The Normal Distribution, Sampling, and the CLT

Why the Bell Curve Shows Up Everywhere

Jake Anderson

Outline

Outline

A Question

Think about the distribution of **household income** in the United States.

A Question

Think about the distribution of **household income** in the United States.

- Most households earn between $40,000 and $100,000
- A long right tail stretches past $1,000,000
- The distribution is **clearly not symmetric**

A Question

Think about the distribution of **household income** in the United States.

- Most households earn between $40,000 and $100,000
- A long right tail stretches past $1,000,000
- The distribution is **clearly not symmetric**

Yet in econometrics we constantly assume errors are **normally distributed** (symmetric, bell-shaped).

A Question

Think about the distribution of **household income** in the United States.

- Most households earn between $40,000 and $100,000
- A long right tail stretches past $1,000,000
- The distribution is **clearly not symmetric**

Yet in econometrics we constantly assume errors are **normally distributed** (symmetric, bell-shaped).

Why is that justified? The answer is the Central Limit Theorem, which we will build toward today.

A Question

Think about the distribution of **household income** in the United States.

- Most households earn between $40,000 and $100,000
- A long right tail stretches past $1,000,000
- The distribution is **clearly not symmetric**

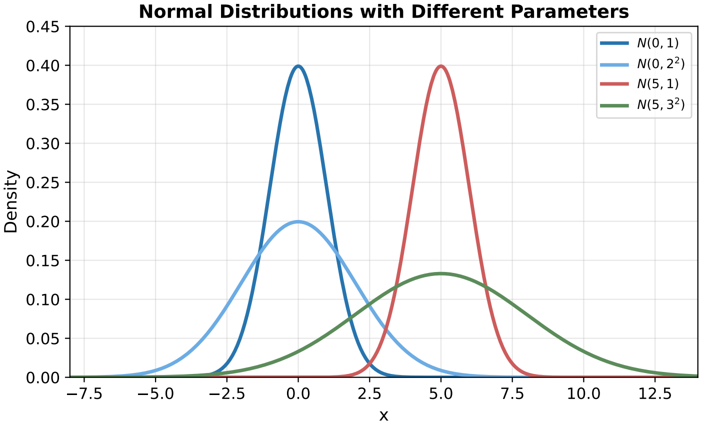
Yet in econometrics we constantly assume errors are **normally distributed** (symmetric, bell-shaped).

Why is that justified? The answer is the Central Limit Theorem, which we will build toward today.

But first: what *is* the normal distribution, and why is it so useful?

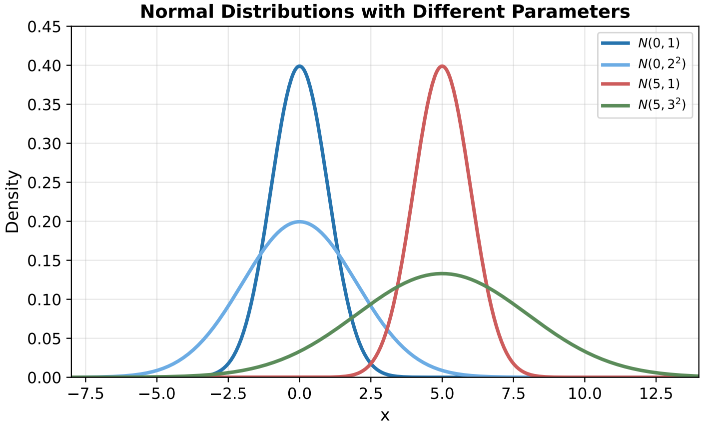
What the Normal Distribution Looks Like

The normal distribution is a **symmetric, bell-shaped** curve. It is completely determined by two parameters: μ (center) and σ^2 (spread).



What the Normal Distribution Looks Like

The normal distribution is a **symmetric, bell-shaped** curve. It is completely determined by two parameters: μ (center) and σ^2 (spread).



Changing μ shifts the curve left or right. Changing σ makes it wider or narrower. The shape is

The Normal PDF

The function that produces that bell shape is:

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

The Normal PDF

The function that produces that bell shape is:

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

We write $X \sim N(\mu, \sigma^2)$.

The Normal PDF

The function that produces that bell shape is:

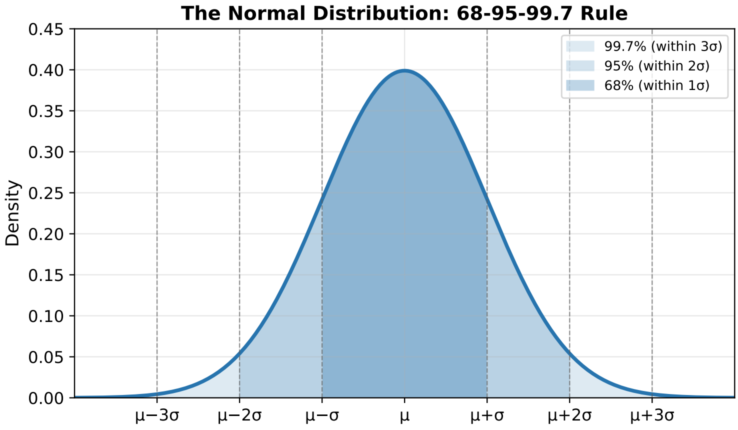
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < \infty$$

We write $X \sim N(\mu, \sigma^2)$.

