CREST, Institut Polytechnique de Paris

Spring 2026

About This Course

Logistics

- 13 lectures, 1.5 hours each
- Office hours: TBA
- Course website: TBA

Prerequisites

- Probability and statistics (L3 level)
- Linear algebra (matrix operations)
- Calculus (optimization, integration)

What You Will Learn

By the end of this course, you will be able to:

1. **Formulate** econometric models for causal questions
2. **Estimate** parameters using OLS, IV, and panel methods
3. **Evaluate** assumptions and diagnose violations
4. **Interpret** results with statistical and economic rigor
5. **Implement** analyses in R with real data

Course Roadmap

Foundations

1. **Introduction & Probability Review**
2. Simple Regression
3. Multiple Regression: Estimation
4. Multiple Regression: Inference
5. Asymptotics for OLS

Diagnostics

6. Heteroskedasticity
7. Specification & Functional Form

Advanced Methods

8. Instrumental Variables
9. Simultaneous Equations
10. Panel Data
11. Binary Response Models

Extensions

12. Time Series Basics
13. Advanced Topics & Review

Textbooks and Resources

Primary textbook

- Wooldridge (2019) — comprehensive, example-driven

Complementary reading

- Angrist and Pischke (2009) — causal inference perspective
- Stock and Watson (2020) — accessible alternative
- Hansen (2022) — advanced/graduate reference

Software

- R (with RStudio) — introduced from Lecture 2 onward

Assessment

Component	Weight
Problem sets (4–5)	30%
Midterm exam	30%
Final exam	40%

Problem sets combine analytical derivations with R-based empirical exercises. Collaboration is encouraged; write-ups must be individual.

Why Econometrics?

The fundamental problem: economic data are **observational**, not experimental.

Correlation

- People with more education earn more
- Countries with more police have more crime
- Hospital patients are sicker than average

Causation?

- Does education *cause* higher earnings?
- Do police *cause* crime?
- Do hospitals *make* people sick?

> Econometrics provides the tools to move from correlation to causal statements — under clearly stated assumptions.

Motivating Example: Returns to Education

The **Mincer wage equation** (Mincer, 1974):

$$\ln(\text{wage}_i) = \beta_0 + \beta_1 \,\text{educ}_i + \beta_2 \,\text{exper}_i + \beta_3 \,\text{exper}_i^2 + u_i$$

Key questions:

- What does β_1 measure? Under what conditions is it *causal*?
- What problems arise if ability is omitted?
- How do we estimate, test, and interpret β_1?

We will return to this example throughout the course. By Lecture 8 (IV), you will have multiple strategies to estimate the causal return to education.

Part II
Review of Probability

Wooldridge, Appendix B

Random Variables

Random Variable

A **random variable** X is a function from a sample space Ω to $\mathbb{R}$:

$$X : \Omega \to \mathbb{R}$$

Discrete X

- Takes countably many values
- Characterized by its **probability mass function** (PMF):

$$f(x) = P(X = x)$$

- Example: number of children

Continuous X

- Takes uncountably many values
- Characterized by its **probability density function** (PDF):

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x)\, dx = 1$$

- Example: hourly wage

Cumulative Distribution Function

Cumulative Distribution Function (CDF)

The **cumulative distribution function** of X is:

$$F(x) = P(X \leq x)$$

Properties:

1. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
2. F is non-decreasing
3. F is right-continuous
4. $P(a < X \leq b) = F(b) - F(a)$

For continuous X: $f(x) = F'(x)$ and $F(x) = \int_{-\infty}^{x} f(t)\, dt$

Expected Value

Expected Value

Discrete: $\mathbb{E}[X] = \sum_x x \cdot f(x)$
Continuous: $\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx$

Linearity of expectation:

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

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] \quad \text{(always, even if dependent)}$$

For a function $g(X)$: $\mathbb{E}[g(X)] = \int g(x)\, f(x)\, dx$

Variance and Standard Deviation

Variance

$$\operatorname{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2$$

The **standard deviation** is $\operatorname{sd}(X) = \sqrt{\operatorname{Var}(X)}$.

