CO-3

* **Session-15:**

Testing of Hypothesis

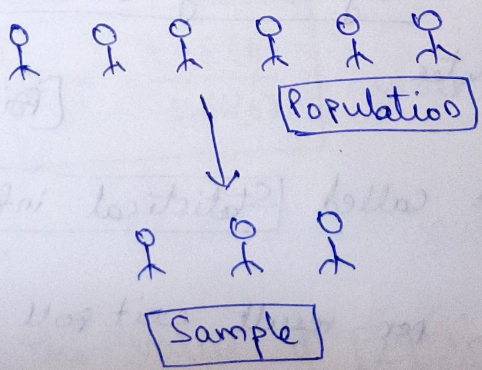
→ **Population and Sample:**

A population is the entire group that you want to draw conclusions about.

- The population refers to the entire grp of individuals items, or events that occurs.

→ **Sample:** Taking a set of things from population is called sample.

Ex:



- Sample is subset of Population.

→ $\bar{x}$ = Sample mean

μ = population Mean

- The formula for $\boxed{\text{Sample mean}}$ is :

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

n = size of sample

- The formula for $\boxed{\text{Population}}$ is :

$$\boxed{\mu = \frac{\sum_{i=1}^{N} x_i}{N}}$$

N = size of entire (population)

→ Testing of Hypothesis:

A statistical ~~ap~~ hypothesis is defined as a statement, which may or may not be true about population parameter.

[For more Refer PPT Theory]

- This process is called $\boxed{\text{Statistical inference}}$.

Ex: Loksabha per result exit poll results by Media

- Steps:

① Formulate Hypotheses → null hypothesis (H0) → alternative Hyp (H1 or Ha)

② Select a significance level (α) → α = 0.05 (or) 0.01

③ choose a statistical Test → t-test, chi, ----

→ **Procedure for Testing of Hypothesis:**

1. Null Hypothesis : set up the Null hypothesis H_0

2. Alternative Hypothesis: set up to H_1, this will enables us to decide whether we have to use a single-tailed (right or left) test (or) two-tailed test.

↳ $\boxed{\mu \neq \mu_0}$

↳ $\boxed{\mu < \mu_0}$ $\boxed{\mu > \mu_0}$

③ level of Significance (α): choose the appropriate α depending on the reliability of the estimates & permissible risk. This is to be deside before the sample drawn ie(α fixed advanced)

→ 0.01 (or) 0.05

④ Test statistic: Compute the test statistic

$$\boxed{Z = \frac{t - E(t)}{S.E.(t)}}$$ under H_0

⑤ Conclusion:

Compare the value of $\boxed{Z}$ in step 4 with the significant value $\boxed{Z_\alpha}$ at the given level of significance 'α'.

→ **Inferential Procedures:**

① Estimation

② Hypothesis testing

- **Null Hypothesis:**

$\boxed{H_0 : \mu = \mu_0}$

[Definition PPT]

- **Alternative Hypothesis:**

Complementary to the Null Hyp--

If $H_0 : \mu = \mu_0$ then A.H should be:

i) $H_1 : \mu \neq \mu_0$ ($\mu > \mu_0$ (or) $\mu < \mu_0$) ↳ $\boxed{\text{Two-tailed}}$

ii) $H_1 : \mu > \mu_0$ → $\boxed{\text{right}}$ tailed

iii) $H_1 : \mu < \mu_0$ → $\boxed{\text{left}}$ tailed

* **One-Tailed Test Formulas:** → Increases / decreases.
If seen in Qn use this.

1. $\boxed{\text{Z-Test:}}$ (Large Sample Size)

$$\boxed{Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}}$$

Z: Test statistic

• Critical Value: look up the CV from the standard normal dist table.

2. T-Test: (small sample size)

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

- Critical Value: look up the t-table,

$df = n - 1.$

$ two-tailed test formulas:

1. Z-Test: (large)

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

2. t-Test: (small)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

n_1, n_2 are values of sample ~~space~~ size

Ex: n_1 = section B value - - -

- $df = n_1 + n_2 - 2$

Problem 1: One-Tailed Test

A Company claims that their new training program increases employee productivity by at least 20 units per hour. To test this claim, a sample of 36 Employees undergoes the training, & their productivity levels are measured. The sample mean productivity increases is 22 unit per hour with a population s deviation of 5 units/hr.

Perform a one-tailed hypothesis test at a significance level of 0.05 to determine if there is evidence to support the Company claim.

Sol: 1. Formula Hypotheses:

i) Null hypothesis (HO): $\mu = 20$ (No significant Increase in productivity)

ii) Alternative Hypothesis (H_a): $\mu > 20$ (Productivity Increased by more than 20)

2. Calculate Test statistic (z-Test):

Given,

Sample Mean ($\bar{x}$) = 22

population standard Deviation (σ) = 5

Sample Size (n) = 36

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$z = \frac{22 - 20}{5 / \sqrt{36}} = 2.4$$

3. Deteemine critical value:

for one-tailed test at a significance level of 0.05, the cv is approx 1.645.

