LECTURE 1.3

Estimators and Sampling Variation

ECON30130 Econometrics I

Sam Deegan
University College Dublin

Autumn 2027

Where We Are

1. Foundations

1.1 Data and Variation

1.2 Conditional Means

1.3 Sampling Variation

2. Simple Regression

2.1 Least Squares

2.2 Units and Fit

2.3 Bias and Precision

3. Multiple Regression

3.1 Estimation

3.2 Inference

3.3 Functional Form

3.4 Dummies

4. Broken Assumptions

4.1 Heteroskedasticity

4.2 Specification

Last Week's Lecture

Lecture 1.2 defined the population average, its algebra and the conditional mean. Its last section compared rules that estimate that average from one sample:

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.

Lecture Plan

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.

> **Wooldridge, Math Refresher C**
> Sections C-2c, C-3, C-5 and C-6 cover sampling variation, intervals and tests (Wooldridge, 2020).

SECTION 1

The Sampling Distribution

Our Sample Mean Sits 0.73 Above the Population Mean

Rule A, the sample mean ($\bar{y}$), gave 12.25 on our 120 people. The population we built has mean 11.52, written μ_y:

- **What Changes:** A different 120 people would give a different sample mean.
- **What Stays Fixed:** The population mean μ_y is one number that no sample changes.

Nothing yet says whether chance alone could produce a gap of 0.73.

Estimates Have a Sampling Distribution

Each sample of 120 gives its own mean, and those means form a distribution:

> **The Sampling Distribution**
> The distribution of $\bar{y}$ across all possible samples is its sampling distribution. Unbiasedness, consistency and efficiency, defined last week, are properties of that distribution.

That distribution's spread says how large a gap like 0.73 typically is.

The Sampling Distribution of the Mean, Simulated

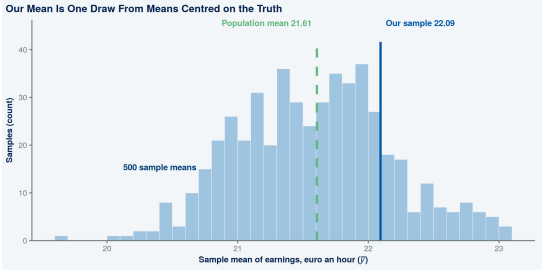

Figure 1: Means of 500 samples of 120, a repeated-sample ghost; truth 21.61, ours 22.09.

> **Our 22.09 Is One Draw**
> Our 22.09 is one draw, above 21.61 by chance rather than bias.

The Sampling Distribution Is Not Observable

In real work you draw once, get one estimate and never see that histogram.

> **Real Data Show One Draw**
> No real dataset shows the sampling distribution. A formula estimates the distribution's spread from one sample.

We wrote this module's population, so we can draw from it repeatedly.

Spread Across People and Spread Across Samples

Earnings spread widely across people, and sample means spread far less across samples:

5.91 euro

the standard deviation of our 120 earnings, across people

0.54 euro

the typical distance of a sample mean from μ_y, across samples

> **Why Means Vary Less Than People**
> High and low earners offset inside an average. Averaging divides the variance by n: 5.91 squared over 120 is 0.29, and the square root of 0.29 is 0.54.

The Standard Error of the Sample Mean

The population spread (σ_y) is unknown, so we estimate it from our sample. The estimated spread of $\bar{y}$ is called its standard error:

$$\text{se}(\bar{y}) = \frac{s_y}{\sqrt{n}} \tag{1}$$

- s_y: the standard deviation of the 120 earnings we drew

> **Variances Add, Standard Deviations Do Not**
> Variances add across independent draws, and standard deviations do not. Four times the data halves the standard error.

How the Standard Error Falls with Sample Size

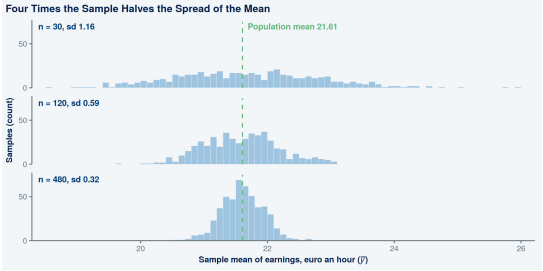

Figure 2: Means of 500 samples at n = 30, 120 and 480, repeated-sample ghosts on shared axes.

> **Same Centre, Narrower as the Sample Grows**
> All three histograms centre on 21.61, and only the width shrinks as *n* rises. The common centre is unbiasedness, and the standard error measures the width.

What Doubling the Sample Size Changes

Suppose we drew 240 people instead of 120, from the same population.

