LECTURE 1.2

Expectation and Conditional Means

ECON30130 Econometrics I

Sam Deegan
University College Dublin

Autumn 2027

Where We Are

Last Week's Lecture

Lecture 1.1 set the module's question and the vocabulary for describing data:

1. **Populations and Samples.**
   The question is about a population; the data is a sample.
2. **Centre and Spread.**
   Where a variable sits, and how far people scatter from its mean.
3. **Moments and Shape.**
   Which tail is longer, and how much sits far out.
4. **Describing Two Variables.**
   Covariance, correlation, and what one number cannot show.

Lecture Plan

1. **Expectation and Its Algebra.**
   An expectation is a population average, and three rules cover sums and constants.

2. **The Conditional Expectation Function.**
   The CEF gives average earnings at each level of schooling, and a fitted line estimates it.

3. **Estimators and Estimates.**
   An estimator is a rule fixed before the draw; an estimate is the number one sample gives.

4. **Unbiasedness, Consistency and Efficiency.**
   Unbiased, consistent and efficient are three separate properties.

SECTION 1

Expectation and Its Algebra

Defining the Expectation of a Random Variable

We take an expectation because nobody can say what one random draw earns. The expectation weights every possible value by how often it occurs:

$$E(y) = \sum_i y_i p_i \tag{1}$$

- y_i: one value that hourly earnings can take
- p_i: the share of the population at that value

> **The Brackets Take an Expression**
> $E(\cdot)$ is an operator: the bracket returns the population average of whatever sits inside it. $E(y)$ reads as "the average of y".

Notation for the Population Mean and the Sample Mean

From here every summary comes as a pair, one population and one sample:

> **The Population Mean Is Fixed, the Sample Mean Is Not**
> $E(y)$ is one fixed number, a property of the distribution, and equals 11.52 here. $\bar{y}$ is the sample average, 12.25 in our 120. Draw a different 120 and $\bar{y}$ is a different number.

Wooldridge calls $E(y)$ the population mean and writes it μ_y. The two are one number, and today's last two sections use the Greek form.

The Three Rules of Expectation Algebra

Everything this module measures is an average of some combination of variables. Three rules deliver those averages without working out a new distribution:

$$E(c) = c \qquad E(a + b\,y) = a + b\,E(y) \qquad E(y + z) = E(y) + E(z) \tag{2}$$

- c: a fixed number with no distribution of its own
- a: a shift added identically to every draw
- b: a scaling applied identically to every draw
- z: a second random variable

Only What Varies Keeps Its Expectation

An expectation averages across draws, so only what changes between draws keeps its E:

- **A Constant:** A fee or a tax rate is the same in every draw, so its average is itself.
- **A Scaling:** Report earnings in cents and the average rescales by 100.
- **A Sum:** The third rule holds whether y and z move together or not.

> **A Constant Need Not Be Known**
> A population slope is one fixed number that nobody knows. The slope still comes out of an expectation, because it is the same in every draw.

Variance Written as an Expectation

The average of the squares never falls below the square of the average. The gap between those two is the variance:

$$\mathrm{Var}(y) = E\Big(\underbrace{\big(y - E(y)\big)^2}_{\text{squared distance from the mean}}\Big) = E(y^2) - \big(E(y)\big)^2 \tag{3}$$

Var Is a Population Operator

$\mathrm{Var}(\cdot)$ returns one population number from whatever sits in its brackets. Lecture 1.1's s_y^2 is the sample version.

Covariance Written as an Expectation

Co-movement of earnings and schooling (x) is an average too, and $\text{Cov}(y, y) = \text{Var}(y)$:

$$\text{Cov}(y, x) = E\Big(\underbrace{\big(y - E(y)\big)\big(x - E(x)\big)}_{\text{the cross-product}}\Big) = E(yx) - E(y)\,E(x) \qquad (4)$$

> **Cov Is a Population Operator**
> $\text{Cov}(\cdot, \cdot)$ returns one population number from the two variables in its brackets. Lecture 1.1's s_{xy} is the sample version.

