Statistics Pre-course
Inference

Ilaria Mozzetta

Tor Vergata University of Rome

September 2026

Part III syllabus

1. Basics of Inference
2. Statistical Models
3. Central Limit Theorem
4. Law of Large Numbers
5. Point Estimation
6. Method of Moments
7. Method of Maximum Likelihood
8. Mean Squared Error and Consistency
9. Interval Estimation
10. Hypothesis Testing

Probability and Inference

Probability

Population $\rightarrow$ **Sample:** Starts from a known population model and predicts the likelihood of observing specific sample outcomes.

Statistical Inference

Sample $\rightarrow$ **Population:** Uses **descriptive statistics** calculated from an observed sample to draw conclusions about the underlying, unobserved population (or Data Generating Process).

The Big Picture

- **Descriptive Statistics** tell us what is happening in the data we *have*.
- **Inferential Statistics** lets us extrapolate those findings to the population we *care about*, accounting for sampling variation.

Inference

Once we collect data from a sample, statistical inference serves two main tasks:

1. Estimation

Using sample data to approximate unknown population characteristics (parameters):

- **Point Estimate:** A single best guess for the parameter (e.g., sample mean $\bar{x}$).
- **Interval Estimate:** A range of plausible values likely containing the true parameter.

2. Hypothesis Testing

Using sample data to assess evidence for or against a specific claim/prediction about the population.

From data to random variables

To analyze data mathematically, we must formalize how data is generated:

1. The Observed Data

We observe numerical outcomes (e.g., test scores, stock returns, device lifetimes).

2. The Random Variable (X)

We model the phenomenon of interest as a **random variable** X that takes values in a set $\mathcal{X}$ (the support).

3. The Data Generating Process (f_X^*)

The behavior of X is governed by an unknown probability law $f_X^*(x)$, called the **true Data Generating Process (DGP)**.

Parametric families

Since searching over *all possible* mathematical functions $f_X^*(x)$ is impossible, we simplify the problem:

Parametric Assumption

We assume the true distribution $f_X^*(x)$ belongs to a known family $\mathcal{F}$ defined by an unknown parameter $\theta \in \Theta$:

$$\mathcal{F} = \{f_X(x \mid \theta) : \theta \in \Theta\}$$

The Core Objective of Inference

We know the functional form (e.g., Normal, Poisson, Bernoulli), but not θ. Our goal is to use our observed sample data to estimate the true value $\theta^* \in \Theta$ that generated the data.

Elements of a statistical model

To characterize a random variable X, we need three basic elements:

- $\mathcal{X}$
- $f_X(x \mid \theta)$
- Θ

Statistical model

A statistical model is the triple:

$$\{\mathcal{X};\ f_X(x \mid \theta);\ \Theta\}$$

Note that

- X might be multidimensional
- $\Theta \subseteq \mathbb{R}$ or $\Theta \subseteq \mathbb{R} \times \mathbb{R}$

Exercise

Task: For each distribution, specify its triple $\{\mathcal{X};\ f_X(x \mid \theta);\ \Theta\}$.

Example: Bernoulli(p)

- **Support:** $\mathcal{X} = \{0, 1\}$
- **PMF:** $f_X(x \mid p) = p^x(1-p)^{1-x}$
- **Parameter Space:** $\Theta = [0, 1]$

Now try these:

1. Geometric(θ)
2. Normal(μ, σ^2)
3. Uniform($0, \theta$)

Random sample & observed sample

Random Sample (Random Variables)

A collection of n random variables $X_1, \ldots, X_n$ is a **Random Sample** if they are **independent and identically distributed (i.i.d.)**:

1. **Independent:** $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$
2. **Identically Distributed:** $f_{X_i}(x_i) = f_X(x_i) \quad \forall i$

Joint Distribution: $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f_X(x_i \mid \theta)$

Observed Sample (Realized Data)

An **observed sample** $(x_1, \ldots, x_n)$ consists of concrete numerical values (a single *realization* of the random vector $(X_1, \ldots, X_n)$).

Statistical model for a random sample

Combining our parametric family with the i.i.d. assumption gives us the complete joint model for an n-dimensional sample:

Joint Statistical Model

$$\left\{ \mathcal{X}^n;\ f_n(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f_X(x_i \mid \theta);\ \theta \in \Theta \right\}$$

- **Support ($\mathcal{X}^n$):** The n-fold Cartesian product of individual supports.
- **Joint Density (f_n):** The product of single-observation marginal densities.
- **Parameter Space (Θ):** The same governing parameter space for all observations.

Example

Let $X_1, \ldots, X_n$ be i.i.d from a Poisson(λ).

The joint distribution takes the form

$$\begin{aligned} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) &= \prod_{i=1}^{n} f_X(x_i) \\ &= \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \\ &= \frac{1}{\prod_{i=1}^{n} x_i!} e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i} \end{aligned}$$

Core concepts in inference

Parameter (θ): A numerical characteristic of the population that we want to learn about (typically unknown).
Examples: Population mean μ, variance σ^2, Poisson rate λ.

Estimator ($T(X_1, \ldots, X_n)$): A function of the random sample used to guess θ.

- **Must NOT depend on any unknown parameters**.
- It is a **random variable** with a sampling distribution.