4. Make a Decision:

z-value greatee than critical value, (2.4>1.645)

we reject null hypothesis.

5. Draw Conclusion:

There is a significant evidence to support the claim that the new training program increases employees.

→ Level of Significance: The LOS, denoted by

→ Standard Error:

$$\boxed{S.E_m = \frac{\sigma}{\sqrt{n}-1}}$$

* Session-16: [concept]

Two types of Errors, Level of Confidence.

→ **Decision Errors**: Refer to the incorrect conclusions that can be made based on the results of a statistical test.

* **Type-1 Error** [WRONGLY REJECTED] (or) False Positive

A Type 1 error occurs when the null Hypothesis (H0) is incorrectly rejected when it is actually true.

- Probability of making Type 1 error Denoted by the significance level $\boxed{\alpha}$.
  ↳ Producer's risk

If $\alpha = 0.05$ then,

5% chance of making Type 1 error.

Example: Concluding that a new drug is effective in treating a diease when atcually has no effect.

Yes → NO

→ Type 2 Error [WRONGLY ACCEPTED]

(or) False Negative

A Type II error occurs when the null hypo (H0) is incorrectly failed to be rejected when it is actually false.

- Prop of Making a Type II error is denoted by $\boxed{\beta}$ → Consumer's risk.

Example: Failing to conclude that a new treatment is effective in treating a disease when it actually has a effect.

No → Yes

→ Other points:

→ lowering the α decreases the likehood of Type 1 errors & vice versa.

Type-1

→ The error of rejecting H_0 (accepting H1) when H_0 is true is called Type I error.

α = prob of T 1 error = Prob of rejecting Ho when H_0 is true

α = P (we are rej Ho when it is true) = (P (reject H_0 | H_0 is true))

Type-2

- Accepting null Hyp (H0), when alternative hyp (H1) is true.
- Also Named as Acceptance error.
- Type II is Dangerous than Type I.
- The error of accepting H_0 (rejecting H_1) when H_1 is true is called Type II error.

β = Prob of Type II error.

= Prob of Accepting H_0 when H_1 is true

= P (Accept H_0 | H_1 is true)

* Level of Confidence:

The confidence with which an experiment rejects a null hypothesis depends upon the level of significance adopted. These may, hence, sometime be termed as LOC.

Level	Amount of Confidence	Interpretation
0.05	95 %	If the experiment is conducted 100 times only 5 occasions the obtained mean fall out side $\mu \pm 1.96SE$
0.01	99 %	$\mu \pm 2.58SE$

* Level of Significance:

The LOS of significance is that prob of chance occurence of observed result up to and below which the prob 'p' of the null hypothesis being correct is considered too ~~small~~ low & the results of the experiment are consired sign.

- LOS depends on our choice (or) α value

$\boxed{0.05 \text{ or } 0.01} \xrightarrow[\text{Percentage}]{\text{In}} 5\% \text{ or } 1\%$

- $P(x \in w | H_0) = \alpha = $ LOS of the test

$P(x \in w | H_1) = \boxed{1-\beta}$ is called the $\boxed{\text{power of test}}$

$P(x \in w | H_0) = 1-\alpha$

$P(x \in w | H_1) = \beta$

→ Example 1: let P be the probability that a coin will fall head in a single toss in order to test $H_0: P = 1/2$ against $H_1: P = 3/4$. the coin is tossed 5 times & H_0 is rejected if more than 3 heads are obtained. obtain the prob of type I & type II error & power of the test

Sol: i) type 1 Error

$P(X=x) = {}^nC_x \, p^x \, q^{n-x}$

Given, $p = \frac{1}{2}$

$q = 1-p = 1-\frac{1}{2} = 1/2$

$n=5$

$\alpha = P(H_0 \text{ is rejected})$ ---> Given greater than 3 rejected so, 4 & 5

$= P(X=4) + P(X=5)$

$= {}^5C_4 \, p^4 q^{5-4} + {}^5C_5 \, p^5 q^{5-5}$

$= {}^5C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^1 + {}^5C_5 (1/2)^5 \left(\frac{1}{2}\right)^0$

$= 3/16$

ii) power of test

$p = \frac{3}{4}$

$q = 1-\frac{3}{4}$

$q = 1/4$

$\beta = P(H_0 \text{ is Accepted})$

$= P(X=0) + P(X=1) + P(X=2)$

$= {}^5C_0 \, p^0 q^{n-0} + {}^5C_1 \, p^1 q^{n-1} + {}^5C_2 \, p^2 q^{n-2} + {}^5C_3 \, p^3 q^{n-3}$

$= 1 \cdot \left(\frac{3}{4}\right)^0 \left(\frac{1}{4}\right)^5 + 5\left(\frac{3}{4}\right)^1 \left(\frac{1}{4}\right)^4 + 10\left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^3 + 10\left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right)^2$

$= \frac{47}{128}$

$POT = 1-\beta = 1 - \frac{47}{128}$

$= \frac{81}{128}$

* Session-17:

Test of Significance for Single Mean

T and Z test

t-Test for Single Mean:

step-1: Define Hypotheses

step 2: Choose a Significance level α.

