Econ 103: Introduction to Econometrics

Lecture 4 — The Normal Distribution, Sampling & the CLT

Ryan Longmuir

UCLA

Summer Session C, 2026

Reading: Stock & Watson (4th ed.), §2.4–2.6; Hill, Griffiths & Lim (5th ed.), Probability Primer §P.7 & Appendix C.

Where we are

Last time (Lecture 3). Means, variances, and how to combine them — including $\mathbb{E}(\bar{X})$ and $\mathrm{Var}(\bar{X}) = \sigma^2/n$, plus standardizing $Z = (X - \mu)/\sigma$.

Today. We close out the probability toolkit and reach the doorstep of inference:

- the Normal distribution — the bell curve — and how to read probabilities off it;
- the sample mean $\bar{Y}$ as a random variable, with its own distribution;
- the Central Limit Theorem — why the Normal shows up everywhere.

The payoff

By the end of today we can say how close $\bar{Y}$ is likely to be to the truth μ. That single fact powers *every* confidence interval and hypothesis test in the rest of the course.

Today's plan

Part 1

The Normal distribution

The Normal distribution

Some distributions are special enough to earn a name. The most important is the **Normal**.

Definition

X is **normally distributed** with mean μ and variance σ^2, written $X \sim N(\mu, \sigma^2)$, if its density is the bell curve

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \qquad -\infty < x < \infty.$$

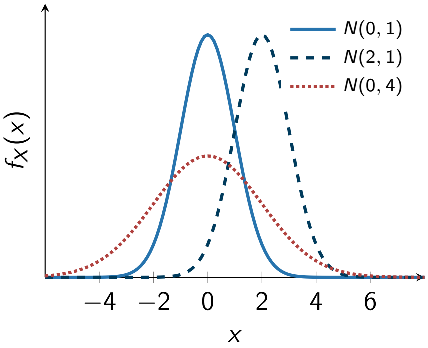
- It is symmetric and centered at μ (so mean = median, skewness = 0).
- μ sets the location; σ^2 sets the spread.

Same family, different μ and σ^2

Changing the parameters slides and stretches the same bell shape:

- μ moves the center left/right.
- larger σ^2 flattens and widens; smaller σ^2 makes it tall and tight.

The total area under each curve is still 1 — a wider curve must be shorter.

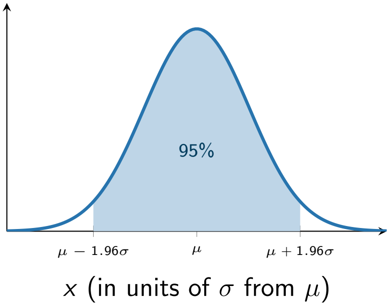


A fact worth memorizing: the 95% rule

For *any* Normal, about 95% of the probability lies within 1.96 standard deviations of the mean:

$$\mathbb{P}(\mu - 1.96\,\sigma \le X \le \mu + 1.96\,\sigma) \approx 0.95.$$

(Round numbers: $\approx 68\%$ within $\pm 1\sigma$, $\approx 95\%$ within $\pm 2\sigma$, $\approx 99.7\%$ within $\pm 3\sigma$.)



Part 2

Standardizing & the standard Normal

The standard Normal Z

Rather than a table for every (μ, σ^2), we convert to one reference distribution.

Standardization

The **standard Normal** is $Z \sim N(0,1)$. If $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1).$$

- This is exactly the standardizing move from Lecture 3 (mean 0, variance 1) — now applied to a Normal.
- Its cdf gets its own symbol: $\Phi(z) = \mathbb{P}(Z \leq z)$, tabulated in the textbook's Statistical Table 1 (and built into R).
- By symmetry, $\mathbb{P}(Z > a) = \mathbb{P}(Z < -a)$.

Reading probabilities off the Normal

Standardize, then look up Φ. For $X \sim N(\mu, \sigma^2)$ and constants $a < b$:

Property (Three rules)

$$\mathbb{P}(X \le a) = \Phi\left(\frac{a-\mu}{\sigma}\right),$$
$$\mathbb{P}(X \ge a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right),$$
$$\mathbb{P}(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Everything reduces to standard-Normal cdf values $\Phi(\cdot)$ — which is why a single table (or one R command, `pnorm`) does all the work.

Worked example

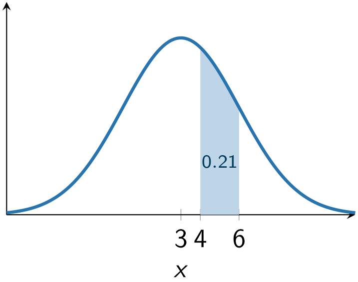
Let $X \sim N(3, 9)$, so $\mu = 3$ and $\sigma = 3$. Find $\mathbb{P}(4 \leq X \leq 6)$.