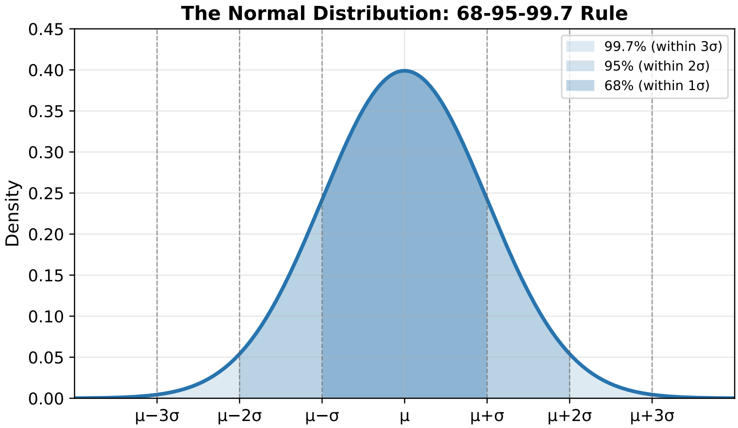
You will never need to evaluate this formula by hand. What you need to know:

- μ determines the **center** (mean = median = mode)
- σ^2 determines the **spread**
- The range is $(-\infty, +\infty)$, but almost all probability is within a few σ of μ

The 68-95-99.7 Rule



The 68-95-99.7 Rule



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

- About **68%** of values fall within 1 standard deviation: $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.68$
- About **95%** within 2: $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$

Outline

The Problem: Every Normal Looks Different

Heights of adult women in the U.S. follow approximately $N(64, 3^2)$ (inches).
SAT math scores follow approximately $N(530, 110^2)$.

The Problem: Every Normal Looks Different

Heights of adult women in the U.S. follow approximately $N(64, 3^2)$ (inches).
SAT math scores follow approximately $N(530, 110^2)$.

These have different units, different means, different variances. How do we compute probabilities without a separate table for each?

The Problem: Every Normal Looks Different

Heights of adult women in the U.S. follow approximately $N(64, 3^2)$ (inches).
SAT math scores follow approximately $N(530, 110^2)$.

These have different units, different means, different variances. How do we compute probabilities without a separate table for each?

The standard normal distribution solves this. If we transform any normal variable so it has mean 0 and variance 1, we only need one table (or one function in software).

Standardization: $Z = (X - \mu)/\sigma$

If $X \sim N(\mu, \sigma^2)$, define:

$$Z = \frac{X - \mu}{\sigma}$$

Standardization: $Z = (X - \mu)/\sigma$

If $X \sim N(\mu, \sigma^2)$, define:

$$Z = \frac{X - \mu}{\sigma}$$

Why does $Z \sim N(0, 1)$? Two steps:

Standardization: $Z = (X - \mu)/\sigma$

If $X \sim N(\mu, \sigma^2)$, define:

$$Z = \frac{X - \mu}{\sigma}$$

Why does $Z \sim N(0, 1)$? Two steps:

Step 1: Z is still normal, because affine transformations preserve normality.

Standardization: $Z = (X - \mu)/\sigma$

If $X \sim N(\mu, \sigma^2)$, define:

$$Z = \frac{X - \mu}{\sigma}$$

Why does $Z \sim N(0, 1)$? Two steps:

Step 1: Z is still normal, because affine transformations preserve normality.

Step 2: Compute its mean and variance using rules you already know:

$$\mathbb{E}[Z] = \mathbb{E}\left[\frac{X - \mu}{\sigma}\right] = \frac{1}{\sigma}\left(\mathbb{E}[X] - \mu\right) = \frac{1}{\sigma}(\mu - \mu) = 0$$

$$\text{Var}(Z) = \text{Var}\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2}\,\text{Var}(X) = \frac{\sigma^2}{\sigma^2} = 1$$

Standardization: $Z = (X - \mu)/\sigma$

If $X \sim N(\mu, \sigma^2)$, define:

$$Z = \frac{X - \mu}{\sigma}$$

Why does $Z \sim N(0, 1)$? Two steps:

Step 1: Z is still normal, because affine transformations preserve normality.

Step 2: Compute its mean and variance using rules you already know:

$$\mathbb{E}[Z] = \mathbb{E}\left[\frac{X - \mu}{\sigma}\right] = \frac{1}{\sigma}\left(\mathbb{E}[X] - \mu\right) = \frac{1}{\sigma}(\mu - \mu) = 0$$

$$\text{Var}(Z) = \text{Var}\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2}\,\text{Var}(X) = \frac{\sigma^2}{\sigma^2} = 1$$

$\Longrightarrow$ Z is normal with mean 0 and variance 1, so $Z \sim N(0, 1)$.