0.05 or 0.01.

step 3: Calculate the t-statistic

$$\boxed{t = \frac{(\bar{x} - \mu)}{s / \sqrt{n}}}$$

μ = population mean

s = S. Deviation

n = sample size

$\bar{x}$ = sample mean.

step 4: Determine the Critical Value or p-value.

- Compare the calculated t-value to the critical t-value

step-5: Make a conclusion

→ z-Test for Single Mean:

Same as t-test Process

$$\boxed{z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}}$$

σ = population standard deviation

→ **Note:**

- $\boxed{n < 30}$ → t-test
- $\boxed{n > 30}$ → z-test
- test statistic, $\boxed{t = \dfrac{\bar{x} - \mu}{s / \sqrt{n}}}$ s = s. deviation

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

→ If calculated $\boxed{|t| > \text{tabulated } t}$, null hypothesis is rejected.

→ If calculated $\boxed{|t| < \text{tabulated } t}$, Ho may be accepted at the LOS adopted.

Confidence Intervals:

for t-test:

i) 95% CI for population mean μ are: $\boxed{\bar{x} \pm t_{0.05} \, s/\sqrt{n}}$

ii) 99% " : $\boxed{\bar{x} \pm t_{0.01} \, s/\sqrt{n}}$

for z-test:

iii) 95% " : $\boxed{\bar{x} \pm 1.96 \, (\sigma/\sqrt{n})}$

iv) 99% " : $\boxed{\bar{x} \pm 2.56 \, (\sigma/\sqrt{n})}$

} large (n > 30)

Problem-1: A manufacturer claims that the mean weight of their cereal boxes is 500 grams. A random sample of 25 cereal boxes is selected, and their weights are measured. The sample mean is 495 grams and the sample standard deviation is 10 grams. Test whether there is enough evidence to reject the manufacturer's claim at a significance level of 0.05.

Sol:

$\mu = 500$ | $\alpha = 0.05$

$\bar{x} = 495$ | $n = 25$

$s = 10$

$\downarrow$

small sample $\rightarrow$ T-test

Null Hypothesis $H_0 : \mu = 500$

Alternative Hyp $H_1 : \mu \neq 500$ (Two-sided)

Degree of freedom : $n-1 = 25-1$

$= 24$

$$|t|_{cal} = \left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right| = \left|\frac{495-500}{10/\sqrt{25}}\right|$$

$$= \left(\frac{-5}{2}\right) = |-2.5|$$

$$t_{cal} = 2.50$$

By taking into the t table,

$t_{table} = 2.064$

$\therefore t_{cal} = 2.50 > 2.064 = t_{tab}$

$\therefore$ Since $|t| > 2.064$, we reject the Null hypothesis.

Problem - 2:

Suppose a researcher wants to test whether a new teaching method $\boxed{\text{improves}}$ students test scores. A random sample of 20 students is selected, & their scores before & after the teaching method are recorded. The mean increases in score is 8 points, with a s.D of 6 points. Test whether there is enough evidence to conclude that the teaching method is effective at a significance level of 0.01.

Sol: Given,

$n = 20$, $\bar{x} = 8$

$s = 6$, $\alpha = 0.01$

$\mu = 0$

Null Hypothesis $H_0 : \mu = 0$ → Improves in Question

Alternative Hyp $H_1 : \mu > 0$ (ONE-SIDED)

Degrees of freedom : $n-1$

$= 20 - 1 = 19$

level of significance $(\alpha) = 0.01$

$$|t|_{cal} = \left| \frac{\bar{x} - \mu}{s / \sqrt{n}} \right| = \left| \frac{8 - 0}{6 / \sqrt{20}} \right|$$

$$|t|_{cal} = \frac{8}{1} \times \frac{\sqrt{20}}{6} = \underline{5.96}$$

$$t_{table} = \underline{2.539}$$

$\boxed{\begin{array}{l} Df = 19 \\ \alpha = 0.01 \\ \text{(oneside)} \end{array}}$

using these check table Value.

$\therefore t_{cal} = 5.96 > 2.539 = t_{table}$

$\therefore t_{cal} > t_{tab}$

$\therefore H_0$ is Rejected.

