Confidence Interval

Problem: A company claims their fries have an average weight of 54 g per pack. You buy 25 random packs and weigh them:

41.4, 42.1, 42.6, 43.1, 44.0, 44.7, 45.4, 46.1, 46.8, 47.5, 48.2, 49.3, 50.0, 50.7, 51.8, 52.5, 53.2, 53.9, 54.6, 55.3, 56.0, 56.9, 57.4, 57.9, 58.6

The total of above 25 samples is 1250 g. So, average weight of single pack is 1250/25 = 50 g.

They seem lighter than the claim.

You calculate the standard deviation of above data and it is 8 g.

Can you *prove statistically,* at 95% confidence that the claim might be false?

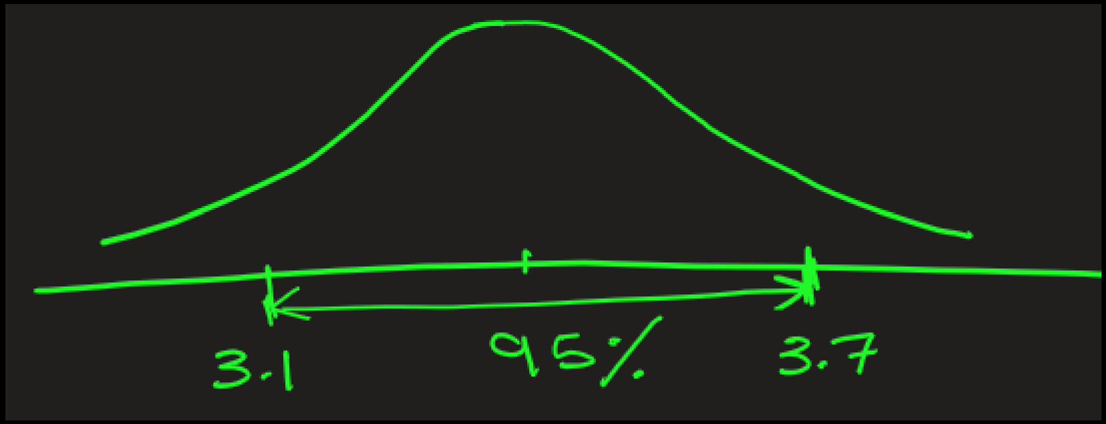
Confidence Interval: Definition

A **confidence interval (CI)** is a **range of values** that is likely to contain the **true population parameter** (like mean, proportion, or difference) with a certain level of confidence.
Confidence is generally expressed as 90%, 95% or 99%.

Example
Suppose you survey 100 students to find their **average study time per day**.
You find the **sample mean = 3.4 hours**, with a **95% confidence interval = [3.1, 3.7]**.

This means that:
- We are 95% confident that the true average study time for all students lies between 3.1 and 3.7 hours.

