Expectation, Variance, and Covariance

Jake Anderson

Outline

Outline

How Much Does Education Affect Wages?

This is one of the central questions in labor economics. You might think: just compare the wages of workers with different education levels.

But there is a problem. Workers who stay in school longer tend to have **fewer years of work experience** at any given age. Education and experience are tangled together.

So when you see that college graduates earn more, is that because of the education itself, or because of something else correlated with education?

$\Longrightarrow$ To sort this out, we need three tools: **expectation** (what is the average wage?), **variance** (how much do wages vary?), and **covariance** (how do education and experience move together?).

Outline

Expected Value: Definition

The **expected value** (or **mean**) of a discrete random variable X is a probability-weighted average of its possible values:

$$\mathbb{E}[X] = \sum_x x \cdot f(x)$$

where $f(x) = P(X = x)$ is the pdf.

Interpretation: If you repeated the experiment infinitely many times, the average outcome would converge to $\mathbb{E}[X]$.

Notation: $\mathbb{E}[X] = \mu_X = \mu$ (the **population mean**).

$\implies \mu$ is a *fixed number* describing the distribution. Do not confuse it with the sample mean $\bar{x}$, which varies from sample to sample.

Expected Value: Example

Suppose you roll a weighted die where:

x	1	2	3	4
$f(x) = P(X = x)$	0.1	0.2	0.3	0.4

$$\begin{aligned}\mathbb{E}[X] &= 1(0.1) + 2(0.2) + 3(0.3) + 4(0.4) \\ &= 0.1 + 0.4 + 0.9 + 1.6 \\ &= 3.0\end{aligned}$$

$\Longrightarrow$ The expected value need not be a value X can actually take. It is the long-run average, not a prediction for any single draw.

Rules for Expected Values

These rules are the foundation for every derivation in this course.

Rule 1: Linearity. For constants a and b:

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$$

Rule 2: Additivity. For any two random variables X and Y:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

This holds **whether or not** X and Y are independent.

Rules 1+2 combined: For constants a, b, c:

$$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$$

Linearity in Action: Hourly Wages

Let H be the number of hours a worker puts in per week, with $\mathbb{E}[H] = 35$.

If the hourly wage is $25, weekly earnings are $W = 25 \cdot H$:

$$\mathbb{E}[W] = \mathbb{E}[25 \cdot H] = 25 \cdot \mathbb{E}[H] = 25(35) = \$875$$

If you also get a flat $50 weekly bonus, earnings are $W = 25H + 50$:

$$\begin{aligned}\mathbb{E}[W] &= \mathbb{E}[25H + 50] \\ &= 25 \cdot \mathbb{E}[H] + 50 = 875 + 50 = \$925\end{aligned}$$

$\implies$ Constants scale and shift the expected value. No surprises here. The interesting part starts with *spread*.

Wages Callback: Expected Wages by Education

Let W be hourly wages and let *Educ* denote years of education. Suppose:

$$\mathbb{E}[W \mid \mathit{Educ} = 12] = \$18, \qquad \mathbb{E}[W \mid \mathit{Educ} = 16] = \$30$$

The \$12 gap looks like the "return to education." But is it?

Workers with 16 years of education (college grads) entered the workforce later. At age 30, they have about 8 years of experience. Workers with 12 years (high school grads) have about 12 years.

$\Longrightarrow$ The raw wage gap reflects *both* education and experience differences. To separate these effects, we will need covariance. But first, we need a way to measure **spread**.

Expected Value of a Product

What about $\mathbb{E}[XY]$?

If X and Y are independent:

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

The expectation “factors” cleanly.

If X and Y are dependent:

$$\mathbb{E}[XY] \neq \mathbb{E}[X] \cdot \mathbb{E}[Y] \quad \text{in general}$$

The gap between $\mathbb{E}[XY]$ and $\mathbb{E}[X]\mathbb{E}[Y]$ has a name. We will get to it shortly.

Outline

Variance: Definition

The **variance** measures how spread out a distribution is around its mean:

$$\mathrm{Var}(X) = \sigma_X^2 = \mathbb{E}\left[(X - \mu)^2\right]$$

Shortcut formula (usually easier to compute):

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mu^2 = \mathbb{E}[X^2] - [\mathbb{E}[X]]^2$$

The **standard deviation** $\sigma_X = \sqrt{\mathrm{Var}(X)}$ has the same units as X.

Interpretation: Larger variance $\implies$ the values of X are more spread out around μ. Variance is always ≥ 0, and equals zero only when X is constant.