* P-Value Approach: The P-value is the prob of

Refer $\boxed{\text{PPT}}$ for

Notes

Example - 1: A machinist is making engine parts with axle diameters of 0.700 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with a standard deviation of 0.040 inch. Compute the test statistic you would use to test whether the work is meeting the specifications. Also state how you would proceed further.

Sol: Given, $n = 10$, $\mu = 0.700$

$\bar{x} = 0.742$, $\alpha = 0.05$

$s = 0.040$

Null Hypothesis $H_0 : \mu = 0.700$

Alternative Hypothesis $(H_1) : \mu \neq 0.700$

(TWO-SIDED)

L.S : $\alpha = 0.01$

Degree of freedom : $n-1 = 10-1 = 9$

$$|t|_{cal} = \left| \frac{\bar{x} - \mu}{s / \sqrt{n-1}} \right| = \left| \frac{0.742 - 0.700}{0.040 / \sqrt{10-1}} \right|$$

$|t|_{cal} = 3.15$

$t_{tab} = 2.262 \leftarrow \boxed{\begin{matrix} Df = 9 \\ \alpha = 0.05 \end{matrix}}$

$\therefore t_{cal} > t_{tab}$

$\therefore$ H_0 is rejected

∴ Product Not meet the specifications.

Example - 2:

A random sample of 10 boys had the following I.Q's: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a population mean I.Q of 100? Find a reasonable range in which most of the mean I.Q values of sample of 10 boys lie.

Sol: Given,

$n = 10$

$\mu = 100$

x	$\bar{x} = \frac{\Sigma x}{n}$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
70	97.2	-27.2	739.84
120	97.2	22.8	519.84
110	97.2	12.8	163.84
101	97.2	3.8	14.44
88	97.2	-9.2	84.64
83	97.2	-14.2	201.64
95	97.2	-2.2	4.84
98	97.2	0.8	0.64
107	97.2	9.8	96.04
100	97.2	2.8	7.84
$\Sigma x = 972$			$\Sigma(x_i - \bar{x})^2 = 1833.60$

$\bar{x} = 97.2$

$n = 10$

$\mu = 100$

$\sigma = 14.27$

N.H (H_0) : $\mu = 100$

A.H (H_1) : $\mu \neq 100$ (two-sided)

$\alpha = 0.05$

$df \Rightarrow n-1 = 10-1 = 9$

$$|t|_{cal} = \left| \frac{\bar{x} - \mu}{s/\sqrt{n}} \right| = \left| \frac{97.2 - 100}{14.27/\sqrt{10}} \right| = 0.62$$

$t_{tab} = 2.262$

$\therefore t_{cal} < t_{tab}$

$\therefore H_0$ is accepted.

The 95% confidence limits within which the mean IQ values of

$\bar{x} \pm t_{0.05}\, s/\sqrt{n}$

$= 97.2 \pm 2.262 \times \frac{14.27}{\sqrt{10}}$

$= 97.2 \pm 10.20$

$= [97.20 - 10.20,\ 97.20 + 10.20]$

$= [87,\ 107.4]$

* Example - 3: A sample of 900 members has a mean 3.4 cms & standard deviation 2.61 cms. Is the sample from a large population of mean 3.25 cm and standard deviation 2.61 cms? If the population is normal & its mean is unknown, find the 95% & 98% fudicial $S = 2.61$, Population Mean $= \mu = 3.25$, $\sigma = 2.61$.

$\boxed{\text{Z-TABLE}}$

Sol: Given,

$n = 900$ $\quad \sigma = 2.61$

$\mu = 3.25$

$\bar{x} = 3.4$

Null Hyp (H_0): $\mu = 3.25$

Alter Hype (H_1) $\mu \neq 3.25$ (TWO SIDED)

There is a difference b/w population & sample mean.

L. significance: $\alpha = 0.05$

$$\boxed{z_{cal} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}} = \left| \frac{3.4 - 3.25}{2.61 / \sqrt{900}} \right| = 1.724$$

$z_{tab} = 1.96$

$\therefore z_{cal} < z_{tab}$

$\therefore H_0$ is Accepted

95% fudicial limits for the population mean μ:

$$\boxed{\bar{x} \pm 1.96(\sigma/\sqrt{n})} \xrightarrow{\text{formula}} = 3.40 \pm 1.96(2.61/\sqrt{900})$$

$= 3.40 \pm 0.1705$

$\Rightarrow [3.5705, 3.2295]$

98% fudicial limits for μ:

$$\boxed{\bar{x} \pm 2.33(\sigma/\sqrt{n})} \xrightarrow{\text{formula}} = 3.40 \pm 2.33(2.61/\sqrt{900})$$

$= 3.40 \pm 0.2027$

$\Rightarrow [3.6027, 3.1973]$

* Terminal Questions:

③ The mean weekly sales of soap bars in departmental store were 146.3 bars per store. After an advertising campaign the mean sales in 22 stores for a typical week increased to 153.7 & showed a standard deviation of 17.2. Was advertising Campaign Successful?