Example: Sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Estimate ($\hat{\theta}$ or t): The concrete numerical value of an estimator calculated from observed sample data.

- It is a **fixed number** (a realization of the estimator).

Example: $\overline{x} = 101.7$.

Estimators: an easy case

If the parameter of interest is the population mean $\theta = \mathbb{E}[X]$, then the perfect candidate estimator is the so-called sample mean:

Sample mean

If we collect a sample of observations $X_1, X_2, \ldots, X_n$, then the sample mean is defined as:

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

Why should we use $\overline{X}$ to estimate $\theta = \mathbb{E}[X]$?

Two fundamental asymptotic results give $\overline{X}$ its power:

1. **Law of Large Numbers (LLN):** Guarantees that $\overline{X}$ converges to the true mean $\mathbb{E}[X]$.
2. **Central Limit Theorem (CLT):** Tells us the approximate probability distribution of $\overline{X}$ for large n.

Law of large numbers

Let $X_1, \ldots, X_n$ be an i.i.d. random sample with mean $\mathbb{E}[X_i] = \mu$ and finite variance σ^2.

Weak Law of Large Numbers (WLLN)

As sample size $n \to \infty$, the sample mean $\bar{X}_n$ converges in probability to the true population mean μ:

$$\overline{X} \xrightarrow{p} \mu$$

Takeaway for Inference: As we collect more data ($n \uparrow$), our estimator $\overline{X}$ gets arbitrarily close to the unknown truth μ.

Central limit theorem

Let $X_1, \ldots, X_n$ be i.i.d. from **any distribution** with mean μ and finite variance σ^2.

Central Limit Theorem (CLT)

As $n \to \infty$, the standardized sample mean converges in distribution to a Standard Normal $\mathcal{N}(0, 1)$:

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Takeaway for Inference:

- For large n, $\overline{X} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ regardless of the underlying population distribution.

Variability of estimators

To estimate the population mean IQ (μ) of Tor Vergata students, we interview $n = 10$ students and use the sample mean $\overline{X}$ as our estimator.

1st Observed Sample $(x_1, x_2, \ldots, x_n)$

$$\begin{aligned} x &= (x_1, x_2, \ldots, x_{10}) \\ &= (95, 104, 104, 95, 88, 126, 77, 112, 111, 105) \end{aligned}$$

Calculated Estimate

$$T(x_1, \ldots, x_{10}) = \overline{x}_1 = 101.7$$

Variability of estimators

If we collect a **second sample** of 10 students, we obtain a different result:

2nd Observed Sample $(x_1', x_2', \ldots, x_n')$

$$\begin{aligned} x' &= (x_1', x_2', \ldots, x_{10}') \\ &= (123, 119, 94, 116, 106, 91, 88, 107, 91, 103) \end{aligned}$$

Calculated Estimate

$$T(x_1', \ldots, x_{10}') = \overline{x}_2 = 103.8$$

Fundamental Conclusion

Because an **estimator** $(\overline{X})$ is a function of random variables, it is itself a **random variable**. The values 101.7 and 103.8 are specific numerical **estimates** (realizations).

Comments on estimators

There is no **"universal estimator"**. We must choose our formula based on:

- **The parameter of interest (θ)**
  - A formula designed to estimate a mean (μ) cannot be used to estimate a variance (σ^2).
- **The distribution of the data (f_X)**
  - An estimator that works well for continuous Normal data may not make sense for discrete counting data.

How do we find a point estimator?

The goal of an estimator is to learn about the unknown distribution that generated our data.

There are standard ways to construct an estimator, depending on our approach:

- **Method of Moments:**
  - Find a distribution that shares key features (like the mean) with our observed sample.
- **Maximum Likelihood:**
  - Find a distribution that maximizes the chance of observing the sample we actually collected.

Method of moments

The Matching Game

If a theoretical probability model describes our data, its theoretical summaries (**population moments**) should match our observed summaries (**sample moments**).

Intuitive Logic

- **Theoretical Mean ($\mathbb{E}[X]$):** A formula that depends on parameter θ.
- **Sample Mean ($\bar{X}$):** A single number calculated from our data.
- **The MoM Strategy:** Set them equal ($\mathbb{E}[X] = \bar{X}$) and solve for θ!

What is a Moment?

"Moments" are simply summary measures of a distribution's shape:

Population Moments
(Theoretical / Model)

- **1st Moment:**

$$\mu_1 = \mathbb{E}[X]$$

- **2nd Moment:**

$$\mu_2 = \mathbb{E}[X^2]$$

Sample Moments
(Calculated from Data)

- **1st Sample Moment:**

$$m_1 = \bar{X} = \frac{1}{n}\sum X_i$$

- **2nd Sample Moment:**

$$m_2 = \frac{1}{n}\sum X_i^2$$

Rule of Thumb

- **1 unknown parameter?** Match 1st moments: $\mathbb{E}[X] = \bar{X}$
- **2 unknown parameters?** Match 1st AND 2nd moments!

The 3-step recipe

Goal: Estimate the rate parameter λ of a Poisson(λ) random sample.

