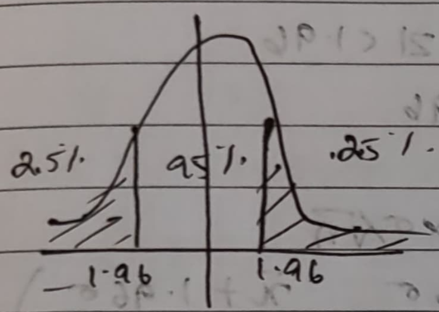
LARGE SAMPLING:

If $\bar{x}$ is the mean of each sample of size n drawn from the population with mean μ and std deviation σ then, $\bar{x}$ is normally distributed with mean μ and std deviation $\frac{\sigma}{\sqrt{n}}$ $\Rightarrow \bar{x}_n N(\mu, \sigma/\sqrt{n})$

$\therefore$ S.N.V $= \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ with mean 0 & S.d = 1



95% $\Rightarrow$ 1.96 to -1.96
99% $\Rightarrow$ 2.57 to -2.57

we know that

$P(z > 1.96 \text{ or } z < -1.96) = 0.05$

i.e $P(|z| > 1.96) = 0.05$

$P(|z| > 2.58) = 0.01$

$P(|z| > 3) = 0.0027$

The levels marked by probabilities 0.05 or 0.01 which decide the significance of an event and are called level of significance and are normally expressed by percentage as 5% level of significance or 1% level of significance (L.O.S)

* **Critical Region**:

→ The limits within which we expect z to lie with specific prob are called confidence limits. Thus, for $P(|z| > 1.96) = 0.05$, ± 1.96 are called confidence limits.

It means we are confident that in 95 cases out of 100, sample mean $\bar{x}$ will be such that Z lies between ± 1.96.

→ The region beyond the confidence level are called critical region.

* <u>Interval estimation to define the population mean 'μ'.</u>

we know that 5% LOS, $|z| < 1.96$

$$\Rightarrow \left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right| < 1.96$$

$$\Rightarrow |\bar{x}-\mu| < 1.96 \times \sigma/\sqrt{n}$$

$$\Rightarrow \mu \in \left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

* $|x-y| < \delta \Rightarrow y \in (x-\delta, x+\delta)$

Q Measurements of weights of random sample of 200 ball bearings ~~should~~ showed the mean of 0.824 N and a std deviation of 0.042 N. Find 95% confidence limits for the mean weight of all ball bearings.

ans: $n = 200$

$\bar{x} = 0.824$

$s = 0.042$

we know that 95% confidence level, μ lies between:

$$\mu \in \left(\bar{x} - 1.96\frac{s}{\sqrt{n}}, \bar{x} + 1.96\frac{s}{\sqrt{n}}\right) \rightarrow \sigma \frac{n}{x}$$

i.e., $\mu \in \left(0.824 - \frac{1.96 \times 0.042}{\sqrt{200}}, 0.824 + \frac{1.96 \times 0.042}{\sqrt{200}}\right)$

$\therefore \mu \in (0.818, 0.829)$

Q A Random sample of size 65 was drawn in the process of estimating mean annual income of 95 families. Mean & S.D were found to be:

$\bar{x}$ = Rs 4730

s = Rs 765

Find 95% confidence interval for the population mean

ans:

$\bar{x} = 4730$

$s = 765$

$n = 65$

$N = 950$

Since the sample is drawn from a ~~large~~ finite population and $\frac{n}{N} = \frac{65}{950} = 0.068$ which is greater than 0.05.

then we find finite population multiplier: $\sqrt{\frac{N-n}{N-1}}$

$\therefore \hat{\sigma}_{\bar{x}} = \sqrt{\frac{N-n}{N-1}} \times \frac{s}{\sqrt{n}}$

$= \sqrt{\frac{950-65}{949}} \times \frac{765}{\sqrt{65}}$

$= 91.63$

$\therefore \mu \in (\bar{x} - 1.96\,\hat{\sigma}_{\bar{x}},\ \bar{x} + 1.96\,\hat{\sigma}_{\bar{x}})$

i.e $\mu \in (4550.41,\ 4909.59)$

* Large sampling: (Z-test) $\Rightarrow n \geq 30$

Small sampling: (T-test)

* Testing the Hypothesis:

Q Random sample of 50 items gives mean 6.2, variance 10.24. Can it be regarded as drawn from normal population with mean 5.4 at 5% LOS.

$\mu = 5.4$
$\bar{x} = 6.2$
$s^2 = 10.24$
$n = 50$

null Hypothesis : (H_0) : $\mu = 5.4$
Alternate hypothesis : (H_a) : $\mu \neq 5.4$

$$|z| = \left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right| = \left|\frac{6.2-5.4}{10.24/\sqrt{50}}\right|$$

$= 1.767 = Z_{cal}$
given; LOS = 5% = 0.05
& value of Z_α = Table value of Z at 5% LOS
$= 1.96$
Since, $Z_{cal} < Z_\alpha$,
we accept null hypothesis (H_0)

Q A Random sample of 400 members is found to have mean of 4.45 cm can it be regarded as sample from population whose mean is 5 cm & variance is 4 cm

ans: $n = 400$ $\quad H_0 : \mu = 5$
$\bar{x} = 4.45$ $\quad H_a : \mu \neq 5$
$\mu = 5$
$\sigma = \sqrt{4} = 2$ $\quad |z| = \left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right| = \left|\frac{4.45-5}{2/\sqrt{400}}\right|$
$= 5.5 = Z_{cal}$
given, LOS = 5% = 0.05
$\therefore$ value of Z_α = Table value of Z at 5% LOS = 1.96
$\because$ $Z_{cal} > Z_{table}$
we Reject H_0.

