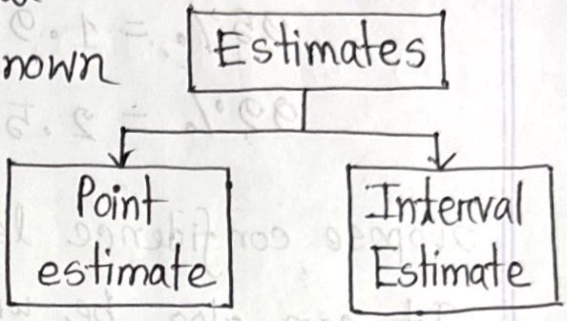
Parameter Estimation

- **Estimation:** Estimation is specific numerical value used to estimate an unknown population.

① **Point estimate:** Single numerical value used to estimate an unknown population parameter.



⇒ Sample mean is a point estimate of a population mean.

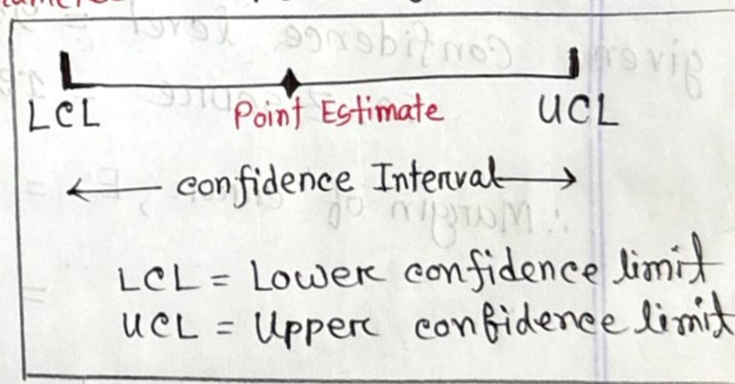
② **Interval Estimate:** Range of values used to estimate an unknown population parameters.

⇒ Interval estimates of population parameters are called confidence intervals (CI)

- **Unbiasedness:** $\boxed{E(\hat{\theta}) = \theta}$ → Parameter → property of an unknown probability distribution

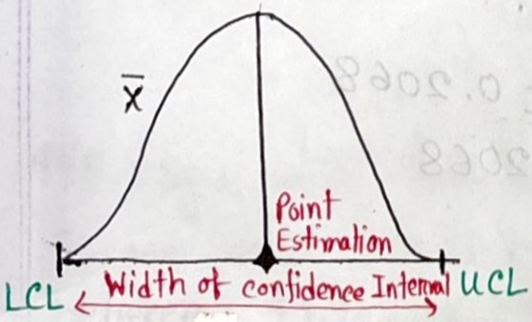
⇒ Sample mean is unbiased estimator of μ.

⇒ The sample variance is an unbiased estimator of σ^2



- **Confidence Intervals:** $\boxed{\bar{x} \pm z\left(\frac{\sigma}{\sqrt{n}}\right)}$

Point estimate ± margin of error



$1 - \alpha = CI$

Common Z scores of Confidence levels :-

$90\% = 1.645$

$95\% = 1.96$

$99\% = 2.576$

Suppose confidence level is $= 95\%$

It can also be written as, $(1-\alpha) = 0.95$

☐ A sample of 11 newly born baby, from a large normal population, has a mean height of 2.20 feet. We know from past testing that the population standard deviation is 0.35 foot. Determine 95% confidence interval for the true mean height of the newly born baby.

given, Confidence level $= 95\%$

so, Z score $= 1.96$

$\therefore$ Margin of error, $E = z\left(\frac{\sigma}{\sqrt{n}}\right)$

$= 1.96 \times \left(\frac{0.35}{\sqrt{11}}\right)$

$\therefore E = 0.2068$

$\therefore$ Confidence Interval $= \bar{X} \pm E$

$= 2.20 \pm 0.2068$

$2.20 - 0.2068 < \mu < 2.20 + 0.2068$

$\therefore 1.9932 < \mu < 2.4068$

So, we are 95% confident that the true mean is between 1.9932 to 2.4068 feet

Though the true mean may or may not be in this interval, 95% of intervals formed in this manner will contain the true mean.

