Sampling Theory

- Large Sampling ($n \geq 30$, no. of sample) → z - test
- Small Sampling ($n \leq 30$) → t - test
- χ - test

Notes :

	Mean/avg	std deviation	Variance
population	μ	σ	σ^2
sample	$\bar{x}$	s	s^2

→ variance = (std. deviation)^2

→ If σ not given then replace it by s

Large sampling (one of case)

$\Downarrow$

$n > 30$

$\downarrow$

z-test

steps : 1] Setting of Hypothesis

Null Hypothesis $(H_0) \rightarrow \mu = \ldots$

Alternate Hypothesis $(H_A) \rightarrow \mu \neq \ldots$

2] Test statistics : $|z| = \left| \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \right| = \ldots$

3] Level of significance (α) :

$\alpha = 5\%$ or $\alpha = 1\%$

$\downarrow$

assume

4] Critical value (z_α)

$z_\alpha = 1.96$ for $\alpha = 5\%$

$z_\alpha = 2.576$ for $\alpha = 1\%$

5] Decision :

If $z < z_\alpha$ ... Null Hypothesis accepted

If $z > z_\alpha$ ... Null Hypothesis rejected

& comment

PAGE No.
DATE / /

1] A random sample of 50 items gives the mean 6.2 and variance 10.24. Can it be regarded as drawn from a normal population with mean 5.4 at 5% level of significance

A. $n = 50$

$\text{mean} = \bar{x} = 6.2$

$s^2 = 10.24$

$s = 3.2$

$\alpha = 5\%$

$z_\alpha = 1.96$

Setting of Hypothesis

Null Hypothesis (H_0) $\rightarrow \mu = 5.4$

Alternate Hypothesis (H_A) $\rightarrow \mu \neq 5.4$

Test statistics

$$|z| = \left| \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \right| = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \right| = \left| \frac{6.2 - 5.4}{\frac{3.2}{\sqrt{50}}} \right| = 1.76$$

LOS : $\alpha = 5\%$

Critical value : $z_\alpha = 1.96$

Decision

$z < z_\alpha$

Null Hypothesis accepted

Yes, the sample can be regarded as drawn from a normal population

2] A random sample of 400 members is found to have a mean of 4.45 cm. Can it be reasonably regarded as a sample from a large population whose mean is 5cm and variance is 4cm

$n = 400$

$\bar{x} = 4.45$

$\mu = 5$

$\sigma^2 = 4$

$\sigma = 2$

Setting of Hypothesis :

Null Hypothesis (H_0) $\rightarrow \mu = 5$

Alternate Hypothesis (H_A) $\rightarrow \mu \neq 5$

Test Statistics :

$$|Z| = \left| \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \right| = \left| \frac{4.45 - 5}{\frac{2}{\sqrt{400}}} \right| = 5.5$$

LOS :

Assume $\alpha = 5\%$.

Critical value :

$z_\alpha = 1.96$

Decision :

$z > z_\alpha$ ... Null Hypothesis rejected.

The sample is not drawn from a large population whose mean is 5cm

3] Can it be concluded that the average life-span of an Indian is more than 70 years, if a random sample of 100 Indians has an average life span of 71.8 years with standard deviation of 8.9 years?

$n = 100$

$\bar{x} = 71.8$

$\mu = 70$

$s = 8.9$

Setting of Hypothesis :

Null Hypothesis (H_0) $\rightarrow \mu = 70$

Alternate Hypothesis (H_A) $\rightarrow \mu \neq 70$

Test statistics:

$$|z| = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \right| = \left| \frac{71.8 - 70}{\frac{8.9}{\sqrt{100}}} \right| = 2.02$$

LOS :

Assume $\alpha = 5\%$

Critical value :

$z_\alpha = 1.96$

Decision

$z > z_\alpha$

Null Hypothesis rejected

It cannot be concluded with this sample that the average life-span of an Indian is more than 70 years

4] A tyre company claims that the lives of tyres have mean 42000 km with SD of 4000 km. A change in the production process is believed to result in better product. A test sample of 81 new tyres has a mean life of 42,500km. Test at 5% LOS that the new product is significantly better than the old one.

$n = 81$

$\mu = 42000$

$\bar{x} = 42500$

$\sigma = 4000$

Setting of Hypothesis :

Null Hypothesis (H_0) $\mu = 42000$ .... No improvement

Alternate Hypothesis (H_A) $\mu \neq 42000$ ... Improved.

Test statistics :

$$|z| = \left| \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \right| = \left| \frac{42500 - 42000}{\frac{4000}{\sqrt{81}}} \right| = 1.125$$

LOS : $\alpha = 5\%$

Critical value :

$z_\alpha = 1.96$

Decision :

$z_\alpha > z$

Null Hypothesis accepted

The new product has no improvement.

DATE / /

5] A random sample of 900 items is found to have a mean of 65·3 cm. Can it be regarded as a sample from a large population whose mean is 66·2 cm. And standard deviation is 5 cm. At 5% level of significance ?