Variance: Example

Using the same distribution as before ($\mu = 3$):

x	1	2	3	4
$f(x)$	0.1	0.2	0.3	0.4

Step 1: Compute $\mathbb{E}[X^2]$:

$$\mathbb{E}[X^2] = 1^2(0.1) + 2^2(0.2) + 3^2(0.3) + 4^2(0.4) = 0.1 + 0.8 + 2.7 + 6.4 = 10$$

Step 2: Apply the shortcut:

$$\text{Var}(X) = \mathbb{E}[X^2] - \mu^2 = 10 - 9 = 1$$

$\implies \sigma_X = \sqrt{1} = 1$. On average, values of X deviate from the mean by about 1 unit.

Variance Under Linear Transformation

The rule you will use constantly:

$$\text{Var}(aX + b) = \text{Var}(aX) + \text{Var}(b) = a^2\,\text{Var}(X) + 0 = a^2\,\text{Var}(X)$$

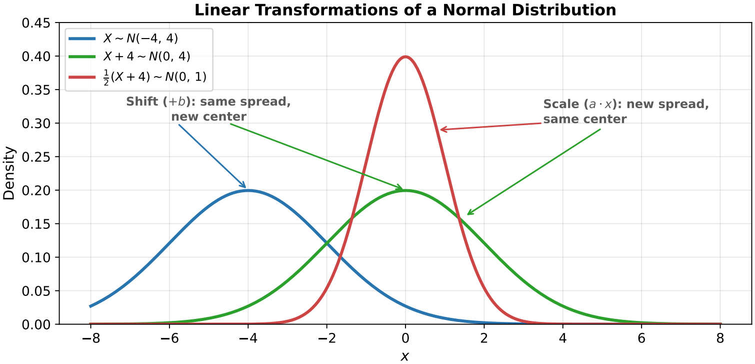
Two things to notice:

1. Adding a constant b **does not change spread**. Shifting a distribution left or right leaves the variance unchanged.
2. The multiplicative constant is **squared**. Not a, but a^2.

Intuition: If you double every value ($a = 2$), deviations from the mean also double, so squared deviations quadruple.

Scaling vs. Shifting: Visualized

Start with $X \sim N(-4, 4)$. First shift, then scale:



- **Shift** $(X + 4)$: moves the center to $\mu = 0$, but the spread stays at $\sigma^2 = 4$
- **Scale** $(\frac{1}{2}(X + 4))$: shrinks the spread to $\sigma^2 = 1$, center stays at $\mu = 0$

$\Longrightarrow$ Adding b changes location, not scale. Multiplying by a changes scale.

Variance Under Linear Transformation: Derivation

Why does this work? Let $Y = aX + b$, so $\mu_Y = a\mu_X + b$.

The deviation from the mean:

$$Y - \mu_Y = (aX + b) - (a\mu_X + b) = a(X - \mu_X)$$

The additive constant cancels. What remains is the scaling factor a.

Squaring and taking expectations:

$$\text{Var}(Y) = \mathbb{E}[(Y - \mu_Y)^2] = \mathbb{E}[a^2(X - \mu_X)^2] = a^2\mathbb{E}[(X - \mu_X)^2] = a^2 \text{Var}(X)$$

Variance: Wages in Different Units

Suppose hourly wages have $\text{Var}(W) = 64$ (in dollars^2), so $\sigma_W = \$8$.

Convert to thousands of dollars: $W_k = W/1000 = (1/1000) \cdot W$:

$$\text{Var}(W_k) = \left(\frac{1}{1000}\right)^2 \text{Var}(W) = \frac{64}{1{,}000{,}000} = 0.000064$$

$$\sigma_{W_k} = \$0.008 \text{ (thousands)} = \$8$$

Everyone gets a \$2 raise (a shift, $W_{\text{new}} = W + 2$):

$$\text{Var}(W + 2) = \text{Var}(W) = 64$$

$\implies$ The raise shifts everyone's wage up equally, so the spread does not change. Variance is driven by the random part, not by constants.

Wages Callback: Why Do Similar Workers Earn Different Amounts?

Two workers both have 16 years of education. One earns \$28/hr, the other \$35/hr.

The variance of wages *within* an education group tells you how much individual wages scatter around the group mean.

Sources of within-group wage variation:

- Different years of experience
- Different industries or occupations
- Ability, negotiation, luck

$\Longrightarrow$ Variance tells us there is a lot we cannot explain with education alone. Can we do better by also accounting for experience? That depends on how education and experience **covary**.