Sol: Given,

$\bar{x} = 146.3$ $\mu = 153.7$

$n = 22$ $\alpha = 0.05$ → our wish value

$s = 17.2$

Null Hyp (H_0): $\mu = 153.7$

A.H (H_1): $\mu > 153.7$ (ONE-SIDED)

$\alpha = 0.01$

D.f : $22 - 1 = 21$

$$|t|_{cal} = \left| \frac{\bar{x} - \mu}{s/\sqrt{n-1}} \right| = \left| \frac{146.3 - 153.7}{17.2/\sqrt{21}} \right|$$

$|t|_{cal} = 2.01$

$t_{tab} = 2.508 \rightarrow$ from T-table.

$\therefore t_{cal} < t_{tab}$

$\therefore H_0$ is Accepted.

* Session-18

T-test & Z-test for DOUBLE MEAN

- $\boxed{\text{T-test}}$ for Double Mean formula:

$$\boxed{t = \frac{|\bar{x}_1 - \bar{x}_2| - |\mu_1 - \mu_2|}{s\sqrt{1/n_1 + 1/n_2}}}$$

(or) σ

(or) $x_1 = x$ first Sample mean

$x_2 = y$ 2nd sample mean

$\mu_1 = \mu_x$

$\mu_2 = \mu_y$

- Degrees of Freedom formula

$$\boxed{DF = n_1 + n_2 - 2}$$

* → Null Hypo:

$$\boxed{t = \frac{(\bar{x} - \bar{y})}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

$\sim t_{n_1+n_2-2}$

→ Remaining part same as previous.

→ $\boxed{\bar{x} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i}$, $\boxed{\bar{y} = \frac{1}{n_2}\sum_{i=1}^{n_2} y_i}$

To find s:

→ $$s^2 = \frac{1}{n_1+n_2-2}\left[\sum_i (x_i - \bar{x})^2 + \sum_i (y_i - \bar{y})^2\right]$$

(or)

$$s^2 = \frac{1}{n_1+n_2-2}\left[(n_1-1)s_1^2 + (n_2-1)s_2^2\right]$$

(or)

$$s^2 = \frac{1}{n_1+n_2-2}\left[n_1 s_1^2 + n_2 s_2^2\right]$$

Example-1:

The following random samples are measurements of the heat-producing capacity of specimens of coal from two mines:

Mine 1: 8260 8350 8130 8070 8340

Mine 2: 7950 7890 7900 7920 8140 7840

Use the 0.01 level of significance to test whether the difference b/w the means of these two samples is significant.

Sol: Step-1:

Set up Null Hypothesis: $\mu_1 - \mu_2 = 0$

Step-2:

Set up Alternative Hypothesis: $\mu_1 - \mu_2 \neq 0$

Step-3:

level of significance $\alpha = 0.01$

Step-4:

Calculations

Degrees of freedom $= n_1 + n_2 - 2$

$= 5 + 6 - 2 = 9$

$n_1 = 5$

$n_2 = 6$

Step-5 :

x	y	$(x_i-\bar{x})$	$\Sigma(x_i-\bar{x})^2$
8260	7950	8260 - 8230 = 30	900
8130	7840	-100	10000
8350	7900	120	14400
8070	7920	-160	25600
8340	8140	110	12100
	7840		
$\Sigma x = 41150$	$\Sigma y = 47640$		$\Sigma(x_i-\bar{x})^2 = 63000$

$$\bar{x} = \frac{\Sigma x}{n_1} = \frac{41150}{5} = 8230$$

$$\bar{y} = \frac{\Sigma y}{n_2} = \frac{47640}{6} = 7940$$

$(y_i-\bar{y})$	$\Sigma(y_i-\bar{y})^2$
7950 - 7940 = -10	100
-50	2500
-40	1600
-20	400
+200	40000
-100	10000
	$\Sigma(y_i-\bar{y})^2 = 54600$

Now,

$$s^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2\right]$$

$$s^2 = \frac{1}{5+6-2}\left[63000 + 54600\right] = 13066.6$$

$s = 114.5$

step-6:

$$t_{cal} = \frac{(\bar{x} - \bar{y}) - 0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t_{cal} = \frac{8230 - 7940}{114.5\sqrt{\frac{1}{5} + \frac{1}{6}}} = \underline{\underline{4.20}}$$

step-7:

$t_{tab} = 3.250$ | $DF = 9$, $\alpha = 0.01$ (Two-tailed)

$\therefore t_{cal} = 4.20 > 3.250 = t_{tab}$

$\therefore t_{cal} > t_{tab}$

$\therefore H_0$ Rejected.

$\therefore$ there is a significant difference b/w two population Means.