The Standard Normal CDF: $\Phi(z)$

The CDF of $Z \sim N(0, 1)$ is denoted $\Phi(z) = P(Z \leq z)$.

The Standard Normal CDF: $\Phi(z)$

The CDF of $Z \sim N(0, 1)$ is denoted $\Phi(z) = P(Z \leq z)$.

For any $X \sim N(\mu, \sigma^2)$, we can compute probabilities by standardizing:

$$P(X \leq a) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X > a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

The Standard Normal CDF: $\Phi(z)$

The CDF of $Z \sim N(0, 1)$ is denoted $\Phi(z) = P(Z \leq z)$.

For any $X \sim N(\mu, \sigma^2)$, we can compute probabilities by standardizing:

$$P(X \leq a) = \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P(X > a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P(a \leq X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

In practice, you look up Φ in a table (Appendix D, Table 1) or use software:

- R: `pnorm(z)`, Python: `scipy.stats.norm.cdf(z)`

Example: Computing a Normal Probability

Suppose test scores follow $X \sim N(70, 100)$, so $\mu = 70$ and $\sigma = 10$.

What is $P(60 \leq X \leq 85)$?

Example: Computing a Normal Probability

Suppose test scores follow $X \sim N(70, 100)$, so $\mu = 70$ and $\sigma = 10$.

What is $P(60 \leq X \leq 85)$**?**

Step 1: Standardize both endpoints.

$$Z_{\text{low}} = \frac{60 - 70}{10} = -1, \qquad Z_{\text{high}} = \frac{85 - 70}{10} = 1.5$$

Example: Computing a Normal Probability

Suppose test scores follow $X \sim N(70, 100)$, so $\mu = 70$ and $\sigma = 10$.

What is $P(60 \leq X \leq 85)$?

Step 1: Standardize both endpoints.

$$Z_{\text{low}} = \frac{60 - 70}{10} = -1, \qquad Z_{\text{high}} = \frac{85 - 70}{10} = 1.5$$

Step 2: Use the standard normal CDF.

$$P(60 \leq X \leq 85) = \Phi(1.5) - \Phi(-1)$$

Example: Computing a Normal Probability

Suppose test scores follow $X \sim N(70, 100)$, so $\mu = 70$ and $\sigma = 10$.

What is $P(60 \leq X \leq 85)$?

Step 1: Standardize both endpoints.

$$Z_{\text{low}} = \frac{60 - 70}{10} = -1, \qquad Z_{\text{high}} = \frac{85 - 70}{10} = 1.5$$

Step 2: Use the standard normal CDF.

$$P(60 \leq X \leq 85) = \Phi(1.5) - \Phi(-1)$$

Step 3: Look up values. $\Phi(1.5) = 0.9332$, $\Phi(-1) = 0.1587$.

$$P(60 \leq X \leq 85) = 0.9332 - 0.1587 = \boxed{0.7745}$$

Example: Computing a Normal Probability

Suppose test scores follow $X \sim N(70, 100)$, so $\mu = 70$ and $\sigma = 10$.

What is $P(60 \leq X \leq 85)$?

Step 1: Standardize both endpoints.

$$Z_{\text{low}} = \frac{60 - 70}{10} = -1, \qquad Z_{\text{high}} = \frac{85 - 70}{10} = 1.5$$

Step 2: Use the standard normal CDF.

$$P(60 \leq X \leq 85) = \Phi(1.5) - \Phi(-1)$$

Step 3: Look up values. $\Phi(1.5) = 0.9332$, $\Phi(-1) = 0.1587$.

$$P(60 \leq X \leq 85) = 0.9332 - 0.1587 = \boxed{0.7745}$$

$\implies$ About 77% of students score between 60 and 85.

Sums of Normal Random Variables

If X_1 and X_2 are **independent** normal random variables, then any linear combination is also normal:

Sums of Normal Random Variables

If X_1 and X_2 are **independent** normal random variables, then any linear combination is also normal:

If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent, then:

$$Y = a_1X_1 + a_2X_2 \sim N\left(a_1\mu_1 + a_2\mu_2,\ a_1^2\sigma_1^2 + a_2^2\sigma_2^2\right)$$

Sums of Normal Random Variables

If X_1 and X_2 are **independent** normal random variables, then any linear combination is also normal:

If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent, then:

$$Y = a_1 X_1 + a_2 X_2 \sim N\left(a_1\mu_1 + a_2\mu_2,\ a_1^2\sigma_1^2 + a_2^2\sigma_2^2\right)$$

More generally, if X_1 and X_2 are normal but *not* independent, with $\text{Cov}(X_1, X_2) = \sigma_{12}$:

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

Sums of Normal Random Variables

If X_1 and X_2 are **independent** normal random variables, then any linear combination is also normal:

If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent, then:

$$Y = a_1 X_1 + a_2 X_2 \sim N\left(a_1\mu_1 + a_2\mu_2,\ a_1^2\sigma_1^2 + a_2^2\sigma_2^2\right)$$

More generally, if X_1 and X_2 are normal but *not* independent, with $\text{Cov}(X_1, X_2) = \sigma_{12}$:

$$Y = a_1 X_1 + a_2 X_2 \sim N\left(a_1\mu_1 + a_2\mu_2,\ a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + 2a_1a_2\sigma_{12}\right)$$

$\implies$ The normal family is "closed" under addition. This will be essential when we study the sampling distribution of $\bar{X}$.