$n = 900$

$\bar{x} = 65 \cdot 3$

$\mu = 66 \cdot 2$

$\sigma = 5$

Setting of Hypothesis :

Null Hypothesis $(H_0) \longrightarrow \mu = 66 \cdot 2$

Alternate Hypothesis $(H_A) \longrightarrow \mu \neq 66 \cdot 2$

Test statistics :

$$|Z| = \left| \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \right| = \left| \frac{65 \cdot 3 - 66 \cdot 2}{\frac{5}{\sqrt{100}}} \right| = 5 \cdot 4$$

LOS : 5%

Critical value

$Z_\alpha = 1 \cdot 96$

Decision

$Z_\alpha < Z$

Null Hypothesis rejected

It cannot be regarded as a sample from this population

6] A machine is set to produce metal plates of thickness 1.5 cm with standard deviation of 0.2 cm. A sample of 100 plates produced by machine gave an average thickness of 1.52 cm. Is the machine fulfilling the purpose?

$n = 100$

$\bar{x} = 152$

$\mu = 1.5$

$\sigma = 0.2$

Setting of Hypothesis

Null Hypothesis (H_0) $\Rightarrow$ $\mu = 1.5$

Alternate Hypothesis (H_A) $\rightarrow$ $\mu \neq 1.5$

Test statistics :

$$|z| = \left| \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \right| = \left| \frac{1.52 - 1.50}{\frac{0.2}{\sqrt{100}}} \right| = 1$$

LOS :

$\alpha = 5\%$ (assumed)

Critical value :

$z_\alpha = 1.96$

$z < z_\alpha$

Null Hypothesis accepted

$\therefore$ Yes the machine is fulfilling its purpose.

Small Sampling ... $n \leq 30$

↓

t · test

(I) one case of sample

(II) Two case of sample

(I) One case of sample :-

Steps : 1] Setting of Hypothesis

Null Hypothesis $(H_0) \Rightarrow \mu = \ldots$

Alternate Hypothesis $(H_A) \Rightarrow \mu \neq \ldots$

2) Test statistics

$$|t| = \left| \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n-1}}} \right| = \ldots$$

3] LOS : same

4] Critical value :

$t_\alpha \to \alpha$, $\nu = n - 1$ } check in t · chart

t_α : Chart Reading

ν : degree of freedom

$\alpha \to$, $\nu \downarrow$

5) Decision

If $t < t_\alpha$ ... Null Hypothesis accepted

If $t > t_\alpha$ ... Null Hypothesis rejected & Comment

Note :- 1) If σ not given then use S

2) If σ & S both not given then get $S = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$

Sum of squares of deviation

1] A random sample of size 16 from a normal population showed a mean of 103.75 and sum of squares of deviations from the mean 843.75 cm^2. Can we say that the population has a mean of 108.75 cm.

$n = 16$

$\bar{x} = 103.75$

$\Sigma(x_i - \bar{x})^2 = 843.75$

$\bar{\mu} = 108.75$ cm

Setting of Hypothesis

Null Hypothesis $(H_0) \Rightarrow \mu = 108.75$ cm

Alternate Hypothesis $(H_A) \Rightarrow \mu \neq 108.75$ cm

Test statistics :

$$|t| = \left| \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n-1}}} \right| = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \right| = \left| -2.66 \right| = 2.66$$

for s,

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$$

$$s = \sqrt{\frac{843.75}{16}} = 7.26$$

LOS : $\alpha = 5\%$.

Critical value : $t_\alpha \begin{cases} \alpha = 5\% = 0.05 \\ \nu = n-1 = 15 \end{cases}$

$$t_\alpha = 2.131$$

$$t > t_\alpha$$

Null Hypothesis rejected

The population doesn't have a mean of 108.75 cm

2] Ten individuals are chosen at random from a population and their heights are found to be 63, 63, 64, 65, 66, 69, 69, 70, 70, 71 inches. Discuss the suggestion that the mean height of the universe is 65 inches

Given : $n = 10$

$\mu = 65$

$x_i = 63, 63, 64, 65, 66, 69, 69, 70, 70, 71$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{63+63+64+65+66+69+69+70+70+71}{10}$$

$$\bar{x} = 67$$

solution :

x_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$
63	-4	16
63	-4	16
64	-3	9
65	-2	4
66	-1	1
69	2	4
69	2	4
70	3	9
70	3	9
71	4	16

$\Sigma(x_i-\bar{x})^2 = 88$

$$S = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n}} = \sqrt{\frac{88}{10}} = 2.96$$

Setting of Hypothesis

Null Hypothesis $\Rightarrow \mu = 65$

Alternate Hypothesis $\Rightarrow \mu \neq 65$

Test statistics :

$$|t| = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \right| = \left| \frac{67 - 65}{\frac{2.96}{\sqrt{9}}} \right| = 2.02$$

LOS : $\alpha = 5\%$

Critical value : t_α $\to \alpha = 5\% = 0.05$, $\to \nu = n-1 = 9$

$t_\alpha = 2.262$

$t < t_\alpha$

Null Hypothesis accepted.

The mean height of the universe may be 65 inches.

3] Tests made on breaking strength of 10 pieces of a metal wire gave the following 578, 572, 570, 568, 572, 570, 570, 572, 596 & 584 in kgs. Test if the breaking strength of the metal wire can be assume to be 577 kg?