* Z-Test for DOUBLE MEAN:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \Sigma(\bar{x}_1 - \bar{x}_2)}{S.E.(\bar{x}_1 - \bar{x}_2)}$$ → No Need

→ z test (for large Sample)

$$\boxed{z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}} \rightarrow \boxed{z = \frac{\bar{x} - \bar{y}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

Here,

$$\boxed{\sigma = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}}$$

→

$$\sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

→ If $\sigma_1^2 \neq \sigma_2^2$ then,

$$\boxed{z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}}$$

Refer PPT for more Problem MODELS

Example - 2:

The means of two single large samples of 1,000 & 2,000 members are 67.5 inches & 68.0 inches respectively. Can the samples be regarded as drawn from the same population of standard Deviation 2.5 inches? Test at 5% level of significance.

Sol:

$n_1 = 1000$, $n_2 = 2000$

$\bar{x}_1 = 67.5$ | $\sigma = 2.5$

$\bar{x}_2 = 68.0$ | $\alpha = 5\% = 0.05$

Null Hyp : $H_0 : \mu_1 = \mu_2$

Alternative Hyp : $H_1 : \mu_1 \neq \mu_2$ (Two - tailed)

Test statistics:

$$z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{|67.5 - 68.0|}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}}$$

$$\boxed{z_{cal} = 5.1}$$

Now,

$z_{table} = 1.96$ | $\alpha = 0.05$ Two - tailed

$\therefore z_{cal} > z_{tab}$

$\therefore H_0$ is Rejected

Conclusion: Difference b/w 2 Population Means.

Terminal Question: (S-18)

③ Measuring - -

Sol:

$n_1 = 8$, $n_2 = 10$

$\bar{x} = 9.67$, $\bar{y} = 7.43$

$\sigma_1 = 1.81$, $\sigma_2 = 1.48$

$\alpha = 0.05$, $\mu_1 - \mu_2 = 1.5$

Now,

$H_0 : \mu_1 - \mu_2 = 1.5$

$H_1 : \mu_1 - \mu_2 > 1.5$ (one-tailed)

$\alpha = 0.05$

$$t_{cal} = \frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} = \frac{(9.67-7.43)-1.5}{\sqrt{\frac{(1.81)^2}{8}+\frac{(1.48)^2}{10}}}$$

$t_{cal} = 0.93$

$\therefore t_{tab} = 1.746$

$\therefore t_{cal} < t_{tab}$

$\therefore H_0$ is Accepted.

Session-19

Chi-Square Test

*

→ **What is chi-square test:**

It is a statistical procedure for determining the diff b/w observed & expected Data.

→ It performs two types of functions:

① Goodness of fit:

②

→ chi-squre Test Formula:

$$\boxed{\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} \sim \chi^2_{(n-1)\ d.f.}}$$

→ Conditions for the Validity of χ^2 test:

i) Sample observations should be independent.

ii) Constraints on the cell frequencies, if any

Example 1:

A business - . . .

	Finance	Sales	Human	Tech	
Satisfied	12	38	5	8	63
Dissatisfied	7	19	3	1	30
Total	19	57	8	9	93

We can use chi square to determine whether the results support & reject the bussiness owner's prediction.

Sol:

$\alpha = 0.05$

$Df = (2-1)(4-1) \rightarrow$ (rows - 1) (colum - 1)

$Df = 1 \times 3 = 3$

O_i	e_i $\frac{RT \times CT}{GT}$	$O_i - e_i$	$(O_i - e_i)^2$	$\Sigma\frac{(O_i - e_i)^2}{e_i}$
12	$\frac{63 \times 19}{93} = 12.8$	-0.87	0.7569	0.0588
38	$\frac{63 \times 57}{93} = 38.61$	-0.61	0.3721	0.0096
5	$\frac{63 \times 8}{93} = 5.42$	-0.42	0.1764	0.0325
8	6.09	1.91	3.6481	0.5990
7	6.129	0.871	0.7586	0.1238

o_i	e_i	$o_i - e_i$	$(o_i-e_i)^2$	$\Sigma \frac{(o_i-e_i)^2}{e_i}$
19	18.39	0.61	0.03721	0.0202
3	2.58	0.42	0.1764	0.0684
1	2.90	-1.9	3.61	1.2448
93	93	0	0	$\chi^2 = 2.1571$

$\therefore \chi^2_{cal} = 2.157$

$\chi^2_{tab} = 7.81 \leftarrow$ | $\alpha = 0.05$, $Df = 3$

$\therefore \chi^2_{cal} < \chi^2_{tab}$

$\Rightarrow H_0$ is Accepted.