Step 1: Compute the theoretical mean $\mathbb{E}[X]$

From probability theory, we know that for a Poisson variable:

$$\mathbb{E}[X] = \lambda$$

Step 2: Equate theoretical mean to sample mean

Set population moment equal to sample moment:

$$\mathbb{E}[X] = \bar{X} \implies \lambda = \bar{X}$$

Step 3: Solve for the estimator

$$\hat{\lambda}_{\text{MoM}} = \bar{X}$$

Method of moments: k parameters

What if our distribution has k unknown parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$?

The General Rule: k Unknowns $\Longrightarrow$ k Equations

To estimate k parameters, we match the first k theoretical population moments to the first k sample moments.

The System of k Equations

Equate each theoretical moment to its sample counterpart and solve for $(\hat{\theta}_1, \ldots, \hat{\theta}_k)$:

$$\begin{cases} \mathbb{E}[X] &= \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \\ \mathbb{E}[X^2] &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\ &\vdots \\ \mathbb{E}[X^k] &= \frac{1}{n}\sum_{i=1}^{n} X_i^k \end{cases}$$

Example

Let $X_1, \ldots, X_n$ be a random sample from a population with probability law

$$f_X(x \mid \theta) = \frac{1}{\theta} x^{\frac{1-\theta}{\theta}}, \quad 0 < x < 1$$

Find the MoM estimator for θ.

Exercise

Consider a sample $X_1, \ldots, X_n \sim \mathcal{N}(1, \sigma^2)$. Find the MOM estimator for σ^2.

Example

$$\begin{aligned}\mathbb{E}[X] &= \int_0^1 x \cdot \left(\frac{1}{\theta} x^{\frac{1-\theta}{\theta}}\right) dx = \frac{1}{\theta}\int_0^1 x^{1+\frac{1-\theta}{\theta}}\, dx \\ &= \frac{1}{\theta}\int_0^1 x^{1/\theta}\, dx = \frac{1}{\theta}\left[\frac{x^{\frac{1}{\theta}+1}}{\frac{1}{\theta}+1}\right]_0^1 \\ &= \frac{1}{\theta}\cdot\frac{1}{\frac{1+\theta}{\theta}} = \frac{1}{1+\theta}\end{aligned}$$

Set population moment equal to sample moment $\bar{X}$:

$$\mathbb{E}[X] = \bar{X} \implies \frac{1}{1+\theta} = \bar{X}$$

Solving for θ yields the MoM estimator:

$$\hat{\theta}_{\text{MoM}} = \frac{1}{\bar{X}} - 1 = \frac{1-\bar{X}}{\bar{X}}$$

Maximum likelihood method

Let $X \sim \text{Binomial}(n, p)$. The probability mass function

$$\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

gives us the probability of observing a specific value x.

Assume that we know $n = 10$ and we suddenly observe $x = 8$.

- With $p = 0.5$: $\mathbb{P}(X = 8) = \binom{10}{8}(0.5)^8(0.5)^2 = 0.043$
- With $p = 0.7$: $\mathbb{P}(X = 8) = \binom{10}{8}(0.7)^8(0.3)^2 = 0.233$

For $x = 8$ a value of the parameter $p = 0.7$ seems more likely than $p = 0.5$.

Maximum likelihood method

When we fix a realization x and we consider the p.m.f. as a function of the unknown parameter p, the p.m.f. $\binom{n}{x} p^x (1-p)^{n-x}$ gives us a measure of how compatible p is.

This is called the **Likelihood** of p.

NB The Likelihood tells us how plausible a value of the parameter is, but it does not measure its probability.

The Core Question

Probability: Given parameter θ, what is the chance of seeing data **x**?
Likelihood: Given observed data **x**, how plausible is parameter θ?

Maximum likelihood method

When $n = 10$ and we observe $x = 8$ successes, the probability formula becomes a function of p:

$$f(p) = \mathbb{P}(X = 8 \mid p) = \binom{10}{8} p^8 (1-p)^2$$

$\Rightarrow$ Our goal is to choose the value of p that **maximizes** the probability of observing our data. We want to solve:

$$\hat{p} = \arg\max_p \mathbb{P}(X = 8 \mid p)$$

N.B.

- We are not maximizing p itself.
- We are maximizing the likelihood with respect to p.
- We choose the candidate $\hat{p}$ that makes the data we actually collected ($x = 8$) as likely as possible to happen.

The likelihood function

For an i.i.d. sample $(X_1, \ldots, X_n)$, the **Likelihood Function** is the joint density viewed as a function of θ:

$$\mathcal{L}(\theta \mid x_1, \ldots, x_n) = \prod_{i=1}^{n} f_X(x_i \mid \theta)$$

The Maximum Likelihood Estimator $\hat{\theta}_{\text{MLE}}$ is the parameter value that maximizes the likelihood of observing our sample:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \mathcal{L}(\theta \mid x_1, \ldots, x_n)$$

How to find $\hat{\theta}_{\text{MLE}}$

Steps to find the MLE:

1. **Write the Likelihood Function:** Construct $\mathcal{L}(\theta) = \prod_{i=1}^{n} f_X(x_i \mid \theta)$ using the PDF or PMF of the random sample.
2. **Take the Logarithm:** Compute $\ell(\theta) = \log \mathcal{L}(\theta)$ to convert the product $\prod$ into an easier sum $\sum$.
3. **Differentiate:** Compute the derivative with respect to θ:
$$\frac{d\ell(\theta)}{d\theta}$$
4. **Set to Zero & Solve:** Equate the derivative to zero and isolate θ to find candidate $\hat{\theta}$:
$$\frac{d\ell(\theta)}{d\theta} = 0$$
5. **Verify Maximum:** Check that the second derivative is strictly negative at $\hat{\theta}$:
$$\frac{d^2\ell(\theta)}{d\theta^2}\Big|_{\theta=\hat{\theta}} < 0$$

Example

Let $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$ i.i.d. Find $\hat{\lambda}_{\text{MLE}}$.