> **Question**
> Does the histogram of sample means shift, narrow, do both, or do neither?

Doubling the Sample Narrows the Histogram by Root Two

Doubling n keeps the centre at μ_y and divides the spread of $\bar{y}$ by $\sqrt{2}$:

$$\frac{\sigma_y}{\sqrt{2n}} = \frac{1}{\sqrt{2}} \times \frac{\sigma_y}{\sqrt{n}} \tag{2}$$

Root Two, Not Two
Doubling n cuts the spread by about 29 per cent, not by half. Halving the spread takes four times the data.

SECTION 2

The Central Limit Theorem and Confidence Intervals

The Central Limit Theorem and Its Conditions

Sample means are close to normal in large samples, whatever the population's shape:

- **The Standardised Mean:** We subtract μ_y from $\bar{y}$ and divide by $\sigma_y/\sqrt{n}$.
- **The Large-Sample Shape:** A standardised mean is approximately standard normal in large samples.

> **Large Samples, Finite Variance**
> The theorem needs random sampling and a finite population variance. The approximation improves as n grows.

Averages from a Skewed Population

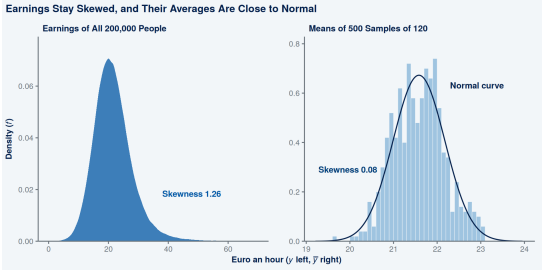

Figure 3: Left: earnings in the population. Right: 500 means of 120, a repeated-sample ghost.

> **The Population Stays Skewed**
> The population keeps its long right tail, while the means' tails are thin. A mean lands far from μ_Y only when many draws deviate together.

What the Central Limit Theorem Does Not Say

The theorem is about averages, and says nothing about the people behind them:

- **Not About One Person:** Earnings stay skewed, and poor workers outnumber rich ones.
- **Not Exact:** The normal shape improves with n and never holds exactly at 120.

> **Earnings Were Never Assumed Normal**
> We never assumed earnings were normal, and they are not. An approximate interval for μ_y needs only the average to be close to normal.

Building a 95 per Cent Interval

A single estimate says nothing about how far it could sit from the truth:

$$\bar{y} \pm \underbrace{1.96 \times \text{se}(\bar{y})}_{\text{the margin of error}} \tag{3}$$

Where 1.96 Comes From

A standard normal distribution puts 2.5 per cent in each tail beyond ± 1.96. The interval spans the 95 per cent between those two points.

Worked example: our 120 people

Coverage Across Repeated Intervals

Figure 4: 100 intervals from 100 samples of 120, a repeated-sample ghost; the line is 21.61.

> **About Five in a Hundred Should Miss, and Six Do**
> Each interval has its own endpoints, and 21.61 stays fixed. The intervals that miss come from the means farthest from 21.61.

Why the 95 per Cent Describes the Procedure

The 95 per cent counts how often this rule's intervals contain μ_y:

- **The Endpoints:** Both endpoints are built from $\bar{y}$, so both change with every sample.
- **The Population Mean:** μ_y is the same number in every sample.

> **Probability Before the Draw, Not After**
> Before the sample is drawn, the random interval covers μ_y with probability 0.95. Once the endpoints are numbers, μ_y either sits inside them or does not.

No sample says whether our own 11.19 to 13.31 covered 11.52.

Wording a Confidence Interval

Only one of these sentences is defensible, and the wrong one sounds better:

This Wording Is Wrong
"There is a 95 per cent chance the true average lies between 11.19 and 13.31."

This Wording Is Correct
"Our estimate is 12.25 euro, with a 95 per cent confidence interval from 11.19 to 13.31."

The first sentence makes μ_y vary; only the endpoints vary (Hoekstra et al., 2014).

SECTION 3

Hypothesis Tests

The Null Hypothesis: a Claim We Try to Contradict

Testing starts from someone else's claim, not from the answer we want. That claim is the null hypothesis (H_0):

- **The Claim:** Somebody asserts a value μ_0 for mean earnings, here ten euro: $H_0 : \mu_y = \mu_0$.
- **The Test:** A t-test measures how far our 12.25 sits from ten euro.
- **The Verdict:** We reject H_0 when a sample mean this far would rarely arise under it.