OR

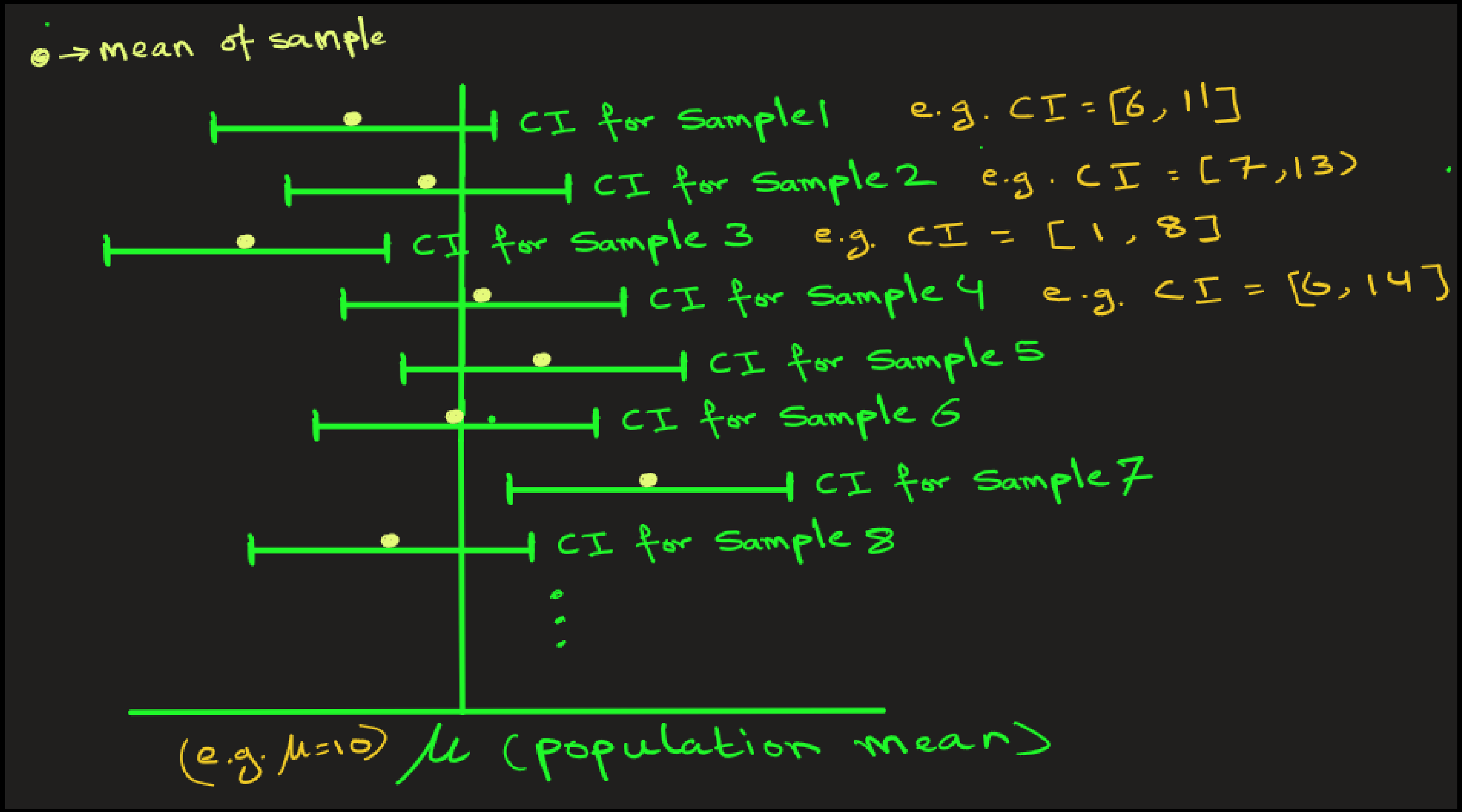


- Given the sample data, there is a **95% probability** that the true mean lies between 3.1 and 3.7.

Confidence Interval: Interpretation

1) **95% CI** means that if we repeatedly sample from the population and calculate CIs, about **95% of these CIs would contain the true mean μ**.

Let's understand this through an example:
Suppose the pop. mean is μ that we attempting to find.
I take sample1 and find its CI = [6,11]
I take sample2 and find its CI = [7,13]
I take sample3 and find its CI = [1,8]
I take sample1 and find its CI = [6,14]
and so on…
95% of these Cis would contain the true mean μ



2) Width of CI tells how precise the estimate is.
The **narrow CI,** say [2, 5], is more precise than **wider CI**, say [2, 30].

Confidence Interval: Formula for z-distribution

Formula for mean:

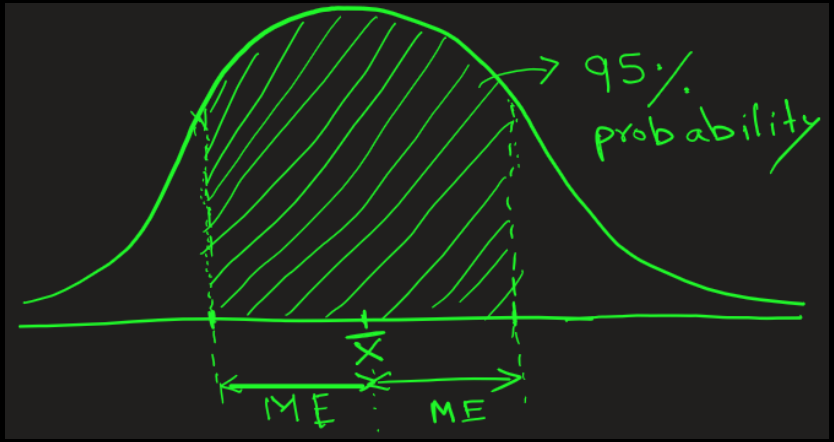
If population SD σ is known:

$$\text{Confidence Interval} = \bar{X} \pm Z \times \frac{\sigma}{\sqrt{n}}$$

where:

- $\bar{X}$ = sample mean
- Z = z-value corresponding to confidence level (1.96 for 95%, 2.58 for 99%)
- σ = population standard deviation
- n = sample size

Term $z\left(\frac{\sigma}{\sqrt{n}}\right)$ is called the **margin of error ME** (AKA **maximum error of the estimate**).



Another way of writing above is

$$\bar{x} - z \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z \cdot \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} - M.E < \mu < \bar{x} + ME$$

where, μ is the population mean.

2 main assumptions for finding a confidence interval for a mean when σ Is Known:

1) The sample is a random sample
2) Either n >= 30 or the population is normally distributed when n < 30

Confidence Interval: Formula For t-distribution

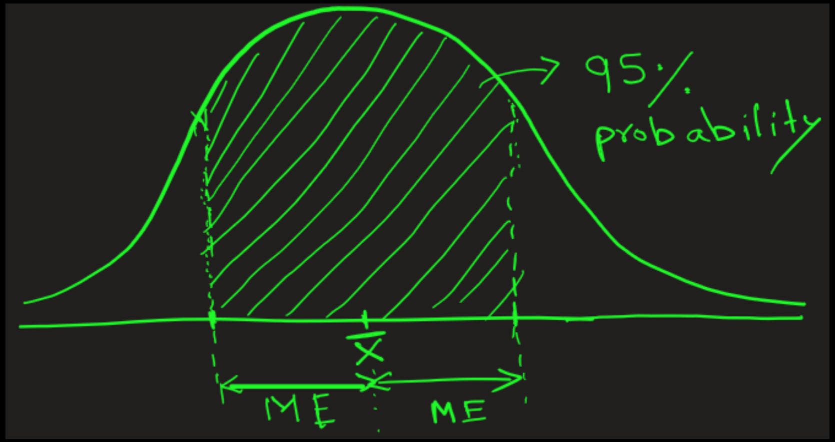
Formula for mean:

Previously we used z-distribution because population SD σ was known. If population SD is unknown, then we use **t-distribution** instead.
The degree of freedom d.f. is n-1. We use d.f. and confidence level to find t-values from the t-distribution table

Margin of error is defined similarly: t$\left(\frac{s}{\sqrt{n}}\right)$

$$\bar{x} - t\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\frac{s}{\sqrt{n}}$$

$$\bar{x} - M.E < \mu < \bar{x} + ME$$



Confidence Interval: When to Use the z or t Distribution

$\sigma \rightarrow$ pop. std. dev

$s \rightarrow$ sample std. dev.

Do we know σ ?