$n = 10$

$\mu = 577$

$x_i = 578, 572, 570, 568, 572, 570, 570, 572, 596, 584$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{5752}{10} = 575.2$$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
578	2.8	7.84
572	-3.2	10.24
570	-5.2	27.04
568	-7.2	51.84
572	-3.2	10.24
570	-5.2	27.04
570	-5.2	27.04
572	-3.2	10.24
596	20.8	432.64
584	8.8	77.44

$$\Sigma (x_i - \bar{x})^2 = 681.6$$

$$S = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}} = \sqrt{\frac{681.6}{10}} = 8.25$$

Setting of Hypothesis

Null Hypothesis (H_0) $\Rightarrow \mu = 577$

Alternate Hypothesis (H_A) $\Rightarrow \mu \neq 577$

Test statistics

$$|t| = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \right| = \left| \frac{575.2 - 577}{\frac{8.25}{\sqrt{9}}} \right| = 0.65$$

LOS : $\alpha = 5\%$

Critical value :

$$t_\alpha \begin{cases} \rightarrow \alpha = 0.05 \\ \rightarrow \nu = n-1 = 9 \end{cases}$$

$t_\alpha = 2.262$ ... from chart

Decision:

$t < t_\alpha$

$\therefore$ Null Hypothesis accepted

Mean Breaking strength of metal wire is 577 kg.

Note :- If pop^n mean 'μ' is not given then take $\mu = 0$
If two different datas of same samples are given then subtract these data to get x_i

PAGE NO.
DATE / /

4] A certain injection administered to 12 patients resulted in the following changes of blood pressure : 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4. Can it be concluded that the injection will be general accompanied by an increase in blood pressure.

$n = 12$

$\mu = 0$ ... assume

$x_i = 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4$

$\Sigma x_i = 31$

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{31}{12} = 2.58$

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
5	2.42	5.85
2	-0.58	6.33
8	5.42	29.37
-1	-3.58	12.81
3	0.42	0.17
0	-2.58	6.65
6	3.42	11.69
-2	-4.58	20.97
1	-1.58	2.49
5	2.42	5.85
0	-2.58	6.65
4	1.42	2.01

$\Sigma (x_i - \bar{x}) = 141.47$

$$S = \sqrt{\frac{(x_i - \bar{x})^2}{n}} = \sqrt{\frac{141.47}{12}} = 3.43$$

Setting of Hypothesis :

Null Hypothesis (H_0) $\Rightarrow \mu = 0$ No rise

Alternate Hypothesis (H_A) $\Rightarrow \mu \neq 0$ rise in BP

test Statistics :

$$|t| = \left| \frac{\bar{x} - \bar{\mu}}{\frac{s}{\sqrt{n-1}}} \right| = \left| \frac{2 \cdot 58 - 0}{\frac{3 \cdot 43}{\sqrt{11}}} \right| = 2 \cdot 49$$

LOS : $\alpha = 5\%$

Critical value : t_α $\rightarrow \alpha = 5\% = 0 \cdot 05$, $\rightarrow \nu = n - 1 = 11$

$t_\alpha = 2 \cdot 201$

Decision :

$t > t_\alpha$

Null Hypothesis rejected

$\therefore$ There is a rise in mean BP.

5] Ten school boys were given a test in statistics and their scores were recorded. They were given a month's special coaching and a second test was given to them in the same subject at the end of the coaching period. Test if the marks given below give evidence to the fact that the students are benefitted by coaching

Marks in Test I : 70, 68, 56, 75, 80, 90, 68, 75, 56, 58
Marks in Test II : 68, 70, 52, 73, 75, 78, 80, 92, 54, 55

$n = 10$

$x_i = T_1 - T_2$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
2	-1.3	1.69
-2	-5.3	28.09
4	0.7	0.49
2	-1.3	1.69
5	1.7	2.89
12	8.7	75.69
-12	-15.3	234.09
17	13.7	187.69
2	-1.3	1.69
3	-0.3	0.09
$\Sigma x_i = 33$		$\Sigma (x_i - \bar{x})^2 = 532.41$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{33}{10} = 3.3$$

$$s = \sqrt{\frac{\Sigma (x_i - \bar{x})}{n}} = \sqrt{\frac{532.41}{10}} = 7.296$$

Setting of Hypothesis

Null Hypothesis $\mu = 0$ Not benefitted

Alternate Hypothesis $\mu \neq 0$ Benefitted

Test statistics

$$|t| = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \right| = \left| \frac{3.3 - 0}{\frac{7.296}{\sqrt{9}}} \right| = 1.356$$

LOS : $\alpha = 5\%$

Critical value : t_α $\rightarrow \alpha = 0.05$, $\rightarrow \nu = 9$

$t_\alpha = 2.262$

$t < t_\alpha$

Null Hypothesis accepted.

Students were not benefitted by caaching

(II) Two case of sample

Steps : 1] Setting of Hypothesis

Null Hypothesis (H_0) $\Rightarrow \mu_1 = \mu_2$

Alternate Hypothesis (H_A) $\Rightarrow \mu_1 \neq \mu_2$

2] Test Statistics

For regular case

$$S_p = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$$

For unbiased SD

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

For different Datas (x_i) or $\Sigma(x_i - \bar{x})$ is given

$$S_p = \sqrt{\frac{\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2}{n_1 + n_2 - 2}}$$

$$SE = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Std error

$$|t| = \left| \frac{\bar{x}_1 - \bar{x}_2}{SE} \right|$$

3] LOS : same

4] Critical value : $t_\alpha \begin{cases} \alpha = \ldots \\ \nu = n_1 + n_2 - 2 = \ldots \end{cases}$ from chart

5] Decision :

same

1] A sample of 8 students of 16 years each shown up a mean systolic blood pressure of 118.4 mm of Hg with SD of 12.17 mm while sample of 10 students of 17 years each showed the mean systolic BP of 121.0 mm with SD of 12.88 during an investigation. The investigator feels that the systolic BP is related to age. Do you think that the data provides enought reasons to support investigators feeling at 5% LOS (Assume the distribution of systolic BP to be normal.

