$	1% more income $\implies$ 0.71% more food

The slope is constant in a linear model. The elasticity is not: it changes as x and $E(y|x)$ change.

$\implies$ Always specify where you evaluate the elasticity ("at the means," "at $x = 25$," etc.).

Outline

Wrapping Up Chapter 2

We started with a variance formula we couldn't use (σ^2 unknown).

So we estimated σ^2 from the residuals, dividing by $N - 2$ to correct the downward bias:

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{N - 2}$$

That gave us **standard errors**, which measure how much b_2 would vary across repeated samples:

$$\text{se}(b_2) = \frac{\hat{\sigma}}{\sqrt{\sum (x_i - \bar{x})^2}}$$

Smaller se $\implies$ less noise, more x-variation, or more data.

We can also use the fitted model for **point predictions** ($\hat{y}_0 = b_1 + b_2 x_0$) and convert the slope to an **elasticity** ($\hat{\varepsilon} = b_2 \cdot \bar{x}/\bar{y}$).

Next: now that we can report precision, what do we *do* with it? $\implies$ confidence intervals and hypothesis tests (Chapter 3).

Thank you!

jakeanderson@g.ucla.edu