Outline

Covariance: Definition

The **covariance** measures the direction of *linear* association between two random variables:

$$\text{Cov}(X, Y) = \sigma_{XY} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

Shortcut formula:

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mu_X \mu_Y = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Sign interpretation:

- $\text{Cov}(X, Y) > 0$: X and Y tend to move **in the same direction**
- $\text{Cov}(X, Y) < 0$: X and Y tend to move **in opposite directions**
- $\text{Cov}(X, Y) = 0$: no *linear* association (there may be nonlinear dependence)

$\implies$ The magnitude of $\text{Cov}(X, Y)$ depends on the units of X and Y, so it does not tell you the *strength* of association. For that, we need correlation.

The $\mathbb{E}[XY]$ Formula, Revisited

Now we can name the gap from earlier. The shortcut formula rearranges to:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] + \text{Cov}(X, Y)$$

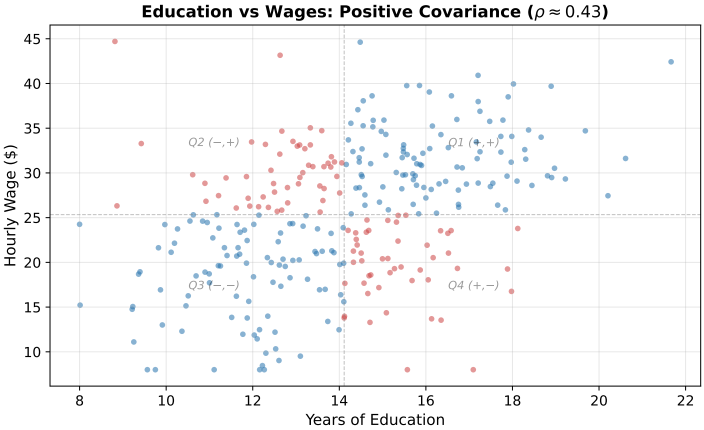
If X and Y are independent: $\text{Cov}(X, Y) = 0$, so $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

If X and Y are dependent: The covariance term is the "correction" for dependence.

$\implies$ This formula shows why independence simplifies so many proofs: it lets you factor expectations of products.

Covariance: Education and Wages

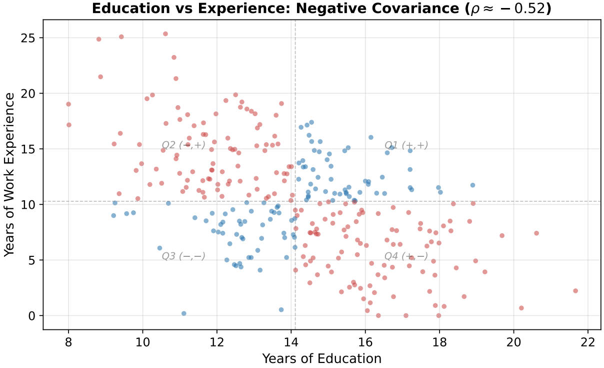
Education and wages tend to move together. Workers with more education earn more on average.



$\text{Cov}(Educ, W) > 0$. Most points cluster in Q1 and Q3 (same-direction deviations from the means).

Covariance: Education and Experience

Here is the key applied insight. At a given age, workers who stayed in school longer have fewer years of work experience.



$\text{Cov}(Educ, Exper) < 0$. This negative covariance is exactly what creates problems when we try to estimate the return to education.

Covariance: Properties

1. Symmetry:

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)$$

2. Variance is a special case:

$$\text{Cov}(X, X) = \mathbb{E}[XX] - \mathbb{E}[X]\mathbb{E}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \text{Var}(X)$$

3. Independence implies zero covariance:
If X and Y are independent, then $\text{Cov}(X, Y) = 0$.

The reverse does not hold: $\text{Cov}(X, Y) = 0$ does **not** guarantee independence. Covariance only detects *linear* association.

$\implies$ Independence is stronger than zero covariance. We will see this distinction again when we study regression assumptions.

Correlation: Unitless Covariance

Covariance depends on the units of X and Y. To get a **unitless** measure that captures both direction and **strength**, standardize by the standard deviations:

$$\rho_{XY} = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

Properties:

- $-1 \leq \rho \leq 1$
- $|\rho| = 1$: perfect linear relationship ($Y = a + bX$ exactly)
- $\rho = 0$: no linear association
- $\rho > 0$: positive linear association; $\rho < 0$: negative

Inverting the formula:

$$\text{Cov}(X, Y) = \rho_{XY} \cdot \sigma_X \cdot \sigma_Y$$

$\Longrightarrow$ This form is useful when you know the correlation and standard deviations.