$n_1 = 8$ $\qquad$ $n_2 = 10$

$\bar{x}_1 = 118.4$ $\qquad$ $\bar{x}_2 = 121.0$

$s_1 = 12.17$ $\qquad$ $s_2 = 12.88$

Setting of Hypothesis

Null Hypothesis $(H_0) \Rightarrow \mu_1 = \mu_2$

Alternate Hypothesis $(H_A) \Rightarrow \mu_1 \neq \mu_2$

Test statistics :

$$S_p = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{8(12.17)^2 + 10(12.88)^2}{8 + 10 - 2}} = 13.33$$

$$\text{std error} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 13.33 \sqrt{\frac{1}{8} + \frac{1}{10}} = 6.32$$

$$|t| = \left| \frac{\bar{x}_1 - \bar{x}_2}{SE} \right| = \left| \frac{118.4 - 121}{6.32} \right| = 0.41$$

LOS : $\alpha = 5\%$

Critical value : t_α $\rightarrow \alpha = 5\% = 0.05$ =
$\rightarrow \nu = n_1 + n_2 - 2 = 16$

$$t_\alpha = 2.120$$

Decision : $t < t_\alpha$

Null Hypothesis accepted

BP is not related to age

2] Sample of two types of electric bulbs were tested for length of life and following data were obtained

	Type I	Type II
No. of samples	8	7
Mean of samples (in hrs)	~~8~~1134	1024
Std Deviation (in hrs)	35	40

Test at 5% level of significance whether the difference in the sample means is significant

$n_1 = 8$ $\quad n_2 = 7$

$\bar{x}_1 = 1134$ $\quad \bar{x}_2 = 1024$

$s_1 = 35$ $\quad s_2 = 40$

Setting of Hypothesis:

Null Hypothesis $(H_0) \Rightarrow \mu_1 = \mu_2$

Alternate Hypothesis $(H_a) \Rightarrow \mu_1 \neq \mu_2$

Test statistics:

$$S_p = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}} = \sqrt{\frac{8(35)^2 + 7(40)^2}{8+7-2}} = 40.19$$

$$SE = S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 40.19\sqrt{\frac{1}{8}+\frac{1}{7}} = 20.800$$

$$|t| = \left|\frac{x_1 - x_2}{SE}\right| = \left|\frac{1134 - 1024}{20.8}\right| = 5.28$$

LOS: $\alpha = 5\%$

Critical value: t_α → $\alpha = 5\% = 0.05$ → $\nu = n_1+n_2-2 = 13$

$t_\alpha = 2.160$

$t > t_\alpha$

Difference is significant.

3) Six guinea pigs injected with 0.5 mg. of a medication took an average 15.4 sec. To fall asleep with an unbiased standard deviation 2.2 sec, while other six guinea pigs injected with 1.5 mg of medication took on an average 11.2 sec to fall asleep with an unbiased standard deviation 2.6 cm. Use 5% LOS to test the null Hypothesis that the difference in dosage has no effect.

$n_1 = 6$ $\quad n_2 = 6$

$\bar{X}_1 = 15.4$ $\quad \bar{X}_2 = 11.2$

$S_1 = 2.2$ $\quad S_2 = 2.6$

Setting of Hypothesis:

Null Hypothesis (H_0) $\Rightarrow \mu_1 = \mu_2$

Alternate Hypothesis (H_A) $\Rightarrow \mu_1 \neq \mu_2$

Test statistics:

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} = \sqrt{\frac{5(2.2)^2 + 5(2.6)^2}{10}} = 2.40$$

$$SE = S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 2.40\sqrt{\frac{1}{6} + \frac{1}{6}} = 1.38$$

$$|t| = \left|\frac{\bar{X}_1 - \bar{X}_2}{SE}\right| = \left|\frac{15.4 - 11.2}{1.38}\right| = 3.04$$

LOS : $\alpha = 5\%$

Critical value: t_α $\rightarrow \alpha = 5\% = 0.05$

$\rightarrow \nu = n_1 + n_2 - 2 = 10$

$t_\alpha = 2.228$

Decision:

$t > t_\alpha$

Null Hypothesis rejected

The Difference in dosage has effect.

PAGE No.
DATE / /

4] The means of two random samples of size 9 & 7 are 196.42 and 198.82 respectively. The sum of squares of deviations from the means are 26.94 and 18.73 respectively. Can the samples be considered to have been drawn from the same population?