→ Example - 2:

The following table represents the no. of boys & the no. of girls who choose each of the 5 possible answers to an item in an scale

	A strongly	Approve	indif	disapp	st dis	total
Boys	25	30	10	25	10	100
girls	10	15	5	15	15	60
total	35	45	15	40	25	160

Sol: $DF = (5-1)(2-1) = 4$

$\alpha = 0.05$

O_i	$E_i = \frac{RT \times CT}{GT}$	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
25	$100 \times 35 / 160 = 21.875$	3.125	9.7656	0.446
30	28.125	1.875	3.5156	0.124
10	9.375	0.625	0.390	0.041
25	25	0	0	0
10	15.625	-5.625	31.6406	2.024
10	13.125	-3.125	9.7656	0.744
15	16.87	-1.87	3.4969	0.186
5	5.620	-0.620	0.3844	0.064
15	15	0	0	0
15	9.375	5.625	31.6406	3.374
				$\Sigma = 7.096$

$\chi^2_{cal} = 7.026$

$\chi^2_{tab} = 9.49$ | $\alpha = 0.05$, $DF = 4$

$\therefore \chi^2_{cal} < \chi^2_{tab}$

$\therefore H_0$ is accepted.

Hence there is no significant difference towards the question b/w boys & girls that they are independent.

Terminal :

① The following table gives the no. of accidents that work place in an industry surveying varies days of the week. Test if the accidents are uniformly distributed over the week.

Days :	Mon	Tue	Wed	Thu	Fri	Sat
No. of Accidents	14	18	12	11	15	14

Sol:

$\alpha = 0.05$ ---> rows no.

Df : $n-1 \rightarrow 6-1=5$

$E_i = \frac{\Sigma o_i}{n}$

$84/6 = 14$

O_i	E_i	$o_i - e_i$	$(o_i-e_i)^2$	$(o_i-e_i)^2/e_i$
14	14	0	0	0
18	14	4	16	1.143
12	14	-2	4	0.28
11	14	-3	9	0.64
15	14	1	1	0.071
14	14	0	0	0

$\Sigma o_i = 84$

$\chi^2 = 2.132$

$\chi^2_{cal} = 2.132$

$\chi^2_{tab} = 11.07$ / $\alpha = 0.05, Df = 5$

$\therefore \chi^2_{cal} = 2.132 < \chi^2_{tab} = 11.07$

$\therefore \chi^2_{cal} < \chi^2_{tab}$

$\therefore H_0$ is Accepted.

③ The figures given below are a) the observed frequencies b) the theoretical frequencies of a normal distribution respectively:

(a) o_i	1	5	20	28	42	22	15	5	2
e_i	1	6	18	25	40	25	18	6	1

Applying the t-test of goodness of fit for above data & comment?

Sol: $\alpha = 0.05$

$Df = 9 - 1 = 8$

o_i	e_i	$(o_i - e_i)$	$(o_i - e_i)^2$	$(o_i - e_i)^2 / e_i$
1	1	0	0	0
5	6	-1	1	$1/6 = 0.166$
20	18	2	4	0.22
28	25	3	9	0.36

o_i	e_i	$(o_i - e_i)$	$(o_i - e_i)^2$	$(o_i - e_i)^2 / e_i$
42	40	2	4	0.1
22	25	-3	9	0.36
15	18	-3	9	0.5
5	6	-1	1	0.166
2	1	1	1	1

$$\chi^2 = 2.872$$

$$\chi^2_{cal} = 2.872$$

$$\chi^2_{tab} = 15.51$$

$\therefore \chi^2_{cal} < \chi^2_{tab}$

$\therefore H_0$ is Accepted.

② Among 64 off springs of a ceitain cross blw guinea pigs 34 were red, 10 were black & 20 were white. According to the genetic model these numbers should be in the ratio 9:3:4. Are the data consistent with the model at 5% level.

Sol:

H_0: Data is consistent with model.

H_1: Data is inconsistant with model.

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
34	$9/16 \times 64 = 36$	-2	4	0.11
10	$3/16 \times 64 = 12$	-2	4	0.33
20	$4/16 \times 64 = 16$	4	16	1

$\chi^2_{cal} = 1.44$

$\chi^2_{tab} = 5.99$ | $\alpha = 0.05$, $Df = 2$

$\therefore \chi^2_{cal} < \chi^2_{tab}$

$\therefore H_0$ is Accepted.

* Session-20

ANNOVA

→ Definition: Analysis of variance is the "Separation of variance ascribable to one group of causes from the variance ascribable to other group.

→ ANOVA test is based on the test statistic F.

Validity of F-test :

i) observations are independent

ii) parent population from which observation are taken

~~(?)~~

$$F = \frac{\text{Variance b/w groups } (\alpha^2)}{\text{Variance with in grps}}$$

→ ANOVA One Way Classification Process:

Step-1: $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$

Step2: H_1 : Atleast one of the α's is not equal to 0.

Step 3: Obtain all the sum of raw scores & the squares of raw scores. Write them at the end of each column.