Step 1: Write the Likelihood Function

$$\mathcal{L}(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

Step 2: Take the Logarithm (Log-Likelihood)

$$\begin{aligned}
\ell(\lambda) &= \log\left(\frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}\right) \\
&= \log(e^{-n\lambda}) + \log\left(\lambda^{\sum_{i=1}^{n} x_i}\right) - \log\left(\prod_{i=1}^{n} x_i!\right) \\
&= -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\log(\lambda) - \sum_{i=1}^{n} \log(x_i!)
\end{aligned}$$

Step 3: Differentiate with respect to λ

$$\frac{d\ell(\lambda)}{d\lambda} = \frac{d}{d\lambda}\left[-n\lambda + \left(\sum_{i=1}^{n} x_i\right)\log(\lambda) - \sum_{i=1}^{n}\log(x_i!)\right] = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i$$

Step 4: Set derivative to zero and solve for λ

$$-n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0 \implies n = \frac{1}{\lambda}\sum_{i=1}^{n} x_i \implies \hat{\lambda}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}$$

Step 5: Verify second derivative

$$\frac{d^2\ell(\lambda)}{d\lambda^2} = -\frac{1}{\lambda^2}\sum_{i=1}^{n} x_i < 0 \quad \implies \quad \text{Confirmed Maximum}$$

Exercises

1. Consider a sample $X_1, \ldots, X_n$ of discrete random variables where $X_i \sim \text{Geometric}(\theta)$ with probability mass function given by

$$f_X(x \mid \theta) = \theta(1-\theta)^{x-1} \quad \forall x \in \mathbb{N}_+ \text{ and } 0 < \theta < 1$$

   Find the MLE for θ.

2. Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample drawn from a distribution with probability density function:

$$f(x \mid \theta) = \theta x^{-\theta-1}, \quad x \geq 1, \quad \theta > 0$$

   Find the MLE for θ.

How do we evaluate our estimators?

How do we judge if an estimator T is doing a "good job"? We look at two main qualities:

- **Bias:** On average, are we hitting the true target θ?

$$\text{Bias}(T) = \mathbb{E}[T] - \theta$$

Unbiased means $\text{Bias}(T) = 0 \rightarrow \mathbb{E}[T] = \theta$

- **Variance:** How spread out are our estimates across different samples?

$$\mathbb{V}[T] = \text{Sampling variability}$$

Low variance means our guesses are very similar.

The Overall Quality Score: Mean Squared Error

To pick the winner between two estimators, we combine both qualities into a single score:

$$\text{MSE}(\mathbf{T}) = \mathbb{V}[\mathbf{T}] + [\text{Bias}(\mathbf{T})]^{\mathbf{2}}$$

A great estimator has zero bias and low variance $\Rightarrow$ low MSE.

What if we collect more data?

What happens to our estimator as we collect more and more data?

- **Consistency** simply means: *"More data $\implies$ better guess."*
- As sample size n grows toward infinity, a consistent estimator gets infinitely close to the true value θ.
- **How do we test for consistency?** Check if the overall error (MSE) goes to zero as $n \to \infty$:

$$\lim_{n \to \infty} \text{MSE}(T_n) = 0$$

- This works as long as:
  1. The Bias drops to 0 (asymptotically unbiased).
  2. The Variance drops to 0 (uncertainty vanishes with more data).

Point estimates vs. intervals

Point Estimate: A single guess (e.g., sample mean $\bar{x}$).

- *"The temperature outside is exactly* $22^\circ C$*."*
- **The problem:** It is almost guaranteed to be slightly wrong!

Interval Estimate (Confidence Interval): A range of plausible values.

- *"I am 95% confident the temperature is between* $20^\circ C$ *and* $24^\circ C$*."*

What does "95% Confidence" actually mean?

It does **not** mean the true parameter has a 95% chance of moving!
It means: **If we take 100 different samples and build 100 intervals, roughly 95 of those intervals will catch the true population mean.**

Building a confidence interval

Confidence Interval (CI)

An interval estimator constructed from sample data, defined by two random endpoints $[L(\mathbf{X}), U(\mathbf{X})]$, that captures the true parameter θ with a specified probability $(1 - \alpha)$:

$$\mathbb{P}\left(L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})\right) = 1 - \alpha$$

where $1 - \alpha$ is the level of confidence.