5]

$n_1 = 9$ $\quad n_2 = 7$

$\bar{x}_1 = 196.42$ $\quad \bar{x}_2 = 198.82$

$\Sigma(x_i - \bar{x})^2 = 26.94$ $\quad \Sigma(y_i - \bar{y})^2 = 18.73$

Setting of Hypothesis

Null Hypothesis $(H_0) \Rightarrow \mu_1 = \mu_2$

Alternate Hypothesis $(H_A) \Rightarrow \mu_1 \neq \mu_2$

Test statistics:

$$S_p = \sqrt{\frac{\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2}{n_1+n_2-2}} = \sqrt{\frac{26.94+18.73}{14}} = \cancel{0.76}\ 1.80$$

$$SE = S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 1.80\sqrt{\frac{1}{9}+\frac{1}{7}} = \cancel{0.38}\ 0.90$$

$$|t| = \left|\frac{\bar{x}_1 - \bar{x}_2}{SE}\right| = \left|\frac{196.42-198.82}{0.90}\right| = 2.66$$

LOS : $\alpha = 5\%$

Critical value : $t_\alpha \begin{cases} \rightarrow \alpha = 5\% = 0.05 \\ \rightarrow \nu = n_1+n_2-2 = 14 \end{cases}$

$t_\alpha = 2.145$

Decision:

$t > t_\alpha$

Null Hypothesis ~~accepted~~ rejected

They are not drawn from the same population

5] The heights of six randomly chosen sailors are in inches : 63, 65, 68, 69, 71 and 72. The height of ten randomly chosen soldiers are : 61, 62, 65, 66, 69, 70 and 71, 72, 73. Discuss in light that these data throw on the suggestion that the soldiers on the average are taller than sailors.

$n_1 = 6 \quad x_i = 63, 65, 68, 69, 71, 72$

$n_2 = 10 \quad y_i = 61, 62, 65, 66, 66, 69, 70, 71, 72, 73$

Setting of Hypothesis :

Null Hypothesis $(H_0) \Rightarrow \mu_1 = \mu_2$

Alternate Hypothesis $(H_A) \Rightarrow \mu_1 \neq \mu_2$

Test statistics :-

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	y_i	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
63	-5	25	61	-6.5	42.25
65	-3	9	62	-5.5	30.25
68	0	0	65	-2.5	6.25
69	1	1	66	-1.5	2.25
71	3	9	66	-1.5	2.25
72	4	16	69	1.5	2.25
$\bar{x}_1 = \frac{\Sigma x_i}{N} = \frac{408}{6}$		(60)	70	2.5	6.25
			71	3.5	12.25
$= 68$			72	4.5	20.25
			73	5.5	30.25
			$\bar{x}_2 = 67.5$		(154.5)

$$sp = \sqrt{\frac{(60) + (154.5)}{14}} = 3.91$$

$$SE = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 3.91 \sqrt{\frac{1}{6} + \frac{1}{10}}$$

$$= 2.019$$

$$|t| = \left| \frac{\bar{X}_1 - \bar{X}_2}{SE} \right|$$

$$= \left| \frac{68 - 67.5}{2.019} \right|$$

$$= 0.24$$

LOS : $\alpha = 5\%$

Critical value : t_α → $\alpha = 0.05$, $\nu = 14$

$$t_\alpha = 2.145$$

Decision :

$t < t_\alpha$

Null Hypo accepted

No soldiers taller than sailors

χ^2 - test

① Dependancy between attributes

② Goodness of fit

① Dependancy between attributes [Tabular sums]

Steps : 1] Setting of Hypothesis :

Null Hypothesis (H_0) $\Rightarrow$ there is no relation

Alternate Hypothesis (H_A) $\Rightarrow$ there is relation

2] Test statistics :

$$\chi^2 = \Sigma \left[\frac{(O - E)^2}{E} \right]$$

expected freq. (we calculated)

observed freq (given in question)

For E →

case I

case II

	Yes 1	No 1	Total
Yes 2	E_1	E_2	fix
No 2	adjust	adjust	fix
Total	fix	fix	grand Total

$$E_1 = \frac{(1^{st} \text{ row total}) \cdot (1^{st} \text{ column total})}{\text{Grand Total}}$$

$$E_2 = \frac{(1^{st} \text{ row total}) \cdot (2^{nd} \text{ column total})}{\text{Grand Total}}$$

NOTE :-

Tabular → Dependancy

Table (x_i, f_i) → goodness of fit

LOS : $\alpha = 5\%$ ... assume

Critical value : χ^2_{crit} [→ $\alpha = \ldots$; → $\nu = (r-1)(c-1)$]

r → no. of rows ; c → no. of column

Decision :

$\chi^2 < \chi^2_{crit}$ ... Null Hypo accepted

$\chi^2 > \chi^2_{crit}$ ... rejected & comment

PAGE No.
DATE / /

17. Investigate the association between the darkness of eye colour in father and son from the following data.

		Colour of father's eyes		
		Dark	Not Dark	Total
Colour of son's eyes	Dark	48	90	138
	Not Dark	80	782	862
	Total	128	872	1000

Setting of Hypothesis:

Null Hypothesis ⇒ There is no association

Alternate Hypothesis ⇒ There is association.

Test statistics:

For E →	Dark	Not Dark	Total
Dark	18	120	138
Not Dark	110	752	862
Total	128	872	1000

$$E_1 = \frac{128 \times 138}{1000} = 18$$

$$E_2 = \frac{872 \times 138}{1000} = 120$$

father

O	E	$(O-E)^2$	$(O-E)^2/E$
48	18	900	50
90	120	900	7.5
80	110	900	8.18
782	752	900	1.19

$$\Sigma\left(\frac{(O-E)^2}{E}\right) = 66.87$$

$$\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right]$$

$$\chi^2 = 66.87$$

LOS : $\alpha = 5\%$

Critical value : χ^2_{crit} → $\alpha = 5\% = 0.05$
→ $\nu = (r-1) \times (c-1) = 1$

$\chi^2_{crit} = 3.841$ ... from chart

Decision :

$\chi^2 > \chi^2_{crit}$

Null Hypothesis rejected

There is association between the darkness of eye colour in father and son.