☐ A random sample of $n = 50$ males showed a mean average daily intake of dairy products equal to 756g with a standard deviation of 35g. Find a 95% and 99% confidence interval for the population average?

for 95%, When confidence level is 90%, z-score = 1.96

$$CI = \bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

$$= 756 \pm 1.96 \frac{35}{\sqrt{50}} =$$

$$746.3g < \mu < 765.70g$$

So, we are 95% confident that the true mean is between 746.3g to 765.70g

for 99%, z-score = 2.576

$$CI = \bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

$$= 756 \pm 2.576 \frac{35}{\sqrt{50}}$$

$$443.25g < \mu < 468.75g$$

So, we are 99% confident that, the true mean is between 443.25g to 468.75g

Student's T Distribution

It is a probability distribution that is used to calculate population parameters when than sample size is small and when the population variance (σ) is unknown.

population std

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

Sample std

t-value depends on df

$df = n-1$

- t-test type.
  1. One sample t-test
  2. Independent Two sample t-test
  3. Paired t-test

- A random sample of $n = 25$ has $\bar{x} = 50$ and $s = 8$. Form a 95% confidence interval for μ.

given $n = 25$

$df = 25 - 1 = 24$

We know, $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

$\Rightarrow t \frac{s}{\sqrt{n}} = \bar{x} - \mu$

$\therefore \mu = \bar{x} - \left(t \frac{s}{\sqrt{n}}\right)$

$\alpha = 1 - 0.95$
$= 0.05$

When, 95% CI

$t_{n-1, \frac{\alpha}{2}} = t_{24, \frac{0.05}{2}}$

$\therefore t_{24, 0.025} = 2.064$

The confidence interval,

$$\bar{x} - \left(t \frac{s}{\sqrt{n}}\right) < \mu < \bar{x} + \left(t \frac{s}{\sqrt{n}}\right)$$

$$\Rightarrow 50 - \left(2.064 \times \frac{8}{\sqrt{25}}\right) < \mu < 50 + \left(2.064 \times \frac{8}{\sqrt{25}}\right)$$

$$\Rightarrow 46.6976 < \mu < 53.3024$$

☐ Inferential Statistics

- CI
- Test of Hypothesis
  - Mean Test
    - Single Mean Test → z/t test
    - two Means Test → t test (paired/Independent)
    - More than two means → Analysis of variance (ANOVA) → F test

$\left\{\begin{array}{l} z \text{ depends on } \alpha \\ t \text{ depends on } \alpha, df \end{array}\right\}$

☐ Error in making decisions

① Type-I Error: Innocent → Guilty

$\left(1-\alpha = CI\right.$ সত্যকে সত্য বলা$\left.\right)$

- Reject a true null hypothesis (H_0)
- It is considered as a serious type of error
- Probability of type-1 error is α.
- It is called level of significance of the test.

② Type-II Error: Guilty → Innocent

- Fail to reject a false null hypothesis
- Probability of type-II error is β.

$\left(1-\beta = \text{Power of test}\right.$ মিথ্যাকে মিথ্যা বলা$\left.\right)$

□ One mean test:

- population std (σ) দেওয়া থাকলে z-test. নাহলে t-test
- Null hypothesis কখনোই "$\neq$" নিতে পারবে না

  only $=, \geqslant, \leqslant$

□ Test the claim that the true mean no. of children in BD homes is equal to 3. Assume, $\sigma = 0.8$ sample size, $n = 100$.

two tail test

Process:

- Step-1: Draw the hypothesis

$\Rightarrow$ $H_0 : \mu = 3$

$H_1 : \mu \neq 3$

- Step-2: draw the picture and fix the tail

$\Rightarrow$ $H_1 : \mu \neq 3$

* if α is not given $\alpha = 5\%$

- Step-3: Fix the test

$\Rightarrow$ Here σ is given and it is a single mean test. So, we have to perform z-test.

here, $\mu_0 = 3$

$$\therefore z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{2.84 - 3}{\frac{0.8}{\sqrt{100}}} = -2$$