Why Expectations Do Not Pass Through a Product

An expectation passes through a product exactly when the covariance is zero:

$$E(yx) \;=\; E(y)\,E(x) + \mathrm{Cov}(y, x) \tag{5}$$

Zero Covariance Does Not Make Two Variables Independent

Independent means knowing one variable tells you nothing about the other:

- **Independence Is Enough:** Independent variables always have zero covariance.
- **Zero Covariance Is Not:** If $E(y) = 0$ and $E(y^3) = 0$, then $\mathrm{Cov}(y, y^2) = 0$, yet y fixes y^2.

> **Schooling and Earnings Are Not Independent**
> Last lecture our 120 people gave a positive covariance between schooling and earnings. Schooling and earnings are therefore not independent.

The Variance of a Sum Includes a Covariance Term

Adding two random variables does not add their variances. Squaring a sum produces two equal cross terms, which combine into one doubled term:

$$\mathrm{Var}(y+z) = \mathrm{Var}(y) + \mathrm{Var}(z) + \underbrace{2\,\mathrm{Cov}(y,z)}_{\text{the covariance term}} \tag{6}$$

The Covariance Can Raise or Lower the Spread

A positive covariance makes the sum more variable, and a negative one less. The term drops out when y and z are uncorrelated; independence is not needed.

The Average of Schooling Times Earnings

Average schooling is 13.54 years and average earnings 11.52 euro. Multiplying those two averages gives 155.9.

> **Question**
> Is the average of schooling times earnings also 155.9?

The Average Product Exceeds the Product of Averages

The population covariance of schooling and earnings is 7.02, so the average product is larger:

$$E(yx) \;=\; E(y)\,E(x) + \mathrm{Cov}(y, x) \tag{7}$$

For Earnings Times Earnings the Gap Is the Variance
Put y in place of x and the gap is $\mathrm{Var}(y)$. The gap is zero only if everyone earns the same.

Worked example: our population

SECTION 2

The Conditional Expectation Function

Defining the Conditional Expectation of Earnings Given Schooling

A covariance says only that schooling and earnings vary together. The conditional expectation averages earnings within each level of schooling:

$$E(y \mid x) \tag{8}$$

- $E(y \mid x)$: the average of y among people with that x

The CEF Is a Function of Schooling

$E(y \mid x)$ reads as "the average of y given x", and that function is the CEF. The CEF returns one average at each level of schooling, and those averages need not increase.

The Same Average, Computed One Column at a Time

The conditional mean needs no new arithmetic, only Lecture 1.1's mean:

$$\bar{y}_x = \frac{1}{n_x} \underbrace{\sum_{i:x_i=x} y_i}_{\text{the column total}} \tag{9}$$

- $\bar{y}_x$: the average earnings of our 120 at that x

> **Only That Column Enters the Sum**
> The condition under the sigma keeps only the people whose schooling equals x, and n_x counts them. A condition replaces the usual 1 to n, so nothing sits above the sigma.

Earnings Distributions at Three Levels of Schooling

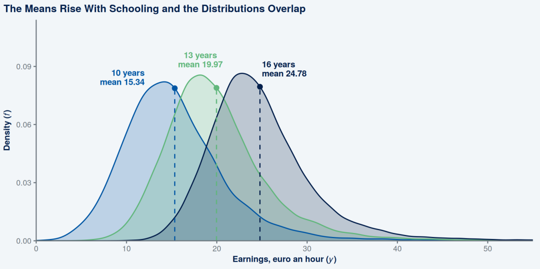

Figure 1: Population earnings at ten, thirteen and sixteen years of schooling.

> **The Three Distributions Overlap**
> About one in six people with sixteen years of schooling earns less than the average at thirteen. The spread around each mean is about 6 euro at all three.