Yes → Use z-distribution & σ in formula

No → Use t-distribution & s in formula

If $n < 30$, the variable should be normally distributed

Confidence Interval: Steps

1. Decide which distribution to use (z or t) and corresponding critical z-value or t-value.

2. Compute the margin of error (ME): $z\left(\frac{\sigma}{\sqrt{n}}\right)$ or $t\left(\frac{s}{\sqrt{n}}\right)$

3. Construct the confidence interval

z- distribution

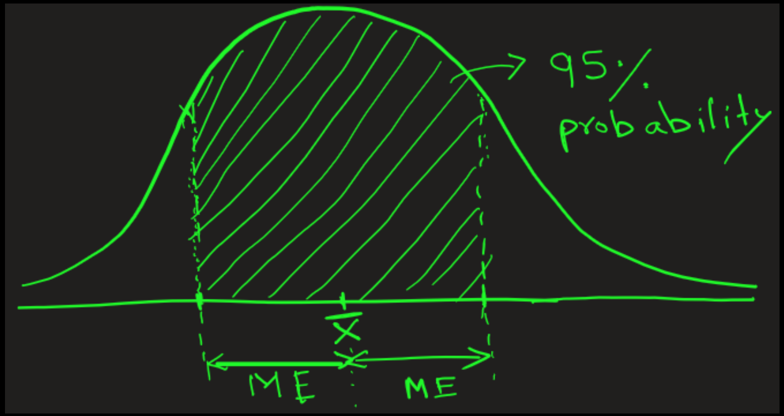
$$\bar{x} - z \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z \cdot \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} - M.E < \mu < \bar{x} + ME$$

t- distibution

$$\bar{x} - t\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\frac{s}{\sqrt{n}}$$

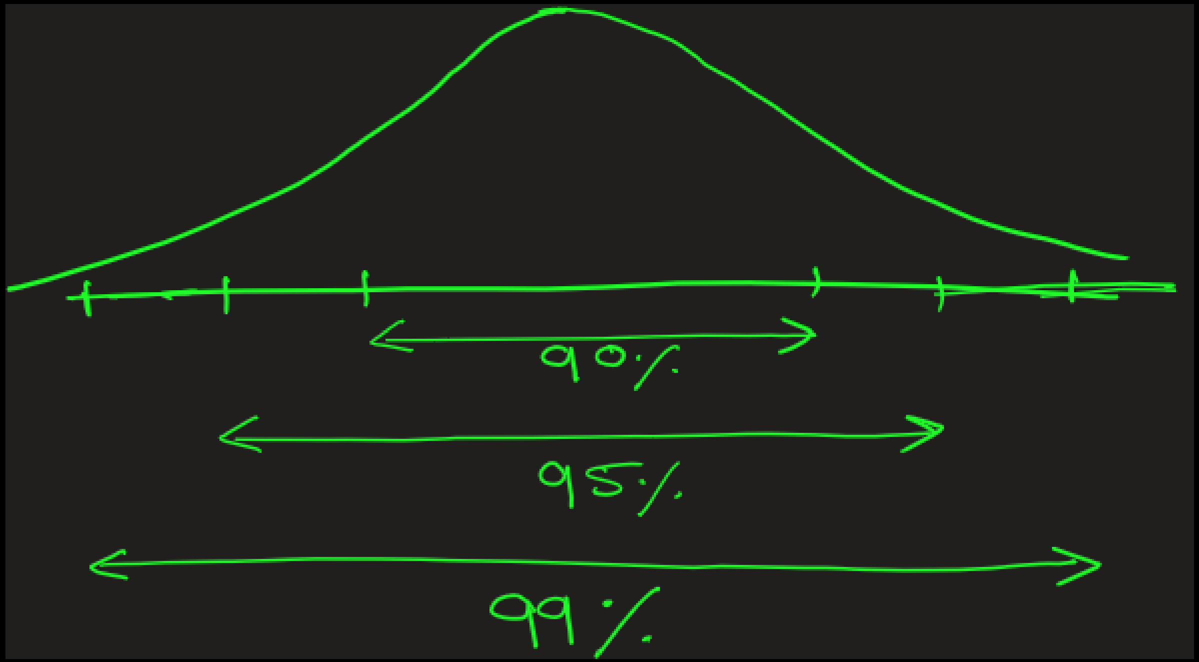
$$\bar{x} - M.E < \mu < \bar{x} + ME$$



Confidence Interval

Cheat Sheet: Common Confidence Levels, their z-values and their meaning

Confidence Level	Z-Value	Meaning
90%	1.645	90% confident true mean lies in interval; narrow interval
95%	**1.96**	**Most commonly used**
99%	2.58	Very high confidence, but wider interval



Confidence Interval: Application

$$\bar{x} - z\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z\cdot\frac{\sigma}{\sqrt{n}}$$

Problem: A factory wants to find average weight μ of a packaged product. You take a random sample of $n = 100$ packages and find the **sample mean** $x = 50$ grams. The **population** standard deviation is known to be $\sigma = 8$ grams.