Properties:

- $\operatorname{Var}(X) \geq 0$, with equality iff X is constant a.s.
- $\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)$
- $\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X, Y)$

Covariance and Correlation

Covariance and Correlation

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\mathsf{sd}(X)\,\mathsf{sd}(Y)}$$

Key facts:

- $-1 \leq \mathrm{Corr}(X, Y) \leq 1$
- $\mathrm{Corr}(X, Y) = 0$ means **uncorrelated** (not necessarily independent)
- $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
- $\mathrm{Cov}(aX + b,\, cY + d) = ac\,\mathrm{Cov}(X, Y)$

Key Properties of $\mathbb{E}$ and Var

Expected value:

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

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i]$$

Variance:

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

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

If X_i **uncorrelated**: $\mathrm{Var}\left(\sum_i X_i\right) = \sum_i \mathrm{Var}(X_i)$; otherwise add $2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$.

Joint, Marginal, and Conditional Distributions

Joint PDF/PMF: $f_{X,Y}(x,y)$ describes the simultaneous behavior of X and Y.

Marginal distribution:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

Conditional distribution:

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad f_X(x) > 0$$

> The conditional distribution $f_{Y|X}$ is the foundation of regression analysis: it describes how Y behaves *given* a particular value of X.

Conditional Expectation

Conditional Expectation

$$\mathbb{E}[Y \mid X = x] = \int y \, f_{Y|X}(y \mid x) \, dy$$

The **conditional expectation function** (CEF) is the map $x \mapsto \mathbb{E}[Y \mid X = x]$.

Why it matters for econometrics:

- The regression function $\mathbb{E}[Y \mid X]$ is the **best predictor** of Y given X
- It minimizes the mean squared prediction error (assuming $\mathbb{E}[Y^2] < \infty$):

$$\mathbb{E}[Y \mid X] = \arg\min_{g(X)} \mathbb{E}\left[(Y - g(X))^2\right]$$

- OLS approximates the CEF with a linear function

Law of Iterated Expectations (LIE)

Law of Iterated Expectations:

$$\mathbb{E}[Y] = \mathbb{E}\left[\mathbb{E}[Y \mid X]\right]$$

The unconditional mean of Y is the average of conditional means, weighted by f_X.

Why this matters:

- Allows us to decompose expectations by conditioning
- Key tool in proving properties of OLS estimators
- Used repeatedly in deriving unbiasedness, omitted variable bias, IV results

Example: $\mathbb{E}[\text{wage}] = \mathbb{E}\left[\mathbb{E}[\text{wage} \mid \text{educ}]\right]$
The overall average wage is the average of education-specific average wages.

Conditional Variance

Conditional Variance

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

Law of Total Variance:

$$\mathrm{Var}(Y) = \underbrace{\mathbb{E}\left[\mathrm{Var}(Y \mid X)\right]}_{\text{within-group}} + \underbrace{\mathrm{Var}\left(\mathbb{E}[Y \mid X]\right)}_{\text{between-group}}$$

Total variance decomposes into average *within-group* variance plus variance of *group means*.

Independence

Independence

X and Y are **independent** ($X \perp\!\!\!\perp Y$) if and only if:

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) \quad \text{for all } x, y$$

Implications of independence:

- $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y] \implies \mathrm{Cov}(X, Y) = 0$
- $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\,\mathbb{E}[h(Y)]$
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Warning: Uncorrelated $\not\Rightarrow$ independent (unless jointly Normal). Independence $\Rightarrow$ uncorrelated always holds.

Common Distributions in Econometrics

Distribution	Notation	Mean	Variance
Normal	$N(\mu, \sigma^2)$	μ	σ^2
Standard Normal	$N(0, 1)$	0	1
Chi-squared	χ^2_k	k	$2k$
Student's t	t_k	$0 \quad (k > 1)$	$\frac{k}{k-2} \quad (k > 2)$
Fisher's F	F_{k_1,k_2}	$\frac{k_2}{k_2-2} \quad (k_2 > 2)$	(complex)

These distributions arise naturally in testing:

- t-distribution $\rightarrow$ individual coefficient tests
- F-distribution $\rightarrow$ joint hypothesis tests
- χ^2-distribution $\rightarrow$ specification tests, MLE

The Normal Distribution

If $X \sim N(\mu, \sigma^2)$, its PDF is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Key properties:

- **Standardization:** $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
- **Linear combinations:** if $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are independent:

$$aX + bY \sim N\big(a\mu_X + b\mu_Y,\ a^2\sigma_X^2 + b^2\sigma_Y^2\big)$$

- **Symmetry:** $\Phi(-z) = 1 - \Phi(z)$, where Φ is the standard Normal CDF

> Under classical assumptions, OLS estimators are *exactly* Normal. Under weaker assumptions, they are *asymptotically* Normal (via CLT).