ईमोल => For tabular sums. If any freq is less than 5 then Yates correction is needed

$$\chi^2 = \Sigma\left[\frac{(|O-E|-0.5)^2}{E}\right]$$

2] Two batches of 12 animals each are given test of inoculation. One batch was inoculated and other was not. The number of dead and surviving animals are given in the following table for both cases. Can the inoculation be regarded as effective against the disease at 5% level of significance (Make Yates correction).

	Dead	Surviving	Total
Inoculated	2	10	12
Not-inoculated	8	4	12
Total	10	14	24

Setting of Hypothesis:

Null Hypothesis (H_0) => Not effective

Alternate Hypothesis (H_A) => effective

Test statistics:

For E →

	Dead	Surviving	Total
Inoculated	5	7	12
Not-inoculated	5	7	12
Total	10	14	24

$$E_1 = \frac{10 \times 12}{24} = 5$$

O	E	$(O-E)^2$	$(O-E)^2/E$	$(\lvert O-E\rvert-0.5)^2$	$\frac{(\lvert O-E\rvert-0.5)^2}{E}$
2	5	9	1.8	6.25	1.25
10	7	9	1.28	6.25	0.89
8	5	9	1.8	6.25	1.25
4	7	9	1.28	6.25	0.89
			$\Sigma = 6.17$		

$$\cancel{\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right] = 6.17} \qquad \chi^2 = \Sigma\left[\frac{(\lvert O-E\rvert-0.5)^2}{E}\right] = 4.28$$

LOS : $\alpha = 5\%$

Critical value : $\chi^2_{crit} \begin{cases} \to \alpha = 5\% = 0.05 \\ \to \nu = (r-1)(c-1) = 1 \end{cases}$

$$\chi^2_{crit} = 3.841$$

$$\chi^2 > \chi^2_{crit}$$

Null Hypothesis rejected.

Inoculation is effective again disease.

3] To test the effect of a new drug, a controlled experiment was conducted. 300 patients were given the new drug while 200 patients were given no drug. On the basis of examination of these persons, the following results were obtained.

	Cured	Condition worsened	Not effect	Total
Given the new drug	200	40	60	300
Not given the drug	120	30	50	200
Total	320	70	110	500

Setting of Hypothesis:

Null Hypothesis (H_0) $\Rightarrow$ There is no effect

Alternate Hypothesis (H_A) $\Rightarrow$ The effect is significant.

Test statistics:

For E $\rightarrow$

	Cured	Condition worsened	Not effect	Total
Given the new drug	192	42	66	300
Not given the drug	128	28	44	200
Total	320	70	110	500

$$E_1 = \frac{320 \times 300}{500} = 192$$

$$E_2 = \frac{70 \times 300}{500} = 42$$

O	E	$(O-E)^2$	$(O-E)^2/E$
200	192	64	0.33
40	42	4	0.09
60	66	36	0.54
120	128	64	0.5
30	28	~~8~~4	0.14
50	44	36	0.81
			$\Sigma = 2.41$

$$\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right] = 2.41$$

LOS : $\alpha = 5\%$

Critical value : $\chi^2_{crit} \begin{cases} \alpha = 5\% = 0.05 \\ \nu = (r-1)(c-1) = (1)(2) = 2 \end{cases}$

$\chi^2_{crit} = 5.991$ ... from chart

$$\chi^2 < \chi^2_{crit}$$

$\therefore$ Null Hypothesis accepted

The effect is significant on patients.

4] The following table gives the result of opinion pole for three parties A, B, C. Test whether the age and the choice of the party are independant at 5% level of significance using χ^2 test

Age	Party A	B	C	Total
20 - 35	25	20	25	70
35 - 50	20	25	35	80
above 50	25	25	30	80
Total	70	70	90	230

Setting of Hypothesis

Null Hypothesis $\Rightarrow$ age and choice has no association

Alternate Hypothesis $\Rightarrow$ age and choice has association

Test statistics:

For E $\rightarrow$

Age	Party A	B	C	Total
20-35	21	21	28	70
35 - 50	24	24	32	80
above 50	25	25	30	80
Total	70	70	90	230

$E_1 = \frac{70 \times 70}{230} = 21$, $E_2 = \frac{70 \times 70}{230} = 21$

$E_4 = \frac{80 \times 70}{230} = 24$, $E_5 = \frac{80 \times 70}{230} = 24$

O	E	$(O-E)^2$	$(O-E)^2/E$
25	21	16	0.76
20	21	1	0.04
25	28	9	0.32
20	24	16	0.67
25	24	1	0.04
35	32	9	0.28
25	25	0	0
25	25	0	0
30	30	0	0
			$\Sigma = 2.11$

$$\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right] = 2.11$$

LOS : $\alpha = 5\%$

Critical value : χ^2_{crit} $\to \alpha = 0.05$, $\to \nu = (r-1)(c-1) = 4$

$$\chi^2_{crit} = 9.488$$

Decision :

$$\chi^2 < \chi^2_{crit}$$

Null Hypothesis accepted

age and choice has no association.

5] In a survey of 200 boys of which 75 were intelligent, 40 has educated fathers, while 85 of the unintelligent boys had uneducated fathers. Do these figures support the ~~sen~~ hypothesis that educated fathers have intelligent boys.