Outline

From Population to Sample

Recall the distinction from last time:

	Population	**Sample**
Quantity	Parameter (fixed, unknown)	Statistic (computed from data)
Mean	$\mu = \mathbb{E}[X]$	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
Variance	$\sigma^2 = \mathrm{Var}(X)$	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
Notation	Greek (μ, σ^2)	Roman ($\bar{X}, S^2$)

From Population to Sample

Recall the distinction from last time:

	Population	**Sample**
Quantity	Parameter (fixed, unknown)	Statistic (computed from data)
Mean	$\mu = \mathbb{E}[X]$	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
Variance	$\sigma^2 = \text{Var}(X)$	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
Notation	Greek (μ, σ^2)	Roman ($\bar{X}, S^2$)

The sample mean $\bar{X}$ is an **estimator**: a function of the random sample $(X_1, \ldots, X_n)$.

From Population to Sample

Recall the distinction from last time:

	Population	**Sample**
Quantity	Parameter (fixed, unknown)	Statistic (computed from data)
Mean	$\mu = \mathbb{E}[X]$	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
Variance	$\sigma^2 = \text{Var}(X)$	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
Notation	Greek (μ, σ^2)	Roman ($\bar{X}, S^2$)

The sample mean $\bar{X}$ is an **estimator**: a function of the random sample $(X_1, \ldots, X_n)$.

Because the sample is random, $\bar{X}$ is itself a random variable. If we drew a different sample, we would get a different $\bar{X}$.

What Is a Sampling Distribution?

Imagine repeating an experiment many times:

1. Draw a random sample of size n from the population
2. Compute $\bar{X}$ from that sample
3. Record it
4. Repeat (thousands of times)

What Is a Sampling Distribution?

Imagine repeating an experiment many times:

1. Draw a random sample of size n from the population
2. Compute $\bar{X}$ from that sample
3. Record it
4. Repeat (thousands of times)

The **sampling distribution of** $\bar{X}$ is the probability distribution of $\bar{X}$ across all possible samples of size n.

What Is a Sampling Distribution?

Imagine repeating an experiment many times:

1. Draw a random sample of size n from the population
2. Compute $\bar{X}$ from that sample
3. Record it
4. Repeat (thousands of times)

The **sampling distribution of** $\bar{X}$ is the probability distribution of $\bar{X}$ across all possible samples of size n.

Two results you can verify with the expectation rules from last time:

$$\mathbb{E}[\bar{X}] = \mu \quad \text{(unbiased: centered at the truth)}$$

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n} \quad \text{(shrinks as } n \text{ grows)}$$

What Is a Sampling Distribution?

Imagine repeating an experiment many times:

1. Draw a random sample of size n from the population
2. Compute $\bar{X}$ from that sample
3. Record it
4. Repeat (thousands of times)

The **sampling distribution of $\bar{X}$** is the probability distribution of $\bar{X}$ across all possible samples of size n.

Two results you can verify with the expectation rules from last time:

$$\mathbb{E}[\bar{X}] = \mu \quad \text{(unbiased: centered at the truth)}$$

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n} \quad \text{(shrinks as } n \text{ grows)}$$

$\Longrightarrow$ Larger samples produce more precise estimates. But what *shape* does the distribution of $\bar{X}$ have?

If the Population Is Normal

When $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, the sample mean is:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

If the Population Is Normal

When $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, the sample mean is:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

This is a linear combination of independent normal random variables (with $a_i = 1/n$ for each). By the “sums of normals” property:

$$\bar{X} \sim N\left(\mu, \ \frac{\sigma^2}{n}\right) \quad \text{for any } n$$

If the Population Is Normal

When $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, the sample mean is:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

This is a linear combination of independent normal random variables (with $a_i = 1/n$ for each). By the "sums of normals" property:

$$\bar{X} \sim N\left(\mu, \ \frac{\sigma^2}{n}\right) \quad \text{for any } n$$

$\implies$ If the population is normal, the sampling distribution of $\bar{X}$ is **exactly normal** for any sample size.

But What If the Population Is Not Normal?

We started with household income: right-skewed, clearly not bell-shaped.

But What If the Population Is Not Normal?

We started with household income: right-skewed, clearly not bell-shaped.

If the normal-population result were the end of the story, it would be nearly useless for econometrics. Real data is messy:

- Wages, income, wealth: heavily right-skewed
- Binary outcomes (employed/not): discrete, not continuous
- Test scores: often left-skewed or bounded

But What If the Population Is Not Normal?

We started with household income: right-skewed, clearly not bell-shaped.

If the normal-population result were the end of the story, it would be nearly useless for econometrics. Real data is messy:

- Wages, income, wealth: heavily right-skewed
- Binary outcomes (employed/not): discrete, not continuous
- Test scores: often left-skewed or bounded

We need a result that works **without assuming the population is normal**.