Standardize the endpoints:

$$\frac{4-3}{3} = 0.33, \qquad \frac{6-3}{3} = 1.$$

Difference of cdf values:

$$\begin{aligned}\mathbb{P}(4 \leq X \leq 6) &= \Phi(1) - \Phi(0.33) \\ &= 0.8413 - 0.6293 = 0.2120.\end{aligned}$$



Shaded area between 4 and 6 under $N(3, 9)$.

Key percentiles you will reuse all term

We will often need the value z_α with $\mathbb{P}(Z \leq z_\alpha) = \alpha$.

α	z_α
0.90	1.28
0.95	1.645
0.975	1.96
0.99	2.33
0.995	2.58

The three to memorize

1.645, 1.96, 2.58.

Because the Normal is symmetric, a two-sided 95% range uses ± 1.96 (leaving 2.5% in each tail). That is the source of the "1.96" in the 95% rule — and of the confidence intervals in Lecture 9.

Part 3

Linear combinations of Normals

Sums of Normals are Normal

Last lecture we found the *mean* and *variance* of a linear combination. For Normals we now get the shape for free.

Closure under linear combinations

If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are jointly normal, then for constants a_1, a_2,

$$a_1 X_1 + a_2 X_2 \sim N\left(a_1\mu_1 + a_2\mu_2,\ a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1 a_2\, \sigma_{12}\right),$$

where $\sigma_{12} = \mathrm{Cov}(X_1, X_2)$.

This is special: most distributions *change shape* when you add them. The Normal does not — any linear combination of jointly normal variables is again Normal. (We will lean on this for the sample mean in a moment.)

Three properties of the (bivariate) Normal

If X and Y are *jointly* normal:

1. each marginal is normal: $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$;
2. zero covariance $\Rightarrow$ independence. (Recall this is *false* in general — it is a special gift of the Normal.)
3. the conditional mean is linear in the conditioning variable:

$$\mathbb{E}(Y \mid X = x) = \alpha + \beta x, \qquad \beta = \frac{\sigma_{XY}}{\sigma_X^2}.$$

Look closely at property 3

$\mathbb{E}(Y \mid X)$ is a straight line, with slope $\mathrm{Cov}(X, Y)/\mathrm{Var}(X)$ — exactly the regression slope previewed in Lecture 3. This **linear regression function** is where the whole course is heading (Lecture 5).

Relatives of the Normal (a signpost)

Three distributions *built from* the Normal run our later inference. We meet them properly when we need them:

- **Chi-squared** χ^2_m: the sum of m independent squared standard Normals. (Appears in variance and joint tests.)
- **Student's** t with m d.f.: bell-shaped but fatter-tailed than the Normal; $\to N(0,1)$ as $m \to \infty$. Used when we estimate σ rather than know it — *Lecture 9*.
- F with (m, n) d.f.: a ratio of scaled chi-squareds, for testing *several* restrictions at once — *Lecture 17*.

For now: just know the t is a slightly wider Normal that we switch to once σ is unknown.

Part 4

Random sampling & the sample mean

Random sampling and i.i.d. data

All our methods rest on *how the data were drawn*.

Simple random sampling

Draw n observations $Y_1, \ldots, Y_n$ at random from a population with mean μ and variance σ^2. Then they are **i.i.d.**:

- identically distributed — each Y_i has the population's distribution (mean μ, variance σ^2);
- independent — knowing one tells you nothing about the others.

Before we look, each Y_i is a *random variable*; after sampling, it is a recorded number. Different draws would give different numbers — the source of **sampling variation**.

The sample mean is a random variable

Our estimator of the population mean μ is the **sample mean**

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Because the Y_i are random, $\bar{Y}$ is itself random. A different sample yields a different $\bar{Y}$. Its distribution — over all possible samples — is the **sampling distribution** of $\bar{Y}$.

Example (HGL hip-width data). Ten samples of size 50 gave sample means ranging from 16.75 to 17.41 — same population, different $\bar{y}$ each time.

The key shift in thinking

We stop asking "is *this* estimate right?" (unanswerable) and ask "how does the *procedure* $\bar{Y}$ behave across samples?" That we *can* answer — via its mean and variance.

Mean and variance of $\bar{Y}$

Apply last lecture's rules to $\bar{Y} = \frac{1}{n}\sum_i Y_i$ with i.i.d. draws.

Mean (linearity — no independence needed):

$$\mathbb{E}(\bar{Y}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i) = \frac{1}{n}(n\mu) = \mu.$$

So $\bar{Y}$ is **unbiased**: on average across samples it equals μ.

Variance (uses independence, so $\mathrm{Cov}(Y_i, Y_j) = 0$):

$$\mathrm{Var}(\bar{Y}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(Y_i) = \frac{\sigma^2}{n}.$$

Standard deviation $\sigma_{\bar{Y}} = \sigma/\sqrt{n}$ — the **standard error**.