Every Confidence Interval follows a simple formula:

$$\text{Interval} = \text{Point Estimate} \pm \text{Margin of Error}$$

Confidence interval for μ (σ known)

Deconstructing the formula

We assume that sample data comes from a **Normally distributed population** $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

Mathematical Formula

When population variance σ^2 is known, the $(1-\alpha)\%$ Confidence Interval for μ is:

$$\mathrm{CI}_{1-\alpha}(\mu) = \left[\bar{x} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

where:

- $\bar{\mathbf{x}}$ is the point estimate (the sample mean)
- $\mathbf{z_{1-\alpha/2}} \cdot \frac{\sigma}{\sqrt{\mathbf{n}}}$ is the margin error, composed of:
  - $\frac{\sigma}{\sqrt{\mathbf{n}}}$ is the standard error (spread of sample means)
  - $\mathbf{z_{1-\alpha/2}}$ is the critical value or quantile of the Normal distribution

The quantile $z_{1-\alpha/2}$

Mathematical Definition

For a Standard Normal variable $Z \sim \mathcal{N}(0,1)$ with Cumulative Distribution Function $\Phi(z) = \mathbb{P}(Z \leq z)$, the quantile $z_{1-\alpha/2}$ satisfies:

$$\Phi\left(z_{1-\alpha/2}\right) = \mathbb{P}\left(Z \leq z_{1-\alpha/2}\right) = 1 - \frac{\alpha}{2}$$

The expression $\Phi(z_{1-\alpha/2}) = 1 - \frac{\alpha}{2}$ states that $z_{1-\alpha/2}$ is the value on the horizontal axis for which our distribution accumulates:

- Exactly $1 - \frac{\alpha}{2}$ of its total probability to the left

Finding z-values

Standard Normal (Z) tables give the cumulative probability $\Phi(z)$ (the area to the left of z). To find a critical value, we read the table in reverse:

1. **Calculate the Target Area:** Determine the total area required to the left:
$$\text{Target Probability} = 1 - \frac{\alpha}{2}$$
2. **Search Inside the Table:** Scan the body of the Z-table to find the probability value closest to $1 - \frac{\alpha}{2}$.
3. **Read the z-Value:** Look at the corresponding row (first decimal) and column (second decimal) to read off $z_{1-\alpha/2}$.

Example

A factory produces light bulbs. The lifetime of these bulbs is normally distributed with a known population standard deviation of $\sigma = 100$ hours. A quality control manager takes a random sample of $n = 25$ bulbs and finds a sample mean lifetime of $\overline{x} = 1050$ hours.

Goal

Construct a 95% Confidence Interval for the true average lifetime μ of all produced light bulbs.

Given Data

- Sample mean: $\overline{x} = 1050$
- Known population standard deviation: $\sigma = 100$
- Sample size: $n = 25 \implies \sqrt{n} = 5$
- Confidence level: $1 - \alpha = 0.95 \implies \alpha = 0.05$

Example

Step 1: Find the Critical Value $z_{1-\alpha/2}$

- Target area: $1 - \frac{\alpha}{2} = 1 - \frac{0.05}{2} = 0.9750$
- From the Z-table: $z_{0.975} = \mathbf{1.96}$

Step 2: Calculate the Standard Error (SE)

$$\text{SE} = \frac{\sigma}{\sqrt{n}} = \frac{100}{\sqrt{25}} = \frac{100}{5} = \mathbf{20}$$

Step 3: Calculate the Margin of Error (ME)

$$\text{ME} = z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot 20 = \mathbf{39.2}$$

Step 4: Construct the Interval

$$\text{CI}_{0.95}(\mu) = [1050 \pm 39.2] = [\mathbf{1010.8}, \mathbf{1089.2}]$$

Interpretation

We are 95% confident that the true population mean lifetime μ of all light bulbs lies between 1010.8 and 1089.2 hours.

Exercise

A technology company produces lithium-ion batteries for smartphones. From long-term testing, the population standard deviation is known to be $\sigma = 4.5$ hours.
An engineer tests a random sample of $n = 36$ batteries under standard usage conditions and records a sample mean lifetime of $\overline{x} = 28.2$ hours.

1. Identify the target area $1 - \frac{\alpha}{2}$ for a 99% confidence interval ($\alpha = 0.01$).
2. Find the critical value $z_{1-\alpha/2}$ using the Z-table.
3. Calculate the Standard Error (SE) and Margin of Error (ME).
4. Construct the 99% Confidence Interval for the true population mean lifetime μ.
5. Interpret the results.

What if we don't know σ^2?

In practice, σ^2 is almost always unknown and must be estimated using the **sample variance** $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

$(1-\alpha)\%$ Confidence Interval

$$\mathrm{CI}_{1-\alpha}(\mu) = \left[\overline{x} - t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}, \overline{x} + t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}\right]$$

where:

- $\overline{\mathbf{x}}$ is the point estimate (the sample mean)
- $\mathbf{t_{n-1,1-\alpha/2}} \cdot \frac{\mathbf{s}}{\sqrt{\mathbf{n}}}$ is the margin error, composed of:
  - $\frac{\mathbf{s}}{\sqrt{\mathbf{n}}}$ is the standard error (spread of sample means)
  - $\mathbf{t_{n-1,1-\alpha/2}}$ is the critical value or quantile of the Student's t distribution