		Fathers		
		Educated ~~Intelligent~~	uneducated	Total
Boys	Intelligent	40	35	75
	unintelligent	40	85	125
	Total	80	120	200

Setting of Hypothesis

Null Hypothesis ⇒ There is association

Alternate Hypothesis ⇒ There is no association

Test statistics:

For $E \rightarrow$

	Educated	uneducated	Total
Intelligent	30	45	75
Unintelligent	50	75	125
Total	80	120	200

$$E_1 = \frac{80 \times 75}{200} = 30$$

$$E_2 = \frac{120 \times 75}{200} = 45$$

O	E	$(O-E)^2$	$(O-E)^2/E$
40	30	100	3.33
35	45	100	2.22
40	50	100	2
85	75	100	1.33
			$\Sigma = 8.88$

$$\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right] = 8.88$$

LOS : $\alpha = 5\%$

Critical value : χ^2_{crit} $\alpha = 5\% = 0.05$, $\nu = (r-1)(c-1) = 1$

$$\chi^2_{crit} = 3.841$$

Decision:

$$\chi^2 > \chi^2_{crit}$$

Null Hypothesis rejected

There is association between education of father and intelligence of son.

II) Goodness of Fit $\Rightarrow$ checking data : corrections

Steps : 1] Setting of Hypothesis

Null Hypothesis (H_0) $\rightarrow$ given case

Alternate Hypothesis (H_A) $\rightarrow$ Opposite of given case

2] Test statistics :

$$\chi^2 = \Sigma \left[\frac{(O-E)^2}{E} \right]$$

For E $\rightarrow$ check note

3] LOS : $\alpha = 5\%$

4] Critical value : χ^2_{crit} $\rightarrow \alpha = \ldots$, $\rightarrow \nu = n-1$

5] Decision :

If $\chi^2 < \chi^2_{crit}$ ... Null Hypothesis accepted

If $\chi^2 > \chi^2_{crit}$ ... Alternate Hypothesis accepted

NOTE :- For E

- For regular case: $E = \dfrac{\text{total}}{\text{no. of data}}$
- For Proportions: $E = \dfrac{\text{given prop}^n \times \text{Total}}{\text{Total prop}^n}$
- For Poisson's: $E = \dfrac{N \cdot e^{-m} m^x}{x_1}$

where,

$\underset{\text{mean}}{m} = \dfrac{\Sigma f_i x_i}{\Sigma f_i}$

1] The following table gives the number of accidents in a city during whether the accidents are uniformly distributed over a week

Day	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Total
No. of accidents	13	15	9	11	12	10	14	84

$n = 7$

Setting of Hypothesis :

Null Hypothesis (H_0) $\Rightarrow$ accidents uniformly distributed

Alternate Hypothesis (H_A) $\Rightarrow$ accidents non-uniformly distributed

Test statistics :

$$E = \frac{\text{total}}{\text{no. of data}} = \frac{84}{7} = 12$$

O	E	$(O-E)^2$	$(O-E)^2/E$
13	12	1	0.08
15	12	9	0.75
9	12	9	0.75
11	12	1	0.08
12	12	0	0
10	12	4	0.33
14	12	4	0.33
			$\Sigma = 2.32$

$$\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right] = 2.32$$

LOS $= \alpha = 5\%$

Critical value $= \chi^2_{\alpha = 0.05,\ \nu = 6}$ $\quad \chi^2_{crit} = 12.592$

$\chi^2 < \chi^2_{crit}$

$\therefore$ Null Hypothesis accepted

Accidents are uniformly distributed.

2] A die was thrown 132 times and the following frequencies were observed

No. obtained	1	2	3	4	5	6	Total
Frequency	15	20	25	15	29	28	132

Setting of Hypothesis

Null Hypothesis (H_0) $\Rightarrow$ Die unbiased

Alternate Hypothesis (H_A) $\Rightarrow$ Die biased.

Setting of Hypothesis:

$E = \frac{\text{total}}{\text{no. of data}} = \frac{132}{6} = 22$

O	E	$(O-E)^2$	$(O-E)^2/E$
15	22	49	2.22
20	22	4	0.18
25	22	9	0.40
15	22	49	2.22
29	22	49	2.22
28	22	36	1.63
			$\Sigma = 8.87$

$$\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right] = 8.87$$

LOS : $\alpha = 5\%$

Critical value : χ^2_{crit} $\alpha = 0.05$, $\nu = 5$

$$\chi^2_{crit} = 11.070$$

Decision : $\chi^2 < \chi^2_{crit}$

Null Hypothesis accepted

The die is unbiased.

3] The number of car accidents in a metropolitian city was found to be 20, 17, 12, 6, 7, 15, 8, 5, 16 & 14 per month respectively. Use χ^2-test to check whether these frequencies are in agreement with the belief that occurence of accidents was the same during 10 Months period. Test at 5% level of significance

Setting of Hypothesis

Null ~~Alternate~~ Hypothesis (H_0) ⇒ accidents occur same no. of times

Alternate Hypothesis (H_A) ⇒ accidents occur different no. of times.