The Conditional Expectation Function in This Population

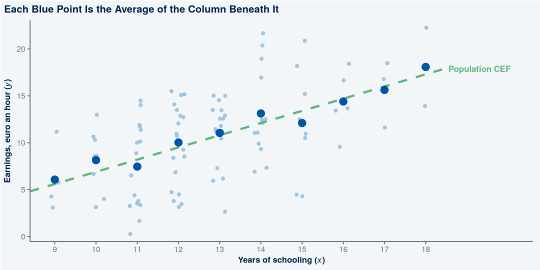

Figure 2: One hundred people drawn for this picture; each blue point is that column's average.

> **Column Averages Trace the CEF**
> Each blue point is the average of the people in that column. The green dashed line is the population CEF, $E(y \mid x)$, which is straight in this population.

One Group's Own Average Is Its Best Single Guess

Name one number for everyone with the same schooling, and score it by its average squared miss:

- **The Group's Own Average:** This guess gives the smallest average squared miss.
- **Any Other Guess:** Any other guess adds its squared distance from that average.

> **The Balance Point Is the Minimum**
> Deviations from a group's average sum to zero, as Lecture 1.1 showed. That zero sum is the condition for the smallest average squared miss.

Why the Minimum Sits at the Conditional Mean

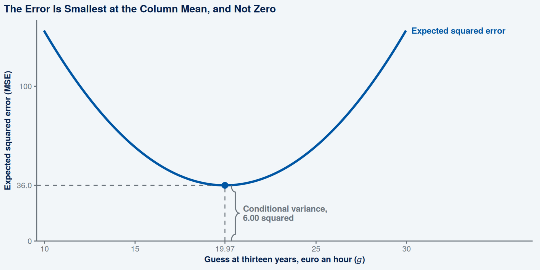

Figure 3: Expected squared error at thirteen years of schooling, against the guess made.

> **The Minimum Is the Conditional Variance**
> The curve is lowest at a guess of 19.97, the column's mean, at a height of 6.00^2. That height is the conditional variance. Earnings still vary inside the column.

The Conditional Mean Minimises Mean Squared Error

A rule $g(x)$ turns each level of schooling into a guess. The expression below picks the rule whose squared misses average out smallest, group by group:

$$E(y \mid x) \;=\; \arg\min_{g} \underbrace{E\big((y - g(x))^2 \big)}_{\text{the mean squared error}} \qquad \textbf{(10)}$$

Both Expected Squares Must Be Finite

The comparison covers every rule g with $E(g(x)^2)$ finite, straight or not. $E(y^2)$ must be finite too, or there is no expected squared error to compare.

What the Fitted Line in Lecture 2.1 Approximates

The CEF belongs to the population, and we only ever hold a sample:

- **Our 120:** Only a handful of our people sit at each level of schooling.
- **The Fitted Line:** Lecture 2.1 fits a line to those people to estimate the CEF.
- **Mastering Metrics:** Angrist & Pischke (2015) build regression theory on the CEF.

> **The Line Gives a Guess Where Few People Sit**
> A line uses all the columns at once, so the line still gives a guess where only a handful of people sit. This population's CEF is straight, and real ones need not be. Where the CEF bends, every straight line has a larger mean squared error.

Description Versus Causation in the Conditional Mean

The CEF reports two observed averages, and neither says schooling did anything to anybody:

$$E(y \mid x = x_H) \qquad\qquad E(y \mid x = x_L) \tag{11}$$

- x_H, x_L: a higher and a lower level of schooling, such as sixteen and thirteen years

> **Observing a Gap Is Not Explaining It**
> "Three more years raises earnings by 3.90" claims schooling changed those earnings. $E(y \mid x)$ only compares two groups of people who already differ in many ways. The causal claim is about years a person never spent in school.

Worked example: our population

What a Causal Claim Needs on Top of the CEF

Turning a gap in the CEF into an effect takes three more things:

- **A Mechanism:** A story has to say how schooling itself raises a person's pay.
- **Rival Stories:** Anything else raising both schooling and earnings has to be ruled out.
- **Chance:** The gap has to be too large for luck to explain.