We assume that sample data comes from a **Normally distributed population** $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

Student's t

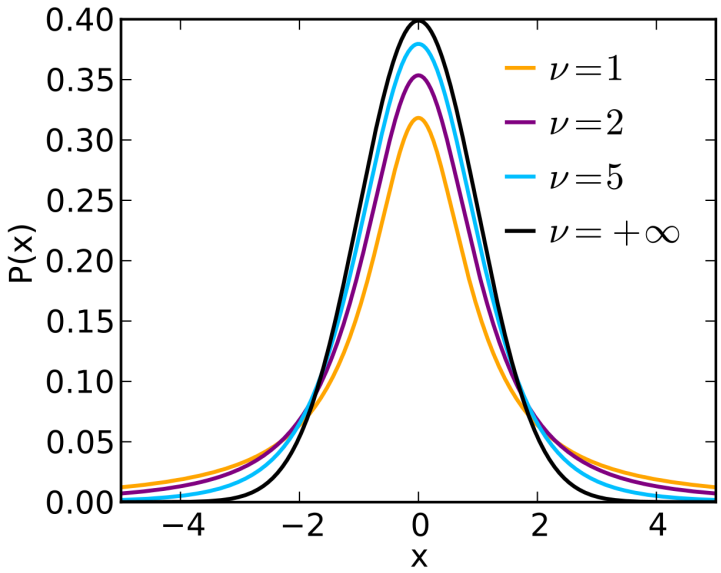
- **What is it?** A bell curve designed specifically for **small samples** ($n < 30$).
- **Why not just use the Normal distribution?**
  - With small data, we don't know the true population variance σ^2.
  - Estimating σ^2 with sample variance S^2 adds uncertainty.
- **The Solution:** The t-distribution adds "fat tails" to account for this uncertainty, making extreme values more likely.
- **Convergence to Normal:** As $\nu \to \infty$, $t(\nu)$ converges to the Standard Normal distribution:

$$t_{n-1} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as } n \to \infty$$

Degrees of Freedom ($\nu = n - 1$)

The shape of the t-distribution depends entirely on its degrees of freedom:

- As ν increases, the tails of the t-distribution become lighter.
- **Practical Rule:** For $\nu \geq 30$, the t-distribution is virtually indistinguishable from $\mathcal{N}(0, 1)$.



The quantile $t_{n-1,1-\alpha/2}$

Mathematical Definition

For a random variable $T \sim t_{n-1}$ following a Student's t-distribution with $\nu = n - 1$ degrees of freedom, the quantile $t_{n-1,1-\alpha/2}$ is the cutoff value satisfying:

$$\mathbb{P}\left(T_{n-1} \leq t_{n-1,1-\alpha/2}\right) = 1 - \frac{\alpha}{2} \quad \Longleftrightarrow \quad \mathbb{P}\left(T_{n-1} > t_{n-1,1-\alpha/2}\right) = \frac{\alpha}{2}$$

How to Find It in a t-Table:
Unlike Z-tables, t-tables are organized by upper-tail area ($\alpha/2$) and degrees of freedom:

1. **Identify Degrees of Freedom:** Calculate $\nu = n - 1$ and locate this row.
2. **Identify Tail Area:** Determine $\frac{\alpha}{2}$ (e.g., for 95% confidence, $\alpha = 0.05 \implies \frac{\alpha}{2} = 0.025$) and locate this column.
3. **Read the Value:** The intersection gives the critical value $t_{n-1,1-\alpha/2}$.

Example

A quality control manager monitors the weight of cereal boxes. The population variance σ^2 is unknown.
A random sample of $n = 16$ boxes is weighed, yielding:

- Sample mean weight: $\overline{x} = 502$ grams
- Sample standard deviation: $s = 8$ grams

Goal

Construct a 95% Confidence Interval for the true average weight μ of all cereal boxes.

Summary of Given Data

- Sample mean: $\overline{x} = 502$
- Sample std dev: $s = 8$
- Sample size: $n = 16 \implies$ Degrees of Freedom $\nu = n - 1 = 15$
- Confidence level: $1 - \alpha = 0.95 \implies \alpha = 0.05 \implies \frac{\alpha}{2} = 0.025$

Example

Step 1: Find the Critical Value $t_{n-1,1-\alpha/2}$

- Degrees of freedom: $\nu = 16 - 1 = 15$
- Upper-tail area: $\frac{\alpha}{2} = 0.025$
- From the t-table: $t_{15,\,0.975} = \mathbf{2.131}$

Step 2: Calculate the Estimated Standard Error (SE)

$$\text{SE} = \frac{s}{\sqrt{n}} = \frac{8}{\sqrt{16}} = \frac{8}{4} = \mathbf{2}$$

Step 3: Calculate the Margin of Error (ME)

$$\text{ME} = t_{n-1,1-\alpha/2} \cdot \frac{s}{\sqrt{n}} = 2.131 \cdot 2 = \mathbf{4.262}$$

Step 4: Construct the Interval

$$\text{CI}_{0.95}(\mu) = [502 \pm 4.262] = [\mathbf{497.74},\ \mathbf{506.26}]\ \text{grams}$$

Exercise

An automotive engineer tests nitrogen oxide (NOx) emissions from a new engine design. Population variance σ^2 is unknown.
A test sample of $n = 10$ engines is measured, yielding:

- Sample mean emission: $\overline{x} = 1.42$ g/km
- Sample standard deviation: $s = 0.15$ g/km

1. Determine the degrees of freedom ($\nu = n - 1$) and upper-tail area $\frac{\alpha}{2}$ for a 90% confidence interval ($\alpha = 0.10$).
2. Find the critical value $t_{n-1,\,1-\alpha/2}$ from the t-table.
3. Calculate the estimated Standard Error (SE) and Margin of Error (ME).
4. Construct the 90% Confidence Interval for the true mean emission μ.
5. Interpret your results.