Test statistics

$$E = \frac{\text{Total}}{\text{no. of data}} = \frac{120}{10} = 12$$

O	E	$(O-E)^2$	$(O-E)^2/E$
20	12	64	5.33
17	12	25	2.08
12	12	0	0
6	12	36	3
7	12	25	2.08
15	12	9	0.75
8	12	16	1.33
5	12	49	4.08
16	12	16	1.33
14	120	4	0.33
			20.33

LOS = α = 5%

Critical value ⇒ χ^2_{crit} $\begin{cases} \alpha = 0.05 \\ \nu = 9 \end{cases}$ $\chi^2_{crit} = 16.919$

Decision: $\chi^2 > \chi^2_{crit}$

Null Hypothesis rejected

Accidents do not occur uniformly per month.

4] Theory predicts that the proportion of beans in the four groups A, B, C, D should be 9:3:3:1. In an experiment among 1600 beans the numbers in the four groups were 882, 313, 287 and 118. Does the experimental results support the theory?

Beans	Freq
A	882
B	313
C	287
D	118

$n = 4$

A , B , C , D

9 : 3 : 3 : 1

Setting of Hypothesis:

Null Hypothesis (H_0) $\Rightarrow$ expt should support theory

Alternate Hypothesis (H_a) $\Rightarrow$ expt doesn't support theory.

Test statistics

$$\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right]$$

for E $\rightarrow$ $E = \dfrac{\text{given proportion} \times \text{total}}{\text{Total proportion}}$

$$E_A = \frac{9}{16} \times 1600 = 900$$

$$E_B = \frac{3}{16} \times 1600 = 300$$

$$E_C = \frac{3}{16} \times 1600 = 300$$

$$E_D = \frac{1}{16} \times 1600 = 100$$

O	E	$(O-E)^2$	$(O-E)^2/E$
882	900	324	0.36
313	300	169	0.56
287	300	169	0.56
118	100	324	3.24
			$\Sigma = 4.726$

LOS $\Rightarrow$ $\alpha = 5\%$.

Critical value : χ^2_{crit} $\rightarrow \alpha = 0.05$, $\nu = 3$

$$\chi^2_{crit} = 7.815$$

$$\chi^2 < \chi^2_{crit}$$

Decision: Null Hypothesis accepted

Expt supports the theory.

5] In an experiment on pea breeding the following frequencies were obtained

Round & yellow Ry	Wrinkled & yellow ry	Round & Green RY	Wrinkled & Green rY	Total
315	110	108	32	556

Theory predicts that the frequencies should be in proportion of 9:3:3:1. Examine the correspondence between theory & expt using χ^2 test.

Setting of Hypothesis

Null Hypothesis (H_0) $\Rightarrow$ 9 : 3 : 3 : 1 is correct proportion

Alternate Hypothesis (H_a) $\Rightarrow$ 9 : 3 : 3 : 1 is incorrect proportion

Test statistics:

For E $\rightarrow$ $E = \frac{\text{given proportion}}{\text{Total proportion}} \times \text{Total}$

$$E_{Ry} = \frac{9}{16} \times 556 = 312.75$$

$$E_{ry} = \frac{3}{16} \times 556 = 104.25$$

$$E_{RY} = \frac{3}{16} \times 556 = 104.25$$

$$E_{rY} = \frac{1}{16} \times 556 = 34.75$$

O	E	$(O-E)^2$	$(O-E)^2/E$
315	312.75	5.0625	0.016
110	104.25	33.0625	0.317
108	104.25	14.0625	0.134
32	34.75	7.5625	0.217
			$\Sigma = 0.6797$

LOS = α = 5%.

$$\chi^2_{crit} \begin{cases} \alpha = 0.05 \\ \nu = 3 \end{cases}$$

$$\chi^2_{crit} = 7.815$$

Decision: $\chi^2 < \chi^2_{crit}$

Null Hypothesis accepted

The proportions are correct.

7) The number of defects in printed circuit board is hypothesized to follow Poisson distribution. A random sample of 60 printed boards showed the following data

Number of defects	0	1	2	3	Total
Observed frequency :	32	15	9	4	60

Does the Hypothesis of Poisson distribution seen appropriate

Setting of Hypothesis

Null Hypothesis (H_0) => Poison's distribution is appropriate

Alternate Hypothesis (H_A) => Poisson's distribution is not appropriate

Test statistics :

$$\chi^2 = \Sigma \left[\frac{(O-E)^2}{E} \right]$$

For E →

No. of data

$$E = \frac{N e^{-m} m^x}{x!}$$

$$\text{getting } m = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{(32)(0) + (15)(1) + (9)(2) + (3)(4)}{60}$$

$$= 0.75$$

15
18
12
~~45~~ 43
~~60~~ 124

$$E = \frac{(60)\ e^{-0.75}\ (0.75)^x}{x!}$$

For $x = 0$ $E = 28.34$

For $x = 1$ $E = 21.25$

For $x = 2$ $E = 7.97$

For $x = 3$ $E = 1.99$

O	E	$(O-E)^2$	$(O-E)^2/E$
32	28.34	13.39	0.47
15	21.25	39.06	1.83
9	7.97	1.06	0.132
4	1.99	4.04	2.03
			$\Sigma = 4.46$

LOS : $\alpha = 5\%$.

$\chi^2_{crit} \rightarrow \left[\begin{array}{l} \alpha = 0.05 \\ \gamma = 3 \end{array} \right.$

$\chi^2_{crit} = 7.815$

Decision:

$\chi^2 < \chi^2_{crit}$

Null Hypothesis accepted

Poisson's distribution is appropriate.