Find a **95% confidence interval** for the true mean weight.

Ans:

Given: $n = 100$, $\bar{x} = 50$ g, $\sigma = 8$ g, Confidence level = 95%

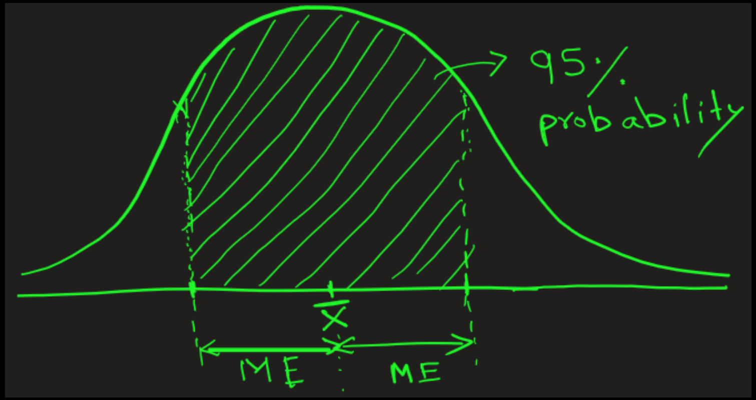
1. Since we know pop. std dev. σ , we use z-distribution. Critical z value for 95%: $z = 1.96$

2. Compute the margin of error (ME): $z\left(\frac{\sigma}{\sqrt{n}}\right)$ = 1.96 x (8/$\sqrt{100}$) = 1.568

3. Construct the confidence interval:

   50 – 1.568 < μ < 50 + 1.568

 => 48.432 g < μ < 51.568 g



Interpretation: “We are 95% confident that the true mean weight μ of the packaged product lies between **48.43 g** and **51.57 g**.”
(Meaning: the procedure used to create this interval will capture the true mean in about 95% of repeated random samples of size 100.)

Confidence Interval: Application

$\bar{x} - z\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z\cdot\frac{\sigma}{\sqrt{n}}$

Problem: Let's redo previous problem but smaller sample size n = 49
A factory wants to find average weight μ of a packaged product. You take a random sample of $n = 49$ packages and find the **sample mean** $\bar{x} = 50$ grams. The **population** standard deviation is known to be $\sigma = 8$ grams.
Find a **95% confidence interval** for the true mean weight.

Ans:

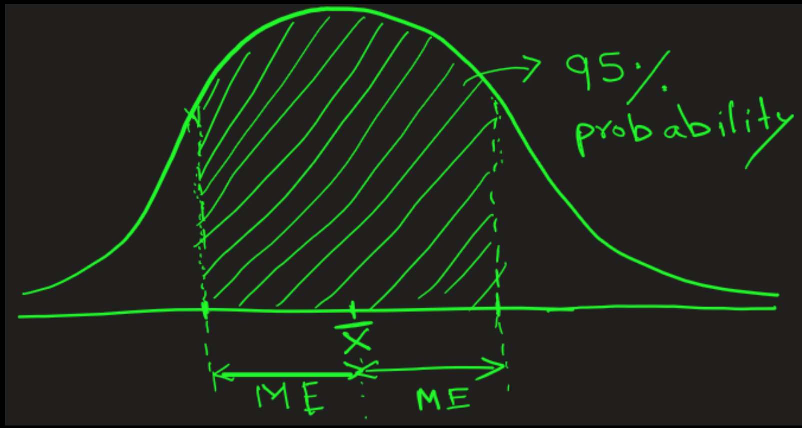
Given: $n = 25$, $\bar{x} = 50$ g, $\sigma = 8$ g, Confidence level = 95%

1. Since we know pop. std dev. σ , we use z-distribution.
   Critical z value for 95%: $z = 1.96$

2. Compute the margin of error (ME): $z\left(\frac{\sigma}{\sqrt{n}}\right)$ = 1.96 x (8/$\sqrt{49}$) = **2.24**

(Note: ME became larger now)

3. Construct the confidence interval:
 50 – 2.24 < μ < 50 + 2.24
 => 47.76 g < μ < 52.24 g



Interpretation: "We are 95% confident that the true mean weight μ of the packaged product lies between **47.76g** and **52.24g**."