Two tail Test

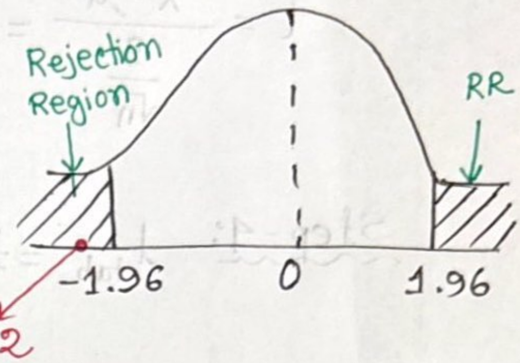
- **Step-4: Find Z_{tab}**

  Z depends on α

  $\alpha = 5\%$

  for two tail $= \frac{\alpha}{2} = 0.025$

  $\therefore$ Probability $= -1.96$



- **Step-5: Decision**

  Decision: Here Z_{cal} falls into the rejection region of H_0. So, we may reject the null hypothesis.

- **Step-6:**

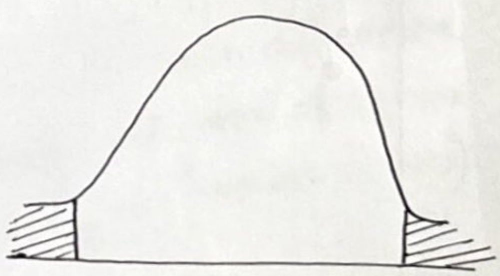
  Conclusion: True mean number of children in BD home is not equal to 3.

□ The average cost of a blood test in a hospital is said to be Tk.168. A random sample of 25 patients resulted in $\bar{x} =$ Tk172.50 and
$s =$ Tk15.40
$\alpha = 0.05$ level

Step-1: $H_0 : \mu = 168$
$H_a : \mu \neq 168$

Step-2: $\alpha = 0.05$

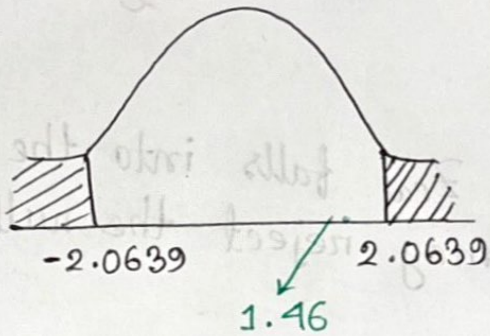
$\frac{\alpha}{2} = 0.025$



Step-3: Here σ is not given, So we have to perform

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{172.50 - 168}{\frac{15.40}{\sqrt{25}}} = 1.46$$

Step-4: $t_{tab} = t_{df, \frac{\alpha}{2}} = t_{24, 0.025} = 2.0639$



Step-5: Decision: Here, t_{tab} falls into the excepted region on H_0. So we may not reject H_0.

Step-6: Conclusion: We have sufficient evidence to prove that, statistically and significantly, the average cost of a blood test in a hospital is 168Tk at 5% confidence level.

Hypothesis Testing

- A hypothesis is a claim (assumption) about a population parameter.

There are 2 types of Hypothesis

(i) Null Hypothesis (H_0)
- → refers to the status Quo
- → always contain '=', $\geq$, $\leq$
- → may or may not be rejected

(ii) Alternative Hypothesis (H_1 / H_a)
- → challenges status Quo.
- → opposite of H_0
- → the hypothesis that researcher wants to prove

Level of significance (α)

We know,

$$CI = 1 - \alpha$$

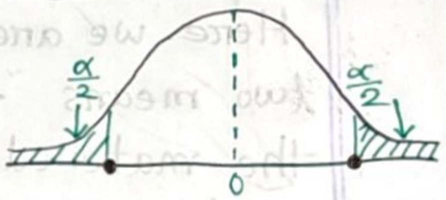
$$\therefore \alpha = 1 - CI$$

- It Defines the rejection region of sampling distribution
- It also provides the critical value of the test.