What if our data are not Normal?

By the **Central Limit Theorem (CLT)**, as long as n is sufficiently large (typically $n \geq 30$), the sampling distribution of $\overline{X}$ is approximately Normal, regardless of the underlying population distribution.

Asymptotic $(1-\alpha)\%$ Confidence Interval

$$\text{CI}_{1-\alpha}(\mu) \approx \left[\overline{x} - z_{1-\alpha/2}\frac{s}{\sqrt{n}}, \overline{x} + z_{1-\alpha/2}\frac{s}{\sqrt{n}}\right]$$

Exercise

An analyst takes a sample of $n = 64$ transactions from an e-commerce platform and calculates a sample mean of $\overline{x} = \$42.50$ with a sample standard deviation of $s = \$16.00$. Calculate the approximate 95% Confidence Interval for the true mean customer spending μ.

Formulating hypotheses

When testing a parameter, we compare our unknown true mean μ to a baseline target value μ_0. The setup depends on whether we are looking for **any difference**, an **increase**, or a **decrease**.

1. Two-Tailed ($\neq$)

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

"Is the true mean different from μ_0?"

2. Right-Tailed ($>$)

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

"Is the true mean strictly greater than μ_0?"

3. Left-Tailed ($<$)

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

"Is the true mean strictly less than μ_0?"

How to evaluate our hypothesis

To test whether our hypothesis holds, we follow a structured decision process:

Step 1: Compute the Test Statistic: Standardize the distance between our estimate $\overline{x}$ and the target μ_0:

$$Z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \quad \left(\text{or } T = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \right)$$

How to decide if our hypothesis is correct

Step 2. Choose an Approach to Make a Decision:

Method A: Rejection Region

- Find the critical value t_{crit} for a given α.
- Construct the **rejection region**.
- **Rule:** Reject H_0 if T falls inside the critical region.

Method B: p-Value

- Compute p-value p.
- Compare p directly to significance level α.
- **Rule:** Reject H_0 if p-value $\leq \alpha$.

Step 3. Formulate the Decision: Either **Reject** H_0 (strong evidence for H_1) or **Fail to Reject** H_0 (insufficient evidence).

Method A: The rejection region approach

Think of the critical region as an "alarm zone." We draw line thresholds on our probability curve. If our sample's test statistic lands inside this zone, it is too unusual to be random noise, so we reject H_0.

Step-by-Step Mechanism

1. **Choose Significance Level (α):** Sets the maximum acceptable risk of making a false alarm (usually $\alpha = 0.05$ or 0.01).
2. **Find the Cutoff:** Look up the critical value from Z or t tables corresponding to α and sample size n.
3. **Define Rejection Region (R):** The set of values where H_0 is rejected.
4. **Compare & Decide:**

$$T_{\text{calc}} \in R \implies \textbf{Reject } H_0 \qquad T_{\text{calc}} \notin R \implies \textbf{Fail to Reject } H_0$$

Rejection regions by test type

The shape of the critical region depends entirely on the direction of our alternative hypothesis H_1:

Case 1: Population Variance σ^2 Known (Z-Distribution)

Alternative Hypothesis (H_1)	**Rejection Region**
$H_1 : \mu \neq \mu_0$	$\lvert Z_{\text{calc}} \rvert \geq z_{1-\alpha/2}$
$H_1 : \mu > \mu_0$	$Z_{\text{calc}} \geq z_{1-\alpha}$
$H_1 : \mu < \mu_0$	$Z_{\text{calc}} \leq -z_{1-\alpha}$

Case 2: Population Variance σ^2 Unknown (t-Distribution)

Alternative Hypothesis (H_1)	**Rejection Region**
$H_1 : \mu \neq \mu_0$	$\lvert T_{\text{calc}} \rvert \geq t_{n-1,\,1-\alpha/2}$
$H_1 : \mu > \mu_0$	$T_{\text{calc}} \geq t_{n-1,\,1-\alpha}$
$H_1 : \mu < \mu_0$	$T_{\text{calc}} \leq -t_{n-1,\,1-\alpha}$

Example

Context: A pharmaceutical company tests whether a new drug **increases** average reaction time beyond the established baseline of $\mu_0 = 100$ ms.