For n = 100, the CI was [**48.43, 51.57**]
For n = 49,   the CI was [**47.76**, **52.24**]
Conclusion: As the sample size decreases, the interval becomes wider, so the estimate is less precise.

Confidence Interval: Application

$$\bar{x} - t\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\frac{s}{\sqrt{n}}$$

Problem: Let's redo previous problem but smaller sample size n = 25

A factory wants to find average weight μ of a packaged product. You take a random sample of $n = 25$ packages and find the **sample mean** $\bar{x} = 50$ grams. The **sample** standard deviation was calculated to be $s = 8$ grams.

Find a **95% confidence interval** for the true mean weight.

Ans:

Given: $n = 25$, $\bar{x} = 50$ g, $s = 8$ g, Confidence level = 95%

1. Since we know sample. std dev. s , we use t-distribution. Here d.f. is 25-1=24.

   Using t-table, we find critical t- value for 95%: $\mathrm{t} = 2.064$

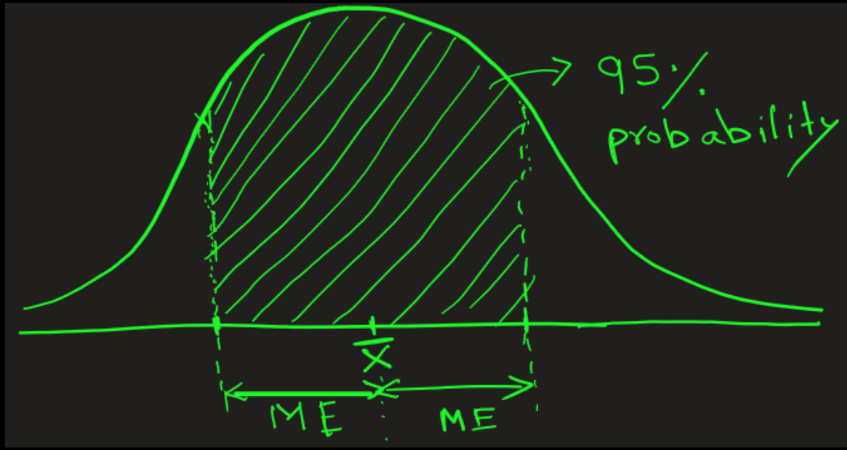
2. Compute the margin of error (ME): $t\left(\frac{s}{\sqrt{n}}\right)$ = 2.064 x (8/$\sqrt{25}$) = 3.3024

   (Note: ME became larger now)

3. Construct the confidence interval:

   50 – 3.30 < μ < 50 + 3.30

   => 46.70 g < μ < 53.30 g



Interpretation: "We are 95% confident that the true mean weight μ of the packaged product lies between **46.70g** and **53.30g**."

Confidence Interval: Why Do We Need CI?

$$\bar{x} - z \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z \cdot \frac{\sigma}{\sqrt{n}}$$

1) Point estimates alone are incomplete

- A sample mean (e.g., 3.5 hours) doesn’t tell us how accurate it is.
- CI gives a range that quantifies the uncertainty.

2) Accounts for sampling variability

- Every sample gives a slightly different result.
- CI shows how much variation to expect.

3) Helps in decision-making

Example: If a new medicine’s 95% CI for improvement = [1.2, 2.8], and it doesn’t include 0, we can infer the drug likely has a real effect.

4) Interpretable in real-world terms

Easier to explain uncertainty than p-values.

EXTRA

Ffnbksdfd
Fdslj;f

Fl;ds