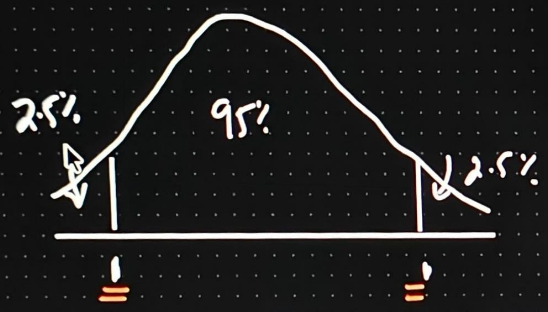
# Confidence Intervals and Margin of Error




C.I = 95%

2

4

## Point Estimate

$$\boxed{\bar{x}} \longrightarrow \boxed{\mu}$$

$\bar{x} = 2.5$ $\mu = 3$

# Confidence Interval

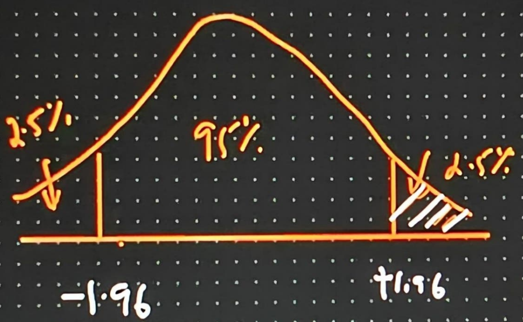
Point Estimate ± Margin of Error

Z test $\Rightarrow$ $$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Eg: On the verbal section of CAT exam, the standard deviation is known to be 100. A sample of 30 test takers has a mean of 520. Construct 95% C.I about the mean.

$$\bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$\alpha = 0.05$




$1 - 0.025 = 0.9750$

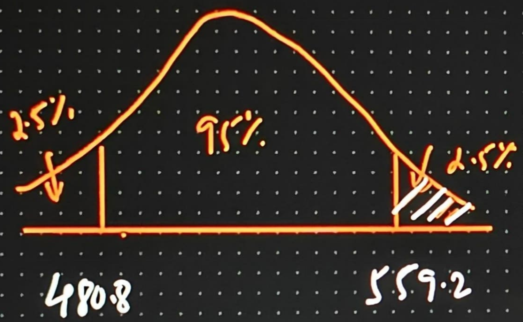
$\pm 1.96 \Rightarrow$ Z table

$$\text{Lower C.I} = 520 - (1.96)\frac{100}{\sqrt{25}} = 480.8$$

$$\text{Higher C.I} = 520 + (1.96)\frac{100}{\sqrt{25}} = 559.2$$

$$x + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$\alpha = 0.05$




$1 - 0.025 = 0.9750$

$+1.96 \Rightarrow$ Z table

$$\text{Lower C.I} = 520 - (1.96)\frac{100}{\sqrt{25}} = 480.8$$

$$\text{Higher C.I} = 520 + (1.96)\frac{100}{\sqrt{25}} = 559.2$$

Lower $C.I = 520 - (1.96)\dfrac{100}{\sqrt{25}} = 480.8$

Higher $C.I = 520 + (1.96)\dfrac{100}{\sqrt{25}} = 559.2$

Conc : I am 95% confident about the mean CAT score is

between 480.8 and 559.2

## 1. Confidence Interval for the Mean of Normally Distributed Data

When we want to find the mean of a population based on a sample we use this method.

- If the sample size is small (less than 30) we use the T-distribution .
- If the sample size is large (more than 30) then we use the Z-distribution.

## 2. Confidence Interval for Proportions

This type is used when estimating population proportions like the percentage of people who like a product. Here we use the sample proportion, the standard error and the critical Z-value to calculate the interval. It gives us the idea where the real value could fall based on sample data.

## 3. Confidence Interval for Non-Normally Distributed Data

If your data isn't normally distributed (doesn't follow a bell curve), use bootstrap methods:

- Randomly resample the data many times
- Calculate intervals from these resamples

## For Calculating Confidence Interval

To calculate a confidence interval you need two key statistics:

- **Mean ($\mu$)**: Average of all sample values
- **Standard Deviation ($\sigma$)**: Shows how much values vary from the mean

Once you have these you can calculate the confidence interval either using t-distribution or z-distribution depend on the sample size whether the population standard deviation is known.