But What If the Population Is Not Normal?

We started with household income: right-skewed, clearly not bell-shaped.

If the normal-population result were the end of the story, it would be nearly useless for econometrics. Real data is messy:

- Wages, income, wealth: heavily right-skewed
- Binary outcomes (employed/not): discrete, not continuous
- Test scores: often left-skewed or bounded

We need a result that works **without assuming the population is normal**.

That result is the Central Limit Theorem.

Outline

An Experiment with a Skewed Population

Take a population that is **nothing like normal**: the Exponential(1) distribution.

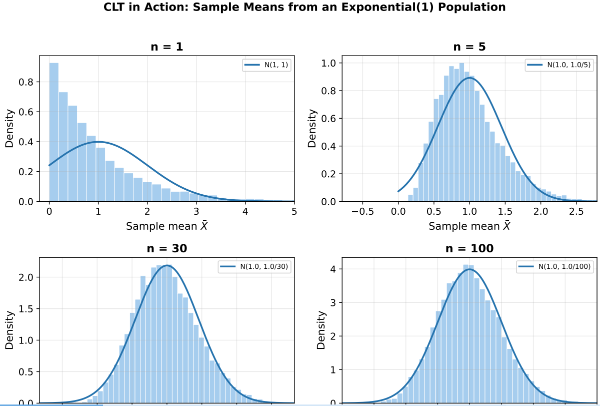
It has mean $= 1$, variance $= 1$, and is heavily right-skewed (probability is concentrated near zero, with a long right tail).

An Experiment with a Skewed Population

Take a population that is **nothing like normal**: the Exponential(1) distribution.

It has mean $= 1$, variance $= 1$, and is heavily right-skewed (probability is concentrated near zero, with a long right tail).

For each sample size n, draw 10,000 samples and plot the histogram of $\bar{X}$:



What the Simulation Shows

- $n = 1$: Each “sample mean” is just one draw from the population. The histogram is right-skewed (it *is* the population distribution).

What the Simulation Shows

- $n = 1$: Each "sample mean" is just one draw from the population. The histogram is right-skewed (it *is* the population distribution).
- $n = 5$: Already less skewed. The distribution of $\bar{X}$ is tightening around $\mu = 1$.

What the Simulation Shows

- $n = 1$: Each "sample mean" is just one draw from the population. The histogram is right-skewed (it *is* the population distribution).
- $n = 5$: Already less skewed. The distribution of $\bar{X}$ is tightening around $\mu = 1$.
- $n = 30$: Approximately bell-shaped. The normal curve fits well.

What the Simulation Shows

- $n = 1$: Each "sample mean" is just one draw from the population. The histogram is right-skewed (it *is* the population distribution).
- $n = 5$: Already less skewed. The distribution of $\bar{X}$ is tightening around $\mu = 1$.
- $n = 30$: Approximately bell-shaped. The normal curve fits well.
- $n = 100$: Very close to normal. Spread has shrunk to $\sigma/\sqrt{n} = 1/10 = 0.1$.

What the Simulation Shows

- $n = 1$: Each "sample mean" is just one draw from the population. The histogram is right-skewed (it *is* the population distribution).
- $n = 5$: Already less skewed. The distribution of $\bar{X}$ is tightening around $\mu = 1$.
- $n = 30$: Approximately bell-shaped. The normal curve fits well.
- $n = 100$: Very close to normal. Spread has shrunk to $\sigma/\sqrt{n} = 1/10 = 0.1$.

The population was nowhere near normal. Yet the distribution of $\bar{X}$ converged to a bell curve as n grew. Why?

The Central Limit Theorem

Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and finite variance σ^2. Then as $n \to \infty$:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ is approximately distributed as } N(0,1)$$

The Central Limit Theorem

Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and finite variance σ^2. Then as $n \to \infty$:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \text{is approximately distributed as} \quad N(0, 1)$$

In words: **no matter what the population looks like**, the standardized sample mean is approximately standard normal for large enough n.

The Central Limit Theorem

Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and finite variance σ^2. Then as $n \to \infty$:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ is approximately distributed as } N(0,1)$$

In words: **no matter what the population looks like**, the standardized sample mean is approximately standard normal for large enough n.

Equivalently:

$$\bar{X} \overset{\text{approx.}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{for large } n$$

The Central Limit Theorem

Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and finite variance σ^2. Then as $n \to \infty$:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ is approximately distributed as } N(0, 1)$$

In words: **no matter what the population looks like**, the standardized sample mean is approximately standard normal for large enough n.

Equivalently:

$$\bar{X} \overset{\text{approx.}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{for large } n$$

The population can be skewed, bimodal, discrete, continuous, bounded, unbounded. It does not need to be normal. The CLT applies as long as the population has **finite variance**.

Example: Applying the CLT

A factory fills cereal boxes. The fill weight per box has $\mu = 368$ g and $\sigma = 15$ g. The distribution of individual box weights is right-skewed (not normal).