No Regression Supplies the Mechanism

Lecture 1.3 measures the role of chance, and Lecture 3.1 controls for ability, one rival story. No technique in this module supplies the mechanism. Human capital theory supplies one: school raises skills, and employers pay for them.

SECTION 3

Estimators and Estimates

The Population Mean Is Not Directly Observable

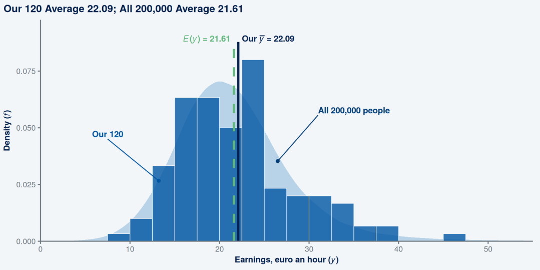

Figure 4: Population earnings pale, our 120 in solid blue, both means marked.

> **Only the Sample Mean Changes**
> Redraw the 120 and $\bar{y}$ leaves 22.09, while $E(y)$ holds at 21.61. The redraw changes our 120, not the distribution behind $E(y)$.

An Estimator Is a Rule Fixed Before the Draw

Every population number today is unobservable, and we hold only 120 people. The sample average is one rule for guessing $E(y)$ from those 120:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{12}$$

On our 120 people $\bar{y}$ comes to 12.25 euro an hour.

An Estimator Is the Rule, an Estimate Is the Number

An estimator turns a sample into a number, by a rule fixed beforehand. An estimate is the number one sample gives.

Reading a Bar, a Hat and a Greek Letter

Each quantity in this lecture has a population version and a sample version. Four marks on a letter tell you which version you are reading:

> **Bar, Roman, Greek and Hat**
> A bar means an average. A Roman letter is a sample quantity. A Greek letter is a population quantity. A hat means estimated from a sample. Whatever sits under the hat is the quantity being estimated.

Two Names Each for the Sample and Population Means

A hat needs a letter to sit on, so the population mean is written μ_y. Our 120 give $\bar{y}$, a sample quantity that estimates μ_y:

$$\hat{\mu}_y = \bar{y} \qquad \mu_y = E(y) \tag{13}$$

- $\hat{\mu}_y$: the estimate of μ_y from our 120

> **Mu-Hat Is a Sample Quantity**
> A hat on a Greek letter still marks a sample quantity. Draw a different 120 and $\hat{\mu}_y$ changes; μ_y does not.

Worked example: our 120 people

Why the Sample Mean Differs from the Population Mean

The module's claims are about μ_y, and our numbers come from 120 people. One random draw of 120 lands near the population mean, not on it:

$$\bar{y} - \mu_y \tag{14}$$

Worked example: our 120 people

The Sample Mean Concentrates on the Population Mean

Add people to a random sample and its mean concentrates on the population mean.

> **The Law of Large Numbers**
> That concentration is the law of large numbers, and it needs a random sample. The law says nothing about 240 people beating 120 on one draw.

At 120 people the gap of 0.73 is still there.

Four Candidate Estimators of the Population Mean

Each of these four rules could be written down before any data arrive:

- **Rule A:** The sample mean averages all 120 earnings.
- **Rule B:** The first-ten rule averages the first ten people drawn and discards the rest.
- **Rule C:** The constant rule reports 12.00 euro whatever the sample shows.
- **Rule D:** The trimmed mean averages the middle ninety per cent of earnings.

Rule C Is an Estimator

Rule C counts as an estimator because the rule is fixed before any data arrive. Rule C ignores the sample entirely, so no amount of data brings its answer closer to μ_Y.

Unbiased, Consistent and Efficient

Two rules can both centre on μ_y and still spread very differently. Three separate properties say which rule to prefer:

- **Unbiased:** At every n, the rule returns μ_y on average across samples.
- **Consistent:** As n grows without bound, the rule concentrates on μ_y.
- **Efficient:** Of two unbiased rules, the efficient one has the smaller sampling variance.

> **Only in the Linear Unbiased Class**
> The sample mean is least variable only among linear unbiased rules. The result also needs a common mean, a common variance and uncorrelated draws.