Given Data

- Sample size: $n = 16$ patients (df $= \nu = 15$)
- Sample mean: $\overline{x} = 108.6$ ms
- Sample standard deviation: $s = 16$ ms
- Significance level: $\alpha = 0.05$

Determine whether there is statistically significant evidence that the drug increases reaction time

Example

1. **Formulate Hypotheses:** Right-tailed test (H_1 tests for an increase):
$$H_0 : \mu \leq 100 \quad \text{vs.} \quad H_1 : \mu > 100$$
2. **Compute Test Statistic (T_{calc}):**
$$T_{\text{calc}} = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{108.6 - 100}{16/\sqrt{16}} = \frac{8.6}{4} = \mathbf{2.15}$$
3. **Find Critical Value & Region (R):** For $\nu = 15$ and upper-tail area $\alpha = 0.05$:
$$t_{\text{crit}} = t_{15,\,0.95} = \mathbf{1.753} \implies R = \{T \mid T \geq 1.753\}$$
4. **Compare & Decide:** Since $T_{\text{calc}} = 2.15 \geq 1.753$, T_{calc} falls **inside** the rejection region R.

Final Conclusion

Reject H_0. There is statistically significant evidence at the 5% significance level that the drug increases average reaction time.

Exercise

A tech company claims that a firmware update reduces the average boot-up time of its laptops below the historic benchmark of $\mu_0 = 45$ seconds. The company knows from long-term production logs that the population standard deviation is fixed at $\sigma = 6$ seconds.

Given Data

- Historical population standard deviation: $\sigma = 6$ seconds
- Sample size: $n = 36$ laptops
- Sample mean boot time: $\overline{x} = 43.1$ seconds
- Significance level: $\alpha = 0.01$

Determine if there is sufficient evidence that the firmware update reduces boot time.

Method B: The p-value approach

In simple words: The p-value measures *how surprising* our sample result is under the assumption that the null hypothesis H_0 is completely true.

Step-by-Step Mechanism

1. **Compute the Test Statistic:** Calculate Z_{calc} or T_{calc} as usual.
2. **Calculate the p-Value:** Find the probability of obtaining a test statistic as extreme as (or more extreme than) our calculated value under H_0.
3. **Compare with α & Decide:** Evaluate the p-value against our significance level α.

p-value $\leq \alpha \implies$ **Reject** H_0 $\qquad$ p-value $> \alpha \implies$ **Fail to Reject** H_0

N.B.

Rejection region approach and p-Value approach will always lead to the exact same conclusion for a given significance level α.

p-VALUE FORMULAS BY TEST TYPE

The exact calculation of the p-value depends on the sign of H_1.

Case 1: Population Variance σ^2 Known (Z-Distribution)

Alternative Hypothesis (H_1)	**p-Value Formula**
$H_1 : \mu \neq \mu_0$	$p = 2 \cdot \mathbb{P}(Z \geq \lvert Z_{\text{calc}} \rvert)$
$H_1 : \mu > \mu_0$	$p = \mathbb{P}(Z \geq Z_{\text{calc}})$
$H_1 : \mu < \mu_0$	$p = \mathbb{P}(Z \leq Z_{\text{calc}})$

Case 2: Population Variance σ^2 Unknown (t-Distribution)

Alternative Hypothesis (H_1)	**p-Value Formula**
$H_1 : \mu \neq \mu_0$	$p = 2 \cdot \mathbb{P}(T_{n-1} \geq \lvert T_{\text{calc}} \rvert)$
$H_1 : \mu > \mu_0$	$p = \mathbb{P}(T_{n-1} \geq T_{\text{calc}})$
$H_1 : \mu < \mu_0$	$p = \mathbb{P}(T_{n-1} \leq T_{\text{calc}})$

Example

A pharmaceutical company tests whether a new drug **increases** average reaction time beyond the established baseline of $\mu_0 = 100$ ms.

Given Data

- Sample size: $n = 16$ patients (df $= \nu = 15$)
- Sample mean: $\overline{x} = 108.6$ ms
- Sample standard deviation: $s = 16$ ms
- Significance level: $\alpha = 0.05$

Determine whether there is statistically significant evidence that the drug increases reaction time.

Example

1. **Formulate Hypotheses:** Right-tailed test (H_1 tests for an increase above 100 ms):

$$H_0 : \mu \leq 100 \quad \text{vs.} \quad H_1 : \mu > 100$$

2. **Compute Test Statistic (T_{calc}):**

$$T_{\text{calc}} = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{108.6 - 100}{16/\sqrt{16}} = \frac{8.6}{4} = \mathbf{2.15}$$

3. **Calculate p-Value:** For a right-tailed t-test with $\nu = 15$, check the upper tail area:

$$p = \mathbb{P}(T_{15} \geq 2.15) \approx \mathbf{0.024} \quad (2.4\%)$$

4. **Compare & Decide:** Since p-value $= 0.024 \leq \alpha = 0.05$, we **Reject** H_0.

Final Conclusion

Reject H_0. There is statistically significant evidence at the 5% level that the drug increases average reaction time.

Exercise

A manufacturing company tests whether a new automated assembly process **decreases** average production time below the standard baseline of $\mu_0 = 45$ minutes.

Given Data

- Historical population standard deviation: $\sigma = 8$ minutes (known)
- Sample size: $n = 64$ units
- Sample mean: $\overline{x} = 43.1$ minutes
- Significance level: $\alpha = 0.05$

Determine whether there is statistically significant evidence that the new process decreases average production time.