A quality inspector weighs $n = 36$ boxes. What is the probability that $\bar{X} > 375$ g?

Example: Applying the CLT

A factory fills cereal boxes. The fill weight per box has $\mu = 368\,\text{g}$ and $\sigma = 15\,\text{g}$. The distribution of individual box weights is right-skewed (not normal).

A quality inspector weighs $n = 36$ boxes. What is the probability that $\bar{X} > 375\,\text{g}$?

By the CLT: $\bar{X} \overset{\text{approx.}}{\sim} N\left(368,\ \frac{15^2}{36}\right) = N(368,\ 6.25)$

Example: Applying the CLT

A factory fills cereal boxes. The fill weight per box has $\mu = 368$ g and $\sigma = 15$ g. The distribution of individual box weights is right-skewed (not normal).

A quality inspector weighs $n = 36$ boxes. What is the probability that $\bar{X} > 375$ g?

By the CLT: $\bar{X} \overset{\text{approx.}}{\sim} N\left(368,\ \frac{15^2}{36}\right) = N(368,\ 6.25)$

Standardize:

$$Z = \frac{375 - 368}{\sqrt{6.25}} = \frac{7}{2.5} = 2.8$$

Example: Applying the CLT

A factory fills cereal boxes. The fill weight per box has $\mu = 368\,\text{g}$ and $\sigma = 15\,\text{g}$. The distribution of individual box weights is right-skewed (not normal).

A quality inspector weighs $n = 36$ boxes. What is the probability that $\bar{X} > 375\,\text{g}$?

By the CLT: $\bar{X} \overset{\text{approx.}}{\sim} N\left(368,\ \frac{15^2}{36}\right) = N(368,\ 6.25)$

Standardize:

$$Z = \frac{375 - 368}{\sqrt{6.25}} = \frac{7}{2.5} = 2.8$$

$$P(\bar{X} > 375) = 1 - \Phi(2.8) = 1 - 0.9974 = \boxed{0.0026}$$

Example: Applying the CLT

A factory fills cereal boxes. The fill weight per box has $\mu = 368$ g and $\sigma = 15$ g. The distribution of individual box weights is right-skewed (not normal).

A quality inspector weighs $n = 36$ boxes. What is the probability that $\bar{X} > 375$ g?

By the CLT: $\bar{X} \overset{\text{approx.}}{\sim} N\left(368,\ \frac{15^2}{36}\right) = N(368,\ 6.25)$

Standardize:

$$Z = \frac{375 - 368}{\sqrt{6.25}} = \frac{7}{2.5} = 2.8$$

$$P(\bar{X} > 375) = 1 - \Phi(2.8) = 1 - 0.9974 = \boxed{0.0026}$$

$\implies$ Even though individual box weights are skewed, we can still compute probabilities about $\bar{X}$ using the normal distribution.

How Large Is “Large Enough”?

The CLT is an asymptotic result ($n \to \infty$), but in practice:

- For **symmetric** populations: $n \geq 5$ often suffices
- For **moderately skewed** populations: $n \geq 30$ is a common rule of thumb
- For **heavily skewed** populations: may need $n \geq 100$ or more

How Large Is "Large Enough"?

The CLT is an asymptotic result ($n \to \infty$), but in practice:

- For **symmetric** populations: $n \geq 5$ often suffices
- For **moderately skewed** populations: $n \geq 30$ is a common rule of thumb
- For **heavily skewed** populations: may need $n \geq 100$ or more

The more the population deviates from normality, the larger n must be for the approximation to be good.

How Large Is "Large Enough"?

The CLT is an asymptotic result ($n \to \infty$), but in practice:

- For **symmetric** populations: $n \geq 5$ often suffices
- For **moderately skewed** populations: $n \geq 30$ is a common rule of thumb
- For **heavily skewed** populations: may need $n \geq 100$ or more

The more the population deviates from normality, the larger n must be for the approximation to be good.

In econometrics, sample sizes are typically in the hundreds or thousands. $\implies$ The CLT approximation is usually excellent.

Why the CLT Is Foundational for Econometrics

Return to our opening example. Household income is right-skewed, not normal. So how can we do inference about average income, or about regression coefficients?

Why the CLT Is Foundational for Econometrics

Return to our opening example. Household income is right-skewed, not normal. So how can we do inference about average income, or about regression coefficients?

The OLS estimator $\hat{\beta}$ is (roughly) a weighted average of the data, similar to a sample mean.

Why the CLT Is Foundational for Econometrics

Return to our opening example. Household income is right-skewed, not normal. So how can we do inference about average income, or about regression coefficients?

The OLS estimator $\hat{\beta}$ is (roughly) a weighted average of the data, similar to a sample mean.

The CLT tells us:

1. $\hat{\beta}$ is approximately normally distributed in large samples

Why the CLT Is Foundational for Econometrics

Return to our opening example. Household income is right-skewed, not normal. So how can we do inference about average income, or about regression coefficients?

The OLS estimator $\hat{\beta}$ is (roughly) a weighted average of the data, similar to a sample mean.