Chi-Squared, t, and F Distributions

Construction from Normal random variables:

1. If $Z_1, \ldots, Z_k \overset{\text{iid}}{\sim} N(0,1)$, then $\sum_{i=1}^{k} Z_i^2 \sim \chi_k^2$
2. If $Z \sim N(0,1)$ and $V \sim \chi_k^2$ are independent, then $\dfrac{Z}{\sqrt{V/k}} \sim t_k$
3. If $V_1 \sim \chi_{k_1}^2$ and $V_2 \sim \chi_{k_2}^2$ are independent, then $\dfrac{V_1/k_1}{V_2/k_2} \sim F_{k_1,k_2}$

Key relationships:

- $t_k^2 = F_{1,k}$ (squaring a t gives an F)
- As $k \to \infty$: $t_k \to N(0,1)$ and $\chi_k^2/k \to 1$

Moment Generating Functions

Moment Generating Function (MGF)

The **moment generating function** of X is:

$$M_X(t) = \mathbb{E}[e^{tX}]$$

defined for t in a neighborhood of zero.

Why useful:

- Moments: $\mathbb{E}[X^k] = M_X^{(k)}(0)$
- **Uniqueness:** if $M_X(t) = M_Y(t)$ in a neighborhood of 0, then $X \overset{d}{=} Y$
- **Sums of independents:** $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$

Example: $X \sim N(\mu, \sigma^2) \Rightarrow M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$

Probability Review: Key Results

1. $\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$ (always)
2. $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$
3. $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$
4. $X \perp\!\!\!\perp Y \;\Rightarrow\; \mathrm{Cov}(X, Y) = 0$ (converse is false)
5. $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]]$ (LIE)
6. $\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y \mid X)] + \mathrm{Var}(\mathbb{E}[Y \mid X])$ (total variance)
7. $\mathbb{E}[Y \mid X] = \arg\min_{g(X)} \mathbb{E}[(Y - g(X))^2]$ (best predictor)

These results will be used repeatedly from Lecture 2 onward.

Part III
Review of Statistical Inference

Wooldridge, Appendix C

Random Sampling

Random Sample

A **random sample** of size n is a collection $\{X_1, X_2, \ldots, X_n\}$ of **independent and identically distributed** (i.i.d.) random variables, each with the same distribution as X.

The i.i.d. assumption means:

- **Identical:** each X_i has the same distribution F
- **Independent:** knowing X_i tells you nothing about X_j $(i \neq j)$

When does i.i.d. hold?

- Cross-sectional surveys with random sampling: typically yes
- Time series data: typically no (observations are dependent)
- Panel data: partially (independent across i, dependent across t)

Estimators as Random Variables

Estimator

An **estimator** $\hat{\theta}_n = g(X_1, \ldots, X_n)$ is a function of the sample. Before the data are observed, $\hat{\theta}_n$ is a **random variable**.

Distinguish:

- **Parameter** θ: fixed, unknown quantity we want to learn
- **Estimator** $\hat{\theta}_n$: random variable (depends on the sample)
- **Estimate** $\hat{\theta}$: realized value from a specific sample

Because $\hat{\theta}_n$ is a random variable, it has a distribution — the **sampling distribution**. All of inference (bias, variance, testing) is about this distribution.

Unbiasedness

Unbiasedness

An estimator $\hat{\theta}_n$ is **unbiased** for θ if:

$$\mathbb{E}[\hat{\theta}_n] = \theta$$

The **bias** is $\text{Bias}(\hat{\theta}_n) = \mathbb{E}[\hat{\theta}_n] - \theta$.