Correlation: Education and Experience

In the U.S. labor market, the correlation between years of education and years of experience (for a given age) is typically around $\rho \approx -0.4$ to -0.6.

Variable pair	ρ	Intuition
Education, Wages	$\approx +0.4$	More school, higher pay
Experience, Wages	$\approx +0.3$	More experience, higher pay
Education, Experience	≈ -0.5	More school, less experience

$\implies$ Education and experience are both positively associated with wages, but negatively associated with *each other*. This creates a tangle we need to unravel.

Wages Callback: Why Covariance Complicates Things

We know two facts:

- More education $\implies$ higher wages ($\text{Cov}(Educ, W) > 0$)
- More education $\implies$ less experience ($\text{Cov}(Educ, Exper) < 0$)
- More experience $\implies$ higher wages ($\text{Cov}(Exper, W) > 0$)

So when we compare workers with different education levels, we are also comparing workers with different experience levels. The raw wage gap between a college graduate and a high school graduate reflects *both* differences.

$\implies$ Because education and experience are negatively correlated, we cannot isolate the effect of one without accounting for the other. Later in the course, we will learn how to do this. For now, the takeaway: **when variables are correlated, you cannot study one in isolation**.

Outline

Variance of a Sum: The Formula

For constants a and b:

$$\text{Var}(aX + bY) = a^2\,\text{Var}(X) + b^2\,\text{Var}(Y) + 2ab\;\text{Cov}(X, Y)$$

$\implies$ The **cross-term** $2ab\;\text{Cov}(X, Y)$ is what makes the variance of a combination depend on the relationship between the variables.

Without the cross-term, you would always predict $\text{Var} = a^2\,\text{Var}(X) + b^2\,\text{Var}(Y)$. This is correct only when $\text{Cov}(X, Y) = 0$.

Variance of a Sum: Derivation

Let $W = aX + bY$, so $\mu_W = a\mu_X + b\mu_Y$:

$$W - \mu_W = a(X - \mu_X) + b(Y - \mu_Y)$$

Squaring:

$$(W - \mu_W)^2 = a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 + 2ab(X - \mu_X)(Y - \mu_Y)$$

Taking expectations of both sides:

$$\text{Var}(W) = a^2\,\text{Var}(X) + b^2\,\text{Var}(Y) + 2ab\;\text{Cov}(X, Y)$$

$\implies$ The cross-term arises from the product of deviations, which is exactly what covariance measures.

Special Cases

Sum ($a = 1, b = 1$):

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X, Y)$$

Difference ($a = 1, b = -1$):

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\,\text{Cov}(X, Y)$$

Note: variances still **add**. Only the covariance term changes sign.

Independent case ($\text{Cov}(X, Y) = 0$):

$$\text{Var}(aX + bY) = a^2\,\text{Var}(X) + b^2\,\text{Var}(Y)$$

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$$

$\implies$ "The variance of a sum is the sum of the variances" is only true when $\text{Cov} = 0$. Forgetting the covariance term is one of the most common errors in this course.

Outline

A Simple Wage Model

Suppose wages depend linearly on education and experience:

$$W = \beta_0 + \beta_1 Educ + \beta_2 Exper + \varepsilon$$

Expected wages:

$$\mathbb{E}[W] = \beta_0 + \beta_1\,\mathbb{E}[Educ] + \beta_2\,\mathbb{E}[Exper]$$

Now consider the variance of the "predicted" part, $\hat{W} = \beta_1 Educ + \beta_2 Exper$:

$$\text{Var}(\hat{W}) = \beta_1^2\,\text{Var}(Educ) + \beta_2^2\,\text{Var}(Exper) + 2\beta_1\beta_2\,\text{Cov}(Educ, Exper)$$

$\implies$ How much wage variation we can explain depends on the variance of each predictor *and* how they covary.

What If You Forgot the Covariance Term?