The CLT tells us:

1. $\hat{\beta}$ is approximately normally distributed in large samples
2. This is true even if the error terms ε_i are not normal

Why the CLT Is Foundational for Econometrics

Return to our opening example. Household income is right-skewed, not normal. So how can we do inference about average income, or about regression coefficients?

The OLS estimator $\hat{\beta}$ is (roughly) a weighted average of the data, similar to a sample mean.

The CLT tells us:

1. $\hat{\beta}$ is approximately normally distributed in large samples
2. This is true even if the error terms ε_i are not normal
3. We can build confidence intervals and hypothesis tests using the normal distribution

Why the CLT Is Foundational for Econometrics

Return to our opening example. Household income is right-skewed, not normal. So how can we do inference about average income, or about regression coefficients?

The OLS estimator $\hat{\beta}$ is (roughly) a weighted average of the data, similar to a sample mean.

The CLT tells us:

1. $\hat{\beta}$ is approximately normally distributed in large samples
2. This is true even if the error terms ε_i are not normal
3. We can build confidence intervals and hypothesis tests using the normal distribution

$\implies$ Even though household income is skewed, the OLS estimate of the return to education is approximately normal in large samples, because $\hat{\beta}$ is a kind of average, and the CLT applies to averages.

Outline

From the CLT to Hypothesis Testing

The CLT says $\bar{X}$ is approximately normal in large samples. That lets us write:

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \overset{\text{approx.}}{\sim} N(0, 1)$$

From the CLT to Hypothesis Testing

The CLT says $\bar{X}$ is approximately normal in large samples. That lets us write:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{\text{approx.}}{\sim} N(0, 1)$$

But there is a problem: σ^2 is unknown. When we replace σ with the sample standard deviation S, the distribution changes.

From the CLT to Hypothesis Testing

The CLT says $\bar{X}$ is approximately normal in large samples. That lets us write:

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \overset{\text{approx.}}{\sim} N(0, 1)$$

But there is a problem: σ^2 is unknown. When we replace σ with the sample standard deviation S, the distribution changes.

Three distributions built from normals handle this and related situations. They are tools we will unpack fully when we reach hypothesis testing and confidence intervals (Topics 9–10). For now, just see how they are constructed from normal random variables.

Three Distributions Built from Normals

Let $Z, Z_1, \ldots, Z_m$ be independent $N(0, 1)$ random variables.

Three Distributions Built from Normals

Let $Z, Z_1, \ldots, Z_m$ be independent $N(0,1)$ random variables.

Chi-squared. $W = Z_1^2 + \cdots + Z_m^2 \sim \chi^2(m)$

- Sum of squared standard normals; always positive, right-skewed
- $\mathbb{E}[W] = m, \quad \mathrm{Var}(W) = 2m$

Three Distributions Built from Normals

Let $Z, Z_1, \ldots, Z_m$ be independent $N(0,1)$ random variables.

Chi-squared. $W = Z_1^2 + \cdots + Z_m^2 \sim \chi^2(m)$

- Sum of squared standard normals; always positive, right-skewed
- $\mathbb{E}[W] = m, \quad \text{Var}(W) = 2m$

Student's t. If $W \sim \chi^2(m)$ is independent of Z: $t = Z/\sqrt{W/m} \sim t(m)$

- Symmetric and bell-shaped like $N(0,1)$, but with **heavier tails**
- As $m \to \infty$, $t(m) \to N(0,1)$

Three Distributions Built from Normals

Let $Z, Z_1, \ldots, Z_m$ be independent $N(0,1)$ random variables.

Chi-squared. $W = Z_1^2 + \cdots + Z_m^2 \sim \chi^2(m)$

- Sum of squared standard normals; always positive, right-skewed
- $\mathbb{E}[W] = m, \quad \text{Var}(W) = 2m$

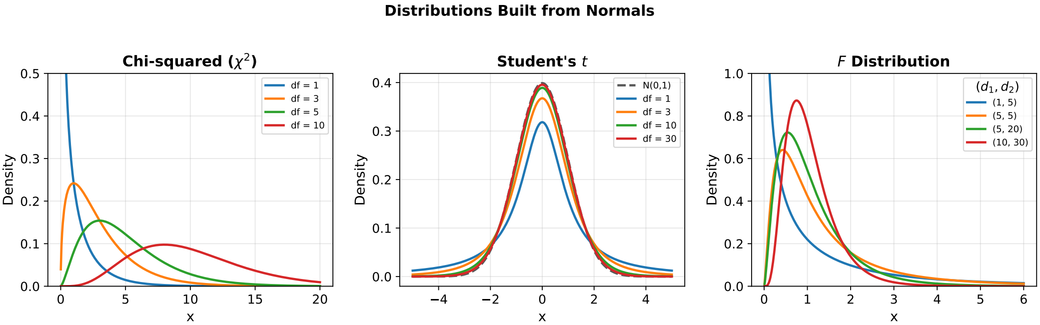
Student's t. If $W \sim \chi^2(m)$ is independent of Z: $t = Z\big/\sqrt{W/m} \sim t(m)$