> **We Reject or Fail to Reject**
> A test rejects the null hypothesis or fails to reject it, never accepts it. Our sample fails to reject every value from 11.19 to 13.31 at five per cent.

Testing the Mean Against a Claimed Value

Our 12.25 sits 2.25 euro above the claimed ten euro. The t-statistic reports that distance in standard errors:

$$t = \frac{\bar{y} - \mu_0}{\text{se}(\bar{y})} \tag{4}$$

- μ_0: the value H_0 asserts for μ_y, here 10.00 euro
- t: how many standard errors $\bar{y}$ sits from μ_0

> **Roman for the Sample, Greek for the Claim**
> $\bar{y}$ is Roman and comes from our 120. μ_0 is Greek because H_0 claims it of the population.

Worked example: our 120 people

Type I and Type II Errors

A test can reach two kinds of wrong conclusion, one for each state of H_0:

	H_0 **is true**	H_0 **is false**
We reject H_0	type I error	correct
We fail to reject H_0	correct	type II error

> **We Choose the Type I Rate**
> Five per cent is a convention we choose, not a fact. A five per cent test rejects five true nulls in every hundred.

The Two Error Rates Trade Off

Lowering the type I rate makes every rejection harder, so more false nulls go unrejected.

> **Failing to Reject Is Weak Evidence**
> A large standard error makes rejection hard, whether the null is true or false. So with a large standard error, a false H_0 often goes unrejected.

Only a larger sample lowers both rates at once.

What a p-Value Measures

A p-value describes what the claim would produce, not whether it is true:

- **What a p-Value Is:** A p-value is the smallest type I rate at which we would reject.
- **A Two-Sided Test:** We count samples landing far from ten euro on either side.
- **Under the Claim:** Fewer than one sample in ten thousand gives $|t|$ above 4.2.

> **Two Common Misreadings**
> A p-value is not the probability the claim is false. A large p-value is no evidence that the claim is true (Wasserstein & Lazar, 2016).

Statistical Significance Is Not the Size of the Gap

A large statistic comes from a large gap or a small standard error:

- **Statistical Significance:** The t-test compares the gap with its own standard error.
- **Practical Significance:** The gap itself is 2.25 euro an hour.
- **Why Both Are Reported:** A large sample makes a tiny gap statistically significant.

> **Significance Is Not Magnitude**
> Whether 2.25 euro an hour matters is a judgement, not a test result.

What Would Have to Be True?

Our 12.25 now carries an interval, 11.19 to 13.31, and a p-value below 0.0001.

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?

A Self-Selected Sample Covers Less Than 95 per Cent

The 95 per cent belongs to the procedure, and the procedure includes random sampling.

> **If the 120 Chose Themselves**
> Suppose the people who answered are those with something to report. Then $\bar{y}$, the interval and t are off in a direction the standard error cannot see.

The rule then covers less than 95 per cent, and nothing in 11.19 to 13.31 shows the shortfall.

Simulating Coverage and Width in the Lab

The lab repeats this lecture's simulation in three steps:

- **Draw:** We draw 120 people, take the mean, and build an interval.
- **Repeat:** We repeat that 500 times and record how often 11.52 falls inside.
- **Change:** We raise *n* and compare the coverage with the width.

> **Coverage Holds, Width Shrinks**
> Raising *n* narrows the intervals and leaves coverage near 95 per cent.

Three Claims About an Estimate

Each section answered one question about our 12.25:

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.

Before Lecture 2.1

1 Play:
The sampling app, at $n = 30$ and then $n = 480$.

2 Apply:
Problem set 1, covering 1.2 and 1.3.

3 Review:
Wooldridge (2020), Math Refresher C-2c, C-3, C-5 and C-6.

> **Problem Set 1 This Week**
> [Out this week, due week n. The last question is the coverage simulation, written as prose not code.]

Next Lecture

Lecture 2.1 fits a line through two variables, and its slope is an estimator like $\bar{y}$:

1. **The Line as a Conditional Expectation.**
   A line is a claim about the average earnings of people with a given schooling.

2. **Least Squares.**
   Minimising the squared gaps picks one line and gives its slope formula.

3. **Fitted Values, Residuals and Sums of Squares.**
   The chosen line splits every person's earnings into a fitted value and a residual.

4. **Goodness of Fit.**
   R-squared measures the share of earnings variation the line accounts for.