SECTION 4

Unbiasedness, Consistency and Efficiency

Unbiasedness Is About the Centre of the Distribution

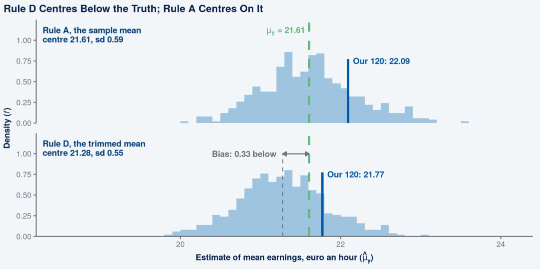

Figure 5: Rules A and D over 500 samples of 120, a repeated-sample ghost; the truth is 21.61.

> **Bias Is a Shifted Centre**
> Rule A centres on 21.61 and rule D centres 0.33 below it. Bias is the gap between a rule's centre and μ_y.

Efficiency Compares Two Unbiased Rules

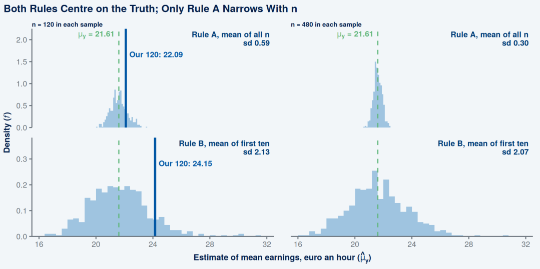

Figure 6: Rules A and B over 500 samples, a repeated-sample ghost; left n = 120, right n = 480.

Same Centre, Different Widths

Both rules centre near 21.61, and rule A's distribution is narrower than rule B's in both panels. Only rule A narrows as *n* grows, which is consistency.

The Four Rules Against the Three Properties

Rule A has all three properties, and each other rule lacks at least one:

- **Rule A:** The sample mean is unbiased, consistent and least variable among linear unbiased rules.
- **Rule B:** The first-ten rule is unbiased but inconsistent, because its spread never falls.
- **Rule C:** The constant rule never varies, and it sits 0.48 euro above μ_y.
- **Rule D:** The trimmed mean varies less than rule A and averages below μ_y.

Mean Squared Error Adds Bias to Variance

A rule's mean squared error (MSE) is its average squared distance from μ_y, and it splits in two:

$$\text{MSE} = \text{Var} + \text{Bias}^2 \tag{15}$$

MSE Ranks Biased and Unbiased Rules Together

Mean squared error can compare a biased rule with an unbiased one. Rule D can score lower on it than rule A.

Why Unbiasedness Alone Is Not Enough

Unbiasedness describes where the rule's distribution centres, not where one estimate lands:

- **Across Samples:** The rule returns μ_y on average over repeated draws.
- **In One Sample:** Unbiasedness does not promise that our 12.25 sits near μ_y.

> **Unbiased and Three Times Wider**
> Rule B is unbiased, and its spread is over three times rule A's. The extra width comes from averaging ten people instead of 120.

Choosing Among Three Imperfect Rules

You want average hourly earnings for Irish workers, and rule A is unavailable.

> **Question**
> Rule B, rule C or rule D: which would you defend, and on what ground?

Each Imperfect Rule Rests on a Different Property

Each defence of rule B, C or D rests on a different property:

- **Rule B:** The first-ten rule is unbiased, but its spread is over three times rule A's.
- **Rule C:** The constant rule never varies, but it sits 0.48 above μ_y.
- **Rule D:** The trimmed mean accepts a small bias for a smaller variance.

What Would Have to Be True?

Today's number is the conditional mean of earnings at each level of schooling.

> **Question**
> What would have to be true for this number to mean what we just said it means? And what is the plausible story where it isn't?

Reading the CEF as an Effect Assumes Stayers and Leavers Alike

Reading $E(y \mid x)$ as an effect is an assumption about who stayed in school:

- **What Must Hold:** Stayers and leavers differ in nothing else that changes earnings.
- **The Plausible Failure:** Angrist & Pischke (2015) doubt that more- and less-educated workers are equally able.