- Symmetric and bell-shaped like $N(0,1)$, but with **heavier tails**
- As $m \to \infty$, $t(m) \to N(0,1)$

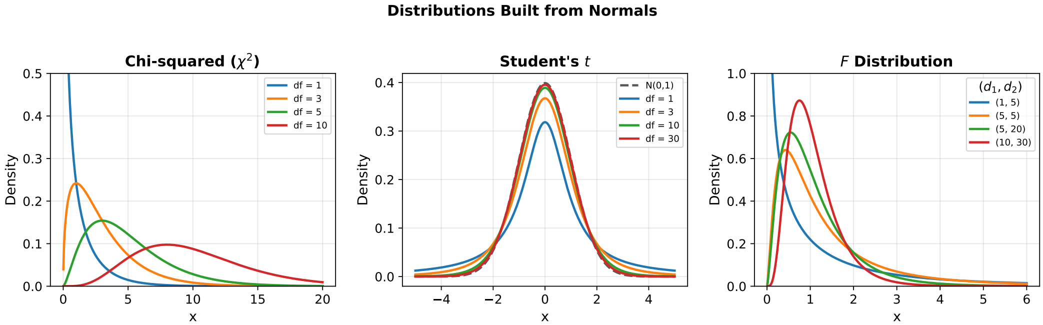
F distribution. If $W_1 \sim \chi^2(m_1)$, $W_2 \sim \chi^2(m_2)$ are independent: $F = \frac{W_1/m_1}{W_2/m_2} \sim F(m_1, m_2)$

- Always positive, right-skewed
- Connection: if $t \sim t(m)$, then $t^2 \sim F(1, m)$

Visualizing the Three Distributions



Visualizing the Three Distributions



The χ^2 and F are always positive and right-skewed. The t is symmetric, with tails that approach the normal as df increases.

Outline

Today's Takeaways

1. The **normal distribution** $N(\mu, \sigma^2)$ is symmetric and bell-shaped; the 68-95-99.7 rule quantifies its spread

Today's Takeaways

1. The **normal distribution** $N(\mu, \sigma^2)$ is symmetric and bell-shaped; the 68-95-99.7 rule quantifies its spread
2. **Standardization** $Z = (X - \mu)/\sigma$ converts any normal to $N(0, 1)$, so one CDF table handles all cases

Today's Takeaways

1. The **normal distribution** $N(\mu, \sigma^2)$ is symmetric and bell-shaped; the 68-95-99.7 rule quantifies its spread
2. **Standardization** $Z = (X - \mu)/\sigma$ converts any normal to $N(0, 1)$, so one CDF table handles all cases
3. The **sampling distribution** of $\bar{X}$ describes how the sample mean varies across repeated samples: $\mathbb{E}[\bar{X}] = \mu$, $\text{Var}(\bar{X}) = \sigma^2/n$

Today's Takeaways

1. The **normal distribution** $N(\mu, \sigma^2)$ is symmetric and bell-shaped; the 68-95-99.7 rule quantifies its spread
2. **Standardization** $Z = (X - \mu)/\sigma$ converts any normal to $N(0, 1)$, so one CDF table handles all cases
3. The **sampling distribution** of $\bar{X}$ describes how the sample mean varies across repeated samples: $\mathbb{E}[\bar{X}] = \mu$, $\text{Var}(\bar{X}) = \sigma^2/n$
4. The **CLT**: for large n, $\bar{X}$ is approximately normal regardless of the population shape (provided the population has finite variance)

Today's Takeaways

1. The **normal distribution** $N(\mu, \sigma^2)$ is symmetric and bell-shaped; the 68-95-99.7 rule quantifies its spread
2. **Standardization** $Z = (X - \mu)/\sigma$ converts any normal to $N(0, 1)$, so one CDF table handles all cases
3. The **sampling distribution** of $\bar{X}$ describes how the sample mean varies across repeated samples: $\mathbb{E}[\bar{X}] = \mu$, $\text{Var}(\bar{X}) = \sigma^2/n$
4. The **CLT**: for large n, $\bar{X}$ is approximately normal regardless of the population shape (provided the population has finite variance)
5. The χ^2, t, and F distributions are **built from normals** and will be our tools for hypothesis testing

Today's Takeaways

1. The **normal distribution** $N(\mu, \sigma^2)$ is symmetric and bell-shaped; the 68-95-99.7 rule quantifies its spread
2. **Standardization** $Z = (X - \mu)/\sigma$ converts any normal to $N(0, 1)$, so one CDF table handles all cases
3. The **sampling distribution** of $\bar{X}$ describes how the sample mean varies across repeated samples: $\mathbb{E}[\bar{X}] = \mu$, $\text{Var}(\bar{X}) = \sigma^2/n$
4. The **CLT**: for large n, $\bar{X}$ is approximately normal regardless of the population shape (provided the population has finite variance)
5. The χ^2, t, and F distributions are **built from normals** and will be our tools for hypothesis testing

$\implies$ The CLT is the reason we can do inference in econometrics without knowing the true error distribution.

Thank you!

jakeanderson@g.ucla.edu