Two tail test

$H_0: \mu = 3$

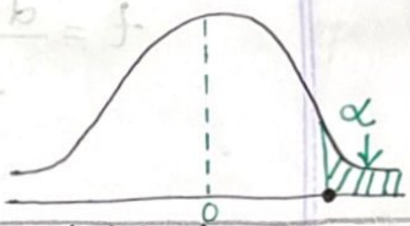
$H_1: \mu \neq 3$



Upper tail test

$H_0: \mu \leq 3$

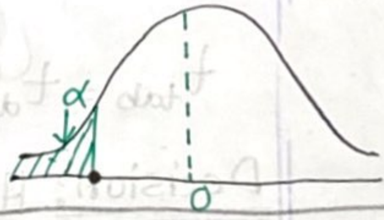
$H_1: \mu > 3$



Lower tail test

$H_0: \mu \geq 3$

$H_1: \mu < 3$



— × —

Matched Pairs : Test Statistic

It compares the mean of two different groups (example - before & after treatment in the same group).

$$\boxed{\therefore t = \frac{\bar{d} - D_0}{\frac{S_d}{\sqrt{n}}}}$$

D_0 = hypothized mean difference

S_d = Sample std dev. of difference

n = Sample size (number of pairs)

Matched Pair

□ Assume you send your Doctors to a training workshop. Has the training made a difference in the number of complaints. You collect the following data:

Doctors	Number of Complaint		Difference (di)
	Before (1)	After (2)	(2) - (1)
DMC	6	4	-2
Mitford	20	6	-14
SMC	3	2	-1
RMC	0	0	0
CMC	4	0	-4
			$\Sigma_{diff} = -21$

$$H_0 : \overbrace{\mu_x - \mu_y}^{D_0} = 0$$

$$H_1 : \mu_x - \mu_y \neq 0$$

$$\alpha = 0.01$$

$$\bar{d} = \frac{\Sigma d_i}{n} = \frac{-21}{5} = -4.2$$

Here we are performing two means test and samples came from the matched population. So, we have to perform matched / paired test.

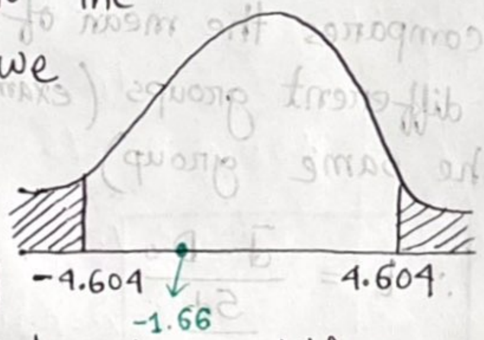
$$S_d = \sqrt{\frac{\Sigma(d_i - \bar{d})^2}{n-1}} = 5.67$$

$$t = \frac{\bar{d} - D_0}{\frac{S_d}{\sqrt{n}}} = \frac{-4.2 - 0}{\frac{5.67}{\sqrt{5}}}$$

$$\therefore t_{cal} = -1.65 \approx -1.66$$

$$t_{tab} = t_{df, \frac{\alpha}{2}} = t_{4, 0.005} = 4.604$$

Decision: Here t_{cal} falls into the accepted region of H_0, So we may not reject H_0.



Conclusion: We have sufficient evidence to prove that statistically & significantly, the training has not made a difference in the number of complaints at 1% significance level.

Pooled Variance t-test

Is there a difference in the ICU patients dying in DMC and SMC per day? Assuming both populations are approximately normal with equal variance, is there a difference in average number of dying ($\alpha = 0.05$)?

	DMC	SMC
Number	21	25
average	3.27	2.53
dying sample std dev	1.30	1.16

$$H_0 : \overbrace{\mu_D - \mu_S}^{D_0} = 0$$

$$H_1 : \mu_D - \mu_S \neq 0$$

Here, we have two means and sample came from the independent population. So, we have to perform independent t-test.

$$\therefore t = \frac{(\bar{x}_1 - \bar{x}_2) - \overbrace{(\mu_1 - \mu_2)}^{D_0}}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 2.040 = t_{cal}$$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)(n_2-1)}$$

$$= \frac{(21-1)(1.30)^2 + (25-1)(1.16)^2}{(21-1)+(25-1)}$$

$$= 1.5021$$

$$t_{cal} = \frac{(3.27 - 2.53) - 0}{\sqrt{1.5021\left(\frac{1}{21} + \frac{1}{25}\right)}} = 2.040$$

$$\therefore t_{tab} = t_{df, \frac{\alpha}{2}} = t_{(21+25-2), 0.025} = t_{44, 0.025}$$

$\therefore t_{tab} =$