### A) Using t-distribution

**Used when:**

- Sample size is small
- Population standard deviation is unknown

**Example:**

*Sample size = 10*
*Mean weight = 240 kg*
*Std deviation = 25 kg*
*Confidence Level = 95%*

**Step-by-Step Process:**

**Step-by-Step Process:**

- **Degrees of Freedom (df): For t-distribution we first calculate the** degrees of freedom: $df = n - 1 = 10 - 1 = 9$
- **Significance Level (α): The confidence level (CL) is 95%** so the significance level is: $\alpha = \frac{1-CL}{2} = \frac{1-0.95}{2} = 0.025$
- **Find t-value from t-distribution table: From the t-table** for df = 9 and α = 0.025 **the t-value is 2.262 which can be find** using the below table.

| (df)/(α) | 0.1 | 0.05 | 0.025 | .. |
|---|---|---|---|---|
| ∞ | 1.282 | 1.645 | 1.960 | .. |
| 1 | 3.078 | 6.314 | 12.706 | .. |
| 2 | 1.886 | 2.920 | 4.303 | .. |
| : | : | : | : | .. |
| 8 | 1.397 | 1.860 | 2.306 | .. |
| 9 | 1.383 | 1.833 | 2.262 | .. |

- **Apply t-value in the formula:** The formula for the confidence interval is: $\mu \pm t\left(\frac{\sigma}{\sqrt{n}}\right)$ Using the values: $240 \pm 2.262 \times \left(\frac{25}{\sqrt{10}}\right)$
- The confidence interval becomes: $(222.117, 257.883)$

Therefore we are 95% confident that the true mean weight of UFC fighters is between 222.117 kg and 257.883 kg.

This can be calculated using Python's scipy and math library to find the t-value and perform the necessary calculations. The stats module provides various statistical functions, probability distributions, and statistical tests.

```
import scipy.stats as stats
import math

mean = 240
std = 25
n = 10
df = n - 1
alpha = 0.025
t = stats.t.ppf(1 - alpha, df)

moe = t * (std / math.sqrt(n))

lower = mean - moe
upper = mean + moe

print(f"Confidence Interval: ({lower:.2f}, {upper:.2f})")
```

**Output:**

**Output:**

*Confidence Interval: (222.1160773511857, 257.8839226488143)*

## B) Using Z-distribution

**Used when:**

- Sample size is large
- Population standard deviation is known

Consider the following example. A random sample of 50 adult females was taken and their RBC count is measured. The sample mean is 4.63 and the standard deviation of RBC count is 0.54. Construct a 95% confidence interval estimate for the true mean RBC count in adult females.

**Step-by-Step Process:**

1. Find the mean and standard deviation given in the problem.
2. Find the z-value for the confidence level: For a 95% confidence interval the z-value is 1.960.
3. Apply z-value in the formula: $\mu \pm z\left(\frac{\sigma}{\sqrt{n}}\right)$

Using the values: some common values in the table given below:

1. Find the mean and standard deviation given in the problem.
2. Find the z-value for the confidence level: For a 95% confidence interval the z-value is 1.960.
3. Apply z-value in the formula: $\mu \pm z\left(\frac{\sigma}{\sqrt{n}}\right)$

Using the values: some common values in the table given below:

| Confidence Interval | z-value |
|---|---|
| 90% | 1.645 |
| 95% | 1.960 |
| 99% | 2.576 |

The confidence interval becomes: (4.480,4.780)

Therefore we are 95% confident that the true mean RBC count for adult females is between 4.480 and 4.780.

Now let's do the implementation of it using Python. But before its implementation we should have some basic knowledge about numpy and scipy.

## Given Data

- Sample size $n = 50$
- Sample mean $\bar{x} = 4.63$
- Standard deviation $\sigma = 0.54$
- Confidence level = 95%
- Z-value for 95% = 1.96

**Since:**

- Sample size is large
- Standard deviation is known

➡ We use Z-distribution.

---

## Formula

$$\text{Confidence Interval} = \bar{x} \pm Z\left(\frac{\sigma}{\sqrt{n}}\right)$$

## Formula

$$\text{Confidence Interval} = \bar{x} \pm Z\left(\frac{\sigma}{\sqrt{n}}\right)$$

## Step 1 – Calculate Standard Error

$$\sqrt{50} \approx 7.07$$

$$\frac{\sigma}{\sqrt{n}} = \frac{0.54}{7.07} \approx 0.0764$$

## Step 2 – Margin of Error

$$ME = 1.96 \times 0.0764 \approx 0.1497$$

## Step 3 – Confidence Interval

### Lower Limit

$$4.63 - 0.1497 \approx 4.48$$

### Upper Limit

$$4.63 + 0.1497 \approx 4.78$$

---

## Final Answer

$$4.48 \le \mu \le 4.78$$

---

## Interpretation

We are 95% confident that the true mean RBC count of adult females lies between **4.48 and 4.78**.