Example: Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} (\mu, \sigma^2)$.

- $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ is unbiased for μ: $\quad \mathbb{E}[\bar{X}_n] = \mu$
- $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ is unbiased for σ^2
- $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ is **biased**: $\mathbb{E}[\tilde{S}^2] = \frac{n-1}{n}\sigma^2$

Efficiency

Efficiency

Among all unbiased estimators, $\hat{\theta}^*$ is **efficient** if it has the **smallest variance**:

$$\mathrm{Var}(\hat{\theta}^*) \leq \mathrm{Var}(\hat{\theta}) \quad \text{for all unbiased } \hat{\theta}$$

Mean squared error combines bias and variance:

$$\text{MSE}(\hat{\theta}_n) = \mathrm{Var}(\hat{\theta}_n) + \left[\text{Bias}(\hat{\theta}_n)\right]^2$$

Bias–variance tradeoff: a slightly biased estimator with much lower variance can have smaller MSE.

Preview: the Gauss–Markov theorem (Lecture 3) establishes OLS as BLUE.

Consistency

Consistency

$\hat{\theta}_n$ is **consistent** for θ if $\hat{\theta}_n \xrightarrow{p} \theta$ as $n \to \infty$, i.e., $\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0$ for all $\varepsilon > 0$.

Sufficient condition: $\text{Bias}(\hat{\theta}_n) \to 0$ and $\mathrm{Var}(\hat{\theta}_n) \to 0$ as $n \to \infty$.

Consistency vs. unbiasedness:

- Unbiasedness is a *finite-sample* property (holds for all n)
- Consistency is an *asymptotic* property (requires $n \to \infty$)
- Unbiased $\not\Rightarrow$ consistent; consistent $\not\Rightarrow$ unbiased

Law of Large Numbers

Weak Law of Large Numbers (WLLN): If $X_1, \ldots, X_n$ are i.i.d. with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) < \infty$, then:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$$

Intuition: as the sample grows, the sample mean converges to the population mean.

Why it matters for econometrics:

- Justifies using sample averages to estimate population quantities
- Underpins consistency of OLS: $\hat{\beta} \xrightarrow{p} \beta$
- Foundation for the method of moments

Central Limit Theorem

Central Limit Theorem (CLT): If $X_1, \ldots, X_n$ are i.i.d. with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 \in (0, \infty)$, then:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$$

Equivalently: $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$

The power of the CLT:

- Works *regardless* of the distribution of X_i
- Convergence is often fast (usable for $n \geq 30$ in practice)
- Justifies Normal-based inference even for non-Normal data

CLT: Visual Intuition

What happens as n grows? Consider sampling from a skewed distribution (e.g., exponential):

Sample size	**Distribution of $\bar{X}_n$**
$n = 1$	Same as original (skewed)
$n = 5$	Less skewed, more concentrated
$n = 30$	Approximately Normal
$n = 100$	Very close to Normal, tightly concentrated

$\bar{X}_n$ is approximately Normal for large n, no matter how non-Normal the population. This is why we can use t-tests and confidence intervals without assuming Normality.

Properties of the Sample Mean and Variance

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} (\mu, \sigma^2)$.

Sample mean: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$

$$\mathbb{E}[\bar{X}_n] = \mu, \qquad \mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$

- $\mathbb{E}[S^2] = \sigma^2$ (unbiased)
- The $n-1$ denominator corrects for using $\bar{X}_n$ instead of μ

Standard error of the mean:

$$\mathsf{SE}(\bar{X}_n) = \frac{S}{\sqrt{n}}$$

Confidence Intervals

Confidence Interval

A $100(1-\alpha)\%$ **CI** for θ is a random interval $[\hat{\theta}_L, \hat{\theta}_U]$ s.t. $P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1-\alpha$.

For the mean (large n or Normal population):

$$\bar{X}_n \pm t_{n-1,\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

Interpretation:

- **Correct:** if we repeated sampling many times, $100(1-\alpha)\%$ of intervals would contain θ
- **Wrong:** “there is a 95% probability that θ is in this interval” — θ is fixed; the interval is random

Hypothesis Testing

Setup:

- **Null hypothesis** H_0: a specific claim about θ (e.g., $H_0 : \theta = \theta_0$)
- **Alternative** H_1: the complement (e.g., $H_1 : \theta \neq \theta_0$)
- **Test statistic:** a function of the data that measures evidence against H_0

Decision rule: reject H_0 if the test statistic falls in the **rejection region**.