Naive approach: Ignore the cross-term and compute

$$\text{Var}(\hat{W}) \stackrel{?}{=} \beta_1^2 \text{ Var}(Educ) + \beta_2^2 \text{ Var}(Exper)$$

Suppose $\beta_1 = 3$ (\$3/yr of education), $\beta_2 = 1.5$ (\$1.50/yr of experience), $\text{Var}(Educ) = 4$, $\text{Var}(Exper) = 16$:

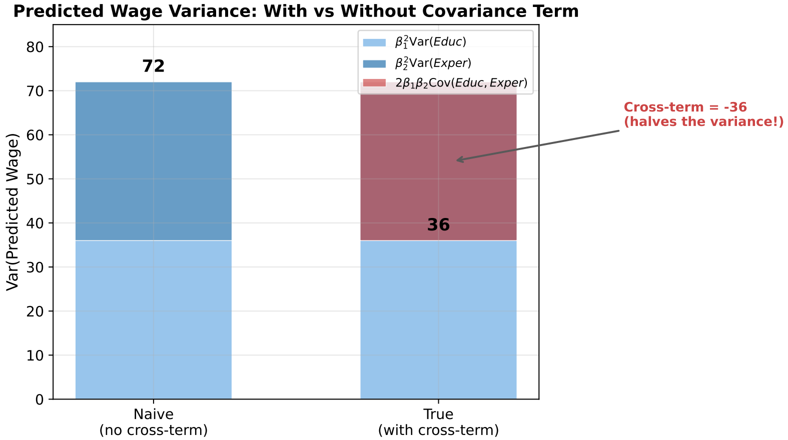
$$\text{Naive prediction:} \quad 9(4) + 2.25(16) = 36 + 36 = 72$$

Actual formula with $\text{Cov}(Educ, Exper) = -4$:

$$\text{Var}(\hat{W}) = 72 + 2(3)(1.5)(-4) = 72 - 36 = 36$$

$\implies$ The naive estimate is **double** the true value. Because education and experience move in opposite directions, they partially offset each other, and predicted wages are less variable than you would expect.

Predicted Wage Variance: Visualized



$\implies$ The negative covariance between education and experience shrinks the variance of predicted wages. Ignoring the cross-term leads you to overestimate how spread out wages should be.

Outline

Formulas to Internalize

Concept	Formula
Expected value	$\mathbb{E}[X] = \sum_x x\, f(x)$
Linearity of $\mathbb{E}$	$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$
Variance	$\text{Var}(X) = \mathbb{E}[X^2] - [\mathbb{E}[X]]^2$
Variance scaling	$\text{Var}(aX + b) = a^2 \,\text{Var}(X)$
Covariance	$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
Correlation	$\rho = \text{Cov}(X, Y)/(\sigma_X \sigma_Y)$
Var of a sum	$\text{Var}(aX + bY) = a^2 \,\text{Var}(X) + b^2 \,\text{Var}(Y) + 2ab \,\text{Cov}(X, Y)$

$\Longrightarrow$ These seven formulas are the algebra behind every estimator, standard error, and test statistic in this course. You will use them in every problem set.

Common Pitfalls

1. $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$, not $a\,\mathrm{Var}(X)$. The constant is **squared**.
2. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ is only true when $\mathrm{Cov}(X, Y) = 0$.
3. $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X, Y)$. Variances **add**; do not subtract $\mathrm{Var}(Y)$.
4. $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ only when X and Y are **independent**.
5. Zero covariance does **not** imply independence. Covariance only detects linear association.

Looking Ahead

These tools connect directly to regression:

- The population regression slope is $\beta_1 = \text{Cov}(X, Y)/\text{Var}(X)$. The OLS estimator $\hat{\beta}_1$ replaces these with sample analogs.
- Unbiasedness proofs use $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
- Standard errors use $\text{Var}(aX + b) = a^2\,\text{Var}(X)$
- Omitted variable bias uses the covariance between regressors

$\Longrightarrow$ Every derivation in Chapters 2 through 16 starts with the rules from today. The education/experience example is an informal version of what we will solve using Ordinary Least Squares (OLS) regression.

Thank you!

jakeanderson@g.ucla.edu

Outline

Derivation: Covariance Shortcut Formula

Start from the definition:

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

Expand the product:

$$= \mathbb{E}[XY - X\mu_Y - \mu_X Y + \mu_X\mu_Y]$$

Apply linearity of $\mathbb{E}$ (distribute term by term):

$$= \mathbb{E}[XY] - \mu_Y\,\mathbb{E}[X] - \mu_X\,\mathbb{E}[Y] + \mu_X\mu_Y$$

Substitute $\mathbb{E}[X] = \mu_X$ and $\mathbb{E}[Y] = \mu_Y$:

$$= \mathbb{E}[XY] - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y$$

$$\boxed{\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]}$$