Read these off

Center at the truth μ; spread $\sigma/\sqrt{n}$ shrinks as n grows. More data $\Rightarrow$ tighter around μ. *These hold for any population distribution.*

If the population is Normal, so is $\bar{Y}$

Since $\bar{Y}$ is a linear combination of the Y_i, closure under linear combinations gives an *exact* result:

$$Y_i \sim N(\mu, \sigma^2) \quad \Longrightarrow \quad \bar{Y} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right).$$

Example (Precision and sample size (HGL))

A normal population with $\sigma^2 = 10$*. With* $n = 40$, $\bar{Y} \sim N(\mu, 0.25)$, *so*

$$\mathbb{P}\big(|\bar{Y} - \mu| \leq 1\big) = \mathbb{P}(-2 \leq Z \leq 2) = 0.954.$$

Raising n *to* 80 *tightens this to* 0.995. *More data, more precision.*

Part 5

Law of large numbers & the CLT

The Law of Large Numbers

LLN

As the sample size grows, the sample mean converges to the population mean:

$$\bar{Y} \xrightarrow{p} \mu \qquad \text{as } n \to \infty.$$

The "law of averages": with many draws, high and low values balance out, and $\bar{Y}$ settles on μ. This is the formal reason large samples are trustworthy — it makes $\bar{Y}$ a **consistent** estimator.

But the LLN only says $\bar{Y}$ *gets close* to μ. To do inference we need the *shape* of its distribution around μ. That is the Central Limit Theorem.

The Central Limit Theorem

CLT

If $Y_1, \dots, Y_n$ are i.i.d. with mean μ and variance σ^2, then the standardized sample mean converges to a standard Normal:

$$\frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} N(0,1) \qquad \text{as } n \to \infty,$$

so for large n, $\quad \bar{Y} \overset{a}{\sim} N\big(\mu, \sigma^2/n\big)$.

The remarkable part: this holds whatever the population distribution — skewed, discrete, fat-tailed, anything (as long as σ^2 is finite). The bell curve emerges from the *averaging*, not from the data's own shape.

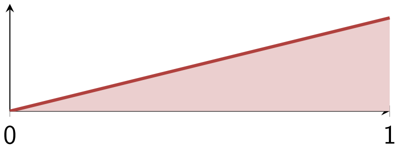
Rule of thumb: $n \geq 30$ is usually enough for the approximation to be good; real samples (hundreds–thousands) make it excellent.

The CLT in action

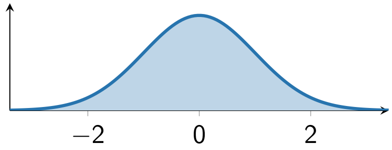
Take a decidedly *non*-Normal population — the triangular density $f(y) = 2y$ on $[0, 1]$ (HGL's example), which is skewed.

Draw samples, average them, standardize. Even for a sample as small as $n = 10$, the histogram of the standardized means is already bell-shaped and centered at 0 — and it only sharpens toward $N(0, 1)$ as n grows.

population: $f(y) = 2y$ (skewed)



std. mean for $n = 10 \to N(0, 1)$



Part 6

The bridge to inference

Why today is the gateway to inference

Put the pieces together. For large n, the CLT gives $\dfrac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \approx N(0,1)$, so using the 95% rule,

$$\mathbb{P}\left(-1.96 \le \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \le 1.96\right) \approx 0.95.$$

Rearrange one way → Lecture 9

Solving the inequality for μ brackets it in a confidence interval $\bar{Y} \pm 1.96\,\sigma/\sqrt{n}$.

Read it another way → Lecture 10

Comparing $\bar{Y}$ to a hypothesized μ gives a hypothesis test: is the standardized gap beyond ± 1.96?

And because regression estimators are (weighted) averages, the *same* CLT logic will make *their* sampling distributions Normal too.

A caution: when "Normal" fails

The Normal is powerful, but assuming it blindly is dangerous.

Example (The Swiss franc, 15 Jan 2015 (Stock & Watson))

The euro fell 17.5% *against the franc in one day — about* 156 *standard deviations. Under a Normal, that has probability* $\sim 10^{-5000}$*: effectively impossible. Yet it happened.*

The lesson: real financial returns have fat tails, so extreme moves are far more common than a Normal predicts. The CLT is about the *sample mean's* distribution, not a license to assume the *data themselves* are Normal.

Recap

The Normal

- $X \sim N(\mu, \sigma^2)$; standardize $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
- probabilities via $\Phi(\cdot)$; 95% within $\pm 1.96\sigma$
- key z: 1.645, 1.96, 2.58
- linear combos of Normals are Normal

Sampling & the CLT

- $\mathbb{E}(\bar{Y}) = \mu$, $\mathrm{Var}(\bar{Y}) = \sigma^2/n$
- normal population $\Rightarrow \bar{Y} \sim N(\mu, \sigma^2/n)$ exactly
- LLN: $\bar{Y} \to \mu$
- CLT: $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$ for *any* population

The bridge

$\mathbb{P}\left(-1.96 \leq \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \leq 1.96\right) \approx 0.95$ becomes confidence intervals (L9) and hypothesis tests (L10). **Next: Lecture 5 — the Simple Linear Regression Model.**

Questions?