General test statistic:

$$T = \frac{\hat{\theta} - \theta_0}{\mathsf{SE}(\hat{\theta})}$$

Under H_0 and regularity conditions: $T \sim t_{n-k}$ (or $\approx N(0, 1)$ for large n).

Type I and Type II Errors

	H_0 **true**	H_0 **false**
Reject H_0	Type I error (α)	Correct (Power)
Fail to reject H_0	Correct ($1-\alpha$)	Type II error (β)

Key concepts:

- **Size** (α): probability of rejecting H_0 when it is true
- **Power** ($1-\beta$): probability of rejecting H_0 when it is false
- Standard choices: $\alpha = 0.10,\ 0.05,\ 0.01$

Tradeoff: decreasing α (fewer false positives) increases β (more false negatives), holding sample size fixed.

The t-Test

One-sample t-test: testing $H_0 : \mu = \mu_0$. Under Normality of X_i:

$$t = \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \quad \text{under } H_0$$

Two-sided ($H_1 : \mu \neq \mu_0$): reject if $|t| > t_{n-1,\,\alpha/2}$. **One-sided** ($H_1 : \mu > \mu_0$): reject if $t > t_{n-1,\,\alpha}$.

> Exact under Normality; approximately valid for large n by the CLT. This is the workhorse test in regression analysis (Lecture 4).

p-Values

p-Value

The p**-value** is the probability, under H_0, of observing a test statistic at least as extreme as the one computed from the data:

$$p = P(|T| \geq |t_{\text{obs}}| \mid H_0)$$

Decision rule: reject H_0 at level α if $p < \alpha$.

Common misinterpretations (**avoid these**):

- "The p-value is the probability that H_0 is true" **Wrong**
- "$p = 0.03$ means a 3% chance the result is due to chance" **Wrong**
- "$p > 0.05$ means H_0 is true" **Wrong**

Correct: the p-value measures the *compatibility* of the data with H_0.

From Statistics to Econometrics

Everything we reviewed today underpins regression analysis:

Statistical concept	Econometric application
Conditional expectation $\mathbb{E}[Y\|X]$	The regression function
Law of iterated expectations	Proving OLS unbiasedness
Variance decomposition	R^2 and model fit
Unbiasedness, efficiency	Gauss–Markov theorem
Consistency, LLN	Large-sample OLS properties
CLT	Asymptotic inference
t-test and F-test	Coefficient testing
Confidence intervals	Inference on β

Solid command of these foundations makes econometrics much more intuitive.

Next Time: The Simple Regression Model

In Lecture 2, we introduce the model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

What we will cover:

- Derivation of the OLS estimator via minimizing $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$
- Interpretation of β_0 and β_1
- Assumptions for unbiasedness: $\mathbb{E}[u \mid X] = 0$
- Properties: unbiasedness and the Gauss–Markov theorem
- First R exercise: estimating the returns to education

Reading: Wooldridge (2019, Chapters 1–2)

Key Takeaways

1. **Econometrics** = statistical methods for causal questions with observational data
2. The **conditional expectation** $\mathbb{E}[Y \mid X]$ is the regression function — the best predictor of Y given X
3. The **LIE** and **law of total variance** are workhorses for deriving estimator properties
4. **Estimators are random variables** — inference is about their sampling distributions
5. The **LLN** justifies consistency; the **CLT** justifies asymptotic inference
6. **Confidence intervals** and p**-values** have precise interpretations — beware of common mistakes

References I

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ.

Hansen, B. E. (2022). *Econometrics*. Princeton University Press, Princeton, NJ.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. Columbia University Press, New York.

Stock, J. H. and Watson, M. W. (2020). *Introduction to Econometrics*. Pearson, New York, 4th edition.

Wooldridge, J. M. (2019). *Introductory Econometrics: A Modern Approach*. Cengage Learning, Boston, MA, 7th edition.