Binned Means in the CEF Explorer

The CEF explorer cuts schooling into one-year bins. Each bin gets one point, at its mean of earnings.

> **Question**
> With our 120 people, will the bin means look like a line?

Bin Means from Our 120 People Do Not Form a Line

The population CEF is straight, and bin means from our 120 are not:

- **Our 120:** With a handful of people per bin, the bin means scatter around the line.
- **All 200,000:** On the whole population, the lab's bin means fall close to the line.

Summary: Expectation, the CEF and Estimators

1. **Expectation and Its Algebra.**
   Three rules carry every derivation of an expectation, and none covers a product.

2. **The Conditional Expectation Function.**
   The CEF is the guess at earnings that minimises mean squared error. A conditional mean describes the population and explains nothing.

3. **Estimators and Estimates.**
   An estimator is a rule fixed before the draw, and our 120 give 12.25.

4. **Unbiasedness, Consistency and Efficiency.**
   Unbiased, consistent and efficient are three separate properties. Rule A, the sample mean, has all three.

Before Lecture 1.3

1. **Play:**
   The CEF explorer, widening the bin from one year to four.

2. **Apply:**
   Problem set 1, covering 1.1 and 1.2, with both appendix derivations.

3. **Review:**
   Wooldridge (2020), Math Refresher B-3 and B-4, then C-2 and C-3a.

Next Lecture

Rule A is the rule we use, and Lecture 1.3 asks how far its answer can sit from $E(y)$:

1. **The Sampling Distribution.**
   The standard error measures how far a sample mean typically sits from the truth.
2. **The Central Limit Theorem and Confidence Intervals.**
   Sample means are close to normal, so one estimate can carry a 95 per cent interval.
3. **Hypothesis Tests.**
   A test says what a claimed mean would produce, not whether the claim is true.

Thank you

sam.deegan@ucdconnect.ie
sam-deegan.com

BACKUP

Appendix

Notation Used in This Lecture

Symbol	What it is	First met
y	hourly earnings in euro	Lecture 1.1
y_i	one value that hourly earnings can take	Defining the Expectation of a Random Variable
y_i	person *i*'s hourly earnings	Lecture 1.1
p_i	the share of the population at that value	Defining the Expectation of a Random Variable
i	which person; in the expectation sum, which value	Lecture 1.1
x	years of schooling	Lecture 1.1
x_i	person *i*'s years of schooling	Lecture 1.1
z	a second random variable	The Three Rules of Expectation Algebra
c	a fixed number with no distribution of its own	The Three Rules of Expectation Algebra
a	a shift added identically to every draw	The Three Rules of Expectation Algebra
b	a scaling applied identically to every draw	The Three Rules of Expectation Algebra
n	the number of people in the sample, 120 here	Lecture 1.1

Notation Used in This Lecture

Symbol	What it is	First met
n_x	how many of our 120 have that x	The Same Average, Computed One Column at a Time
$g(x)$	a rule turning schooling into a guess	The Conditional Mean Minimises Mean Squared Error
$\bar{y}$	the sample average of y, 12.25	Notation for the Population Mean and the Sample Mean
$\bar{y}_x$	the average earnings of our 120 at that x	The Same Average, Computed One Column at a Time
μ_y	the Greek name for $E(y)$, 11.52	Notation for the Population Mean and the Sample Mean
$\hat{\mu}_y$	the estimate of μ_y from our 120	Two Names Each for the Sample and Population Means
a bar	an average	Lecture 1.1
a hat	estimated from a sample	Reading a Bar, a Hat and a Greek Letter
s_y^2	the variance of y in our 120	Lecture 1.1
s_{xy}	the covariance of x and y in our 120	Lecture 1.1
$E(\cdot)$	the population average of what is inside	Defining the Expectation of a Random Variable
$E(y)$	the population mean of earnings, 11.52	Defining the Expectation of a Random Variable