Thank you

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

BACKUP

Appendix

Notation Used in This Lecture

Symbol	What it is	First met
n	how many people are in the sample	Lecture 1.1
$\bar{y}$	rule A, the sample mean of our 120 earnings	Lecture 1.2
μ_y	the population mean of earnings, 11.52	Lecture 1.2
$\hat{\mu}_y$	another name for $\bar{y}$	Lecture 1.2
σ_y	the population spread of earnings	The Standard Error of the Sample Mean
s_y	the standard deviation of the 120 earnings	The Standard Error of the Sample Mean
$\text{se}(\bar{y})$	the standard error of $\bar{y}$	The Standard Error of the Sample Mean
1.96	the standard normal's 2.5 per cent tail point	Building a 95 Per Cent Interval
H_0	the null hypothesis	The Null Hypothesis: A Claim We Try to Contradict
μ_0	the value H_0 asserts for μ_y, here 10.00	The Null Hypothesis: A Claim We Try to Contradict
t	how many standard errors $\bar{y}$ sits from μ_0	Testing the Mean Against a Claimed Value

Worked Example: Our Interval Runs from 11.19 to 13.31 Euro

The interval is the sample mean plus and minus its margin of error:

$$\bar{y} \pm \underbrace{1.96 \times \text{se}(\bar{y})}_{\text{the margin of error}} \tag{5}$$

Our mean of 12.25 and its standard error of 0.54 give both endpoints:

$$12.25 - 1.96 \times 0.54 = 11.19 \qquad 12.25 + 1.96 \times 0.54 = 13.31 \tag{6}$$

> **1.96 Is a Large-Sample Value**
> Earnings are skewed, so 1.96 rests on the central limit theorem. The 95 per cent is approximate.

Back to the slide

Worked Example: Our Mean Sits 4.2 Standard Errors Above Ten

The t-statistic reports the gap from the claimed value in standard errors:

$$t = \frac{\bar{y} - \mu_0}{\text{se}(\bar{y})} \tag{7}$$

Dividing the 2.25 gap by the 0.54 standard error gives our t-statistic:

$$t = \frac{12.25 - 10.00}{0.54} = 4.2 \tag{8}$$

> **A T Beyond About Two Counts Against the Claim**
> A $|t|$ above about two would be unusual if H_0 were true. Our t-statistic is 4.2.

Back to the slide

Derivation: Unbiasedness and Consistency of the Sample Mean

The derivation has two short parts:

- **Part One:** Push the expectation through the sum; every draw has mean μ_y.
- **Part Two:** Do the same to the variance, then let n grow in its denominator.

> **Linearity, Then Independence**
> Linearity of expectation carries part one. Independence of the draws carries part two, and random sampling guarantees it.

[Full derivation. Set as homework; worked solution published with the set.]

Derivation: the Standard Error of the Mean

Compute the variance of the sample mean, then take its square root. Replace the population spread with the sample one.

> **Root N Is Not Assumed**
> The root n appears only at the last step, from taking the square root.

[Full derivation. Set as homework; worked solution published with the set.]

Notes and Tools

- The app, and this lecture's simulation script.
- `mean()`, `sd()` and `replicate()` in R, which are the whole of the lab.
- Simulation as the route in (Chance et al., 2004; delMas et al., 1999; Tintle et al., 2012).
- Two misreadings, a probability on μ_y and the two spreads confused (Greenland et al., 2016; Kalinowski et al., 2008; Sotos et al., 2007).
- Retrieval practice, and why it beats restudy (Crouch & Mazur, 2001; Roediger & Karpicke, 2006).
- What a first course should spend its time on (Angrist & Pischke, 2017).

References

Angrist, J. D., & Pischke, J.-S. (2017). Undergraduate econometrics instruction: Through our classes, darkly. *Journal of Economic Perspectives*, *31*(2), 125–144. https://doi.org/10.1257/jep.31.2.125

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295–323). Springer. https://doi.org/10.1007/1-4020-2278-6_13

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, *69*(9), 970–977. https://doi.org/10.1119/1.1374249

delMas, R. C., Garfield, J., & Chance, B. L. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, *7*(3). https://doi.org/10.1080/10691898.1999.12131279

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

References

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, *21*(5), 1157–1164. https://doi.org/10.3758/s13423-013-0572-3

Kalinowski, P., Fidler, F., & Cumming, G. (2008). Overcoming the inverse probability fallacy: A comparison of two teaching interventions. *Methodology*, *4*(4), 152–158. https://doi.org/10.1027/1614-2241.4.4.152

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Sotos, A. E. C., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, *2*(2), 98–113. https://doi.org/10.1016/j.edurev.2007.04.001

Tintle, N., Topliff, K., VanderStoep, J., Holmes, V.-L., & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum. *Statistics Education Research Journal*, *11*(1), 21–40. https://doi.org/10.52041/serj.v11i1.343

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

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

References