Step 4: obtain grand sums of raw srs. & square of raw square as Σx^2

Step-5: Calculate the Correction Factor (CF) by formula:

$$\boxed{CF = G^2 / N}$$ G = Grand total of all obs.

step-6:

Calculate the sum of squares (S.S.A)

$$\boxed{SSA = \frac{\sum x^2}{N} - cf}$$ (or)

$$\boxed{SSA = \frac{\sum x_1^2}{n_1} + \frac{\sum x_2^2}{n_2} + \cdots \frac{\sum x_k^2}{n_k} - CF}$$

step 7: Calculate the total sum of squares (SST)

$$\boxed{SST = \sum\sum x_{ij}^2 - cf}$$

step-8: Calculate the sum of square within grps & Error sum of square.

$$\boxed{SSE = SST - SSA}$$

step-9: Calculate the Degrees of Freedom as:

DF SSA = $k-1$

DF SST = $n-1$

DF SSE = $n-k$

k = no. of groups

n = total number

Step-10: Find the Mean Sum of squares of 2 vari.. as

b/w group $\boxed{MSSA = \frac{SSA}{k-1}}$

~~SECTION - Q20~~

within group $\boxed{MSSE = \frac{SSE}{n-k}}$

Step-11:

Sources of variation	DF	Sum of Square	Mean sum sqr	Variance ratio	
				F-Calc value	F-tabluted value
B/w grps	K-1	SSA	MSSA = F_1	$\frac{MSSA}{MSSE} = \frac{F_1}{F_2}$	$\frac{F_1}{F_2 \cdot F_{(k-1, N-k)}}$
within grps	N-K	SSE	MSSE = F_2		
Total	N-1	SST			

Step-12: State Conclusion

Example-I

Obs	low	Moderate	High
1	120	61	40
2	68	59	55
3	40	110	73
4	95	75	45
5	83	80	64

House Pricing

Calculate H & report your results with the P-value of 0.05

Sol:

Step-1: Set up the Null Hyp

H_0: There is no significant diff b/w the mean housing prices of 3 crowded places.

Step 2: H_1: There is significant diff b/w the mean H. Prices of 3. crowded Places.

Step-3&4:

Obser	Low	Mod	High	
1	120	61	40	
2	68	59	55	
3	40	110	73	
4	95	75	45	
5	83	80	64	
Total T_i	$t_1 = 406$	$T_2 = 385$	$T_3 = 277$	$G = 1068$
T_i^2	$T_1^2 = 164836$	$T_2^2 = 148225$	$T_3^2 = 76729$	

sum of squares of each element.

step 5:

$$cf = (1068)^2 / 15 = 76041.6$$
$$\downarrow$$
$$5+5+5$$

$$\therefore \boxed{cf = \frac{G^2}{N}}$$

step-6:

$$SSA = \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} - cf$$

$$= \frac{164836}{5} + \frac{148225}{5} + \frac{76729}{5} - 76041.6$$

$$\underline{SSA = 1916.4}$$

step-7:

$$SST = \sum\sum x_{ij}^2 - cf$$

$$SST = (120)^2 + 68^2 + 40^2 + 95^2 + 83^2 + 61^2 + 59^2 + 110^2 + 75^2 + 80^2 + 40^2 + 55^2 + 73^2 + 45^2 + 64^2 - cf$$

$$= 83940 - 76041.6$$

$$SST = 7898.4$$

step-8:

$$SSE = SST - SSA$$

$$SSE = 7898.4 - 1916.4$$

$$SSE = 5982$$

step-9:

$$k = 3$$
$$N = 15$$

DF $SSA = k - 1 = 3 - 1 = 2$

DF $SST = N - 1 = 15 - 1 = 14$

DF $SSE = N - K = 15 - 3 = 12$

Step-10:

$$F_1 = MSSA = \frac{SSA}{K-1} = \frac{1916.4}{2}$$

$$= 958.2$$

$$F_2 = MSSE = \frac{SSE}{N-K} = \frac{5982}{12} = 498.5$$

$$F_{cal} = \frac{F_1}{F_2} = \frac{958.2}{498.5}$$

$$F_{cal} = 1.922$$

$F(v_1, v_2)$

$F(K-1, N-K)$ v_2

$\boxed{2 (v_1) \quad 12}$ see in table

Step-11: ANOVA table

Source	DF	S O Square	Mean sum	Variance ratio	
				F-cal	F-tab
B/w the grps (colums)	3-1 = 2	1916.4	958.2	1.922	3.88
within group	15-3 = 12	5982	5982/12 = 498.5		
Total	15-1=14	7898.4			

$\therefore F_{cal} = 1.92 < 3.88 = F_{tab}$

$\therefore F_{cal} < F_{tab}$ $\therefore H_0$ is Accepted.

Step-12:

There is no significant dif b/w the mean housing prices of 3 places.