Notation Used in This Lecture

Symbol	What it is	First met
$E(y \mid x)$	the CEF	Defining the Conditional Expectation of Earnings Given Schooling
x_H, x_L	two levels of schooling	Description Versus Causation in the Conditional Mean
$\mathrm{Var}(\cdot)$	the population variance	Variance Written as an Expectation
$\mathrm{Cov}(\cdot, \cdot)$	the population covariance	Covariance Written as an Expectation
$\sum_{i=1}^{n}$	add over all n people	Lecture 1.1
$\sum_{i\,:\,x_i=x}$	add over one column	The Same Average, Computed One Column at a Time
$\arg\min_g$	the minimising g	The Conditional Mean Minimises Mean Squared Error
MSE	mean squared error	Mean Squared Error Adds Bias to Variance
Bias	a rule's centre minus μ_y	Unbiasedness Is About the Centre of the Distribution

Worked Example: Our Population's Average Product

The average product is the product of the averages plus the covariance:

$$E(yx) \;=\; E(y)\,E(x) + \mathrm{Cov}(y, x) \tag{16}$$

In our population the product of the averages is 155.94 and the covariance 7.02:

$$E(yx) \;=\; 155.94 + 7.02 \;=\; 162.96 \tag{17}$$

Back to the slide

Worked Example: Two Averages on the CEF

The CEF returns one average at each level of schooling:

$$E(y \mid x = x_H) \qquad E(y \mid x = x_L) \tag{18}$$

We take sixteen years for x_H and thirteen for x_L:

$$E(y \mid x = 16) \qquad E(y \mid x = 13) \tag{19}$$

In our population the CEF is $10.80 + 1.30\,(x - 13)$, so the two averages are:

$$E(y \mid x = 16) = 10.80 + 1.30\,(16 - 13) = 14.70 \qquad E(y \mid x = 13) = 10.80 \tag{20}$$

Back to the slide

Worked Example: Two Means in Our Data

The sample mean has two names, and so does the population mean:

$$\hat{\mu}_y = \bar{y} \qquad \mu_y = E(y) \tag{21}$$

Our 120 give the sample pair, and the population gives the other:

$$\hat{\mu}_y = \bar{y} = 12.25 \qquad \mu_y = E(y) = 11.52 \tag{22}$$

Back to the slide

Worked Example: Our Sample Mean Against the Population Mean

The gap is the sample mean minus the population mean:

$$\bar{y} - \mu_y \tag{23}$$

Our 120 average 12.25 euro an hour, and the population mean is 11.52:

$$\bar{y} - \mu_y = 12.25 - 11.52 = 0.73 \tag{24}$$

Back to the slide

Derivation: the Algebra of Expectations

Write each expectation as its weighted sum, split it, take the constants outside. The step doing the work is splitting the sum, which only addition allows.

Three Rules, the Variance Form Follows

Three rules follow: a constant, a linear function, and a sum. The variance form $E(y^2) - (E(y))^2$ is a consequence of them, not a fourth rule.

Derivation: the Conditional Mean Minimises Expected Squared Error

Add and subtract $E(y \mid x)$ inside the square, then expand. The cross term multiplies a deviation from $E(y \mid x)$ by something fixed once x is known, so it has expectation zero.

The Cross Term Has Mean Zero

Assume $E(y^2)$ and $E(g(x)^2)$ are finite. What remains is the conditional variance plus a squared term, zero only at $g(x) = E(y \mid x)$.

Notes and Tools

The R behind today's lab and exhibits:

- **Binning in R:** `cut()` and `tapply()` do the whole binned-means lab.
- **The Lecture in One Line:** Compare `mean(y[x == 13])` with `mean(y)`.
- **This Lecture's Script:** The CEF explorer holds the code that draws its exhibits.

References

Angrist, J. D., & Pischke, J.-S. (2015). *Mastering ’metrics: The path from cause to effect*. Princeton University Press.

Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach* (7th ed.). Cengage Learning.