Multiple Populations :

→ If $\bar{x}_1$ and $\bar{x}_2$ are means of the samples drawn from population 1 and population 2 respectively with sample size n_1 and n_2. If ~~p~~ std deviation of population 1 & 2 are σ_1 and σ_2 then, $\bar{x}_1 - \bar{x}_2$ follows a normal distribution with mean $\mu_1 - \mu_2$ and std error :

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

note: (i) If $\sigma_1 = \sigma_2 = \sigma$ (known)

then, $S.E = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

(ii) If $\sigma_1 = \sigma_2 = \sigma$ (unknown)

then, $S.E = \sqrt{\frac{S_1^2}{n_2} + \frac{S_2^2}{n_1}}$

where, S_1 & S_2 are unbiased estimates of population std deviation.

(iii) If $\sigma_1 \neq \sigma_2$ (unknown).

then, $S.E = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

Interval Estimation :

$$\mu_1 - \mu_2 \in \left((\bar{x}_1 - \bar{x}_2) - 1.96\, S.E \;,\; (\bar{x}_1 - \bar{x}_2) + 1.96\, S.E \right)$$

Q Find 95% confidence limits for the difference between the means from the data given below:

	Size	mean	S.D.
S-I	400 (n_1)	124 ($\bar{x}_1$)	14 (s_1)
S-II	250 (n_2)	120 ($\bar{x}_2$)	12 (s_2)

Date

ans $S.E = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$= 1.0324$

at 5% LOS,

$\mu_1 - \mu_2 \in \left((\bar{x}_1 - \bar{x}_2) - 1.96\, S.E,\ (\bar{x}_1 - \bar{x}_2) + 1.96\, S.E\right)$

$\mu_1 - \mu_2 \in (1.9765,\ 6.0235)$

Testing Hypothesis:

Q Means of 2 samples of sizes 1000 and 2000 respectively are 67.50 and 68 can the samples be regarded as drawn from the sample population with std deviation 2.5

ans: $H_0 : \mu_1 = \mu_2$

$H_a : \mu_1 \neq \mu_2$

$S.E = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 0.0968$

$|z|_{cal} = \left|\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S.E}\right| = 5.165$

at 0.27% LOS

$z_\alpha = z_{table} = 3$

Since $z_{cal} > z_\alpha$ we Reject H_0

∴ 2 samples are drawn from diff populations.

If LOS is not given take it as 5%

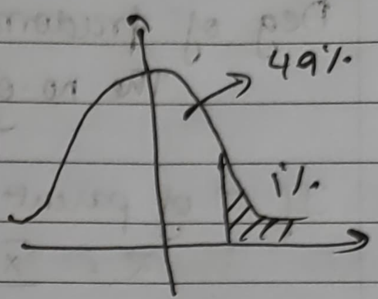
BARFI Page No.
Date

Q The average of marks scored by 36 boys is 72 with std deviation 8 while that of 36 girls is 70 with std deviation 6. Using one tail test at 1% LOS test whether boys performed better than girls.

ans: Ho: $\mu_1 = \mu_2$

Ha: $\mu_1 > \mu_2$

$S.E = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1.67$



$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{S.E} = 1.197$

at 1% LOS

$Z_\alpha = 2.33$

$\because Z_{cal} < Z_\alpha$

Hence, we accept Ho

$\therefore$ Boys do not ~~be~~ perform better than girls.

SMALL SAMPLING:

* If we take large no. of samples of size < 30 then we use Student's t-distribution.

*

	σ of popu known	σ of popu is not known
$n \geq 30$	Z-Test	Z Test
$n < 30$	Z-Test	t-Test

* T-statistics defined as:

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \quad \text{where} \quad S = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n-1}}$$

* Deg of freedom:
The no. of values we are free to choose = n − 1

① If σ of parent population is known,

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

② σ − unknown

then $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$, $s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$

or $t = \frac{\bar{x} - \mu}{\$/\sqrt{n-1}}$, $\$ = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$ = sample S.D

(unbiased estimate of popu S.D)

Interval Estimation for population mean 'μ'

Ex: Sample of size 10 has mean 40 and S.D 10. Construct 99% confidence interval for the population mean.

ans: $n = 10$
$\bar{x} = 40$
$\$ = 10$

$$\hat{\sigma}_{\bar{x}} = \frac{\$}{\sqrt{n-1}} = \frac{10}{\sqrt{9}} = 3.33$$

at 1% LOS, and $n-1 = 10-1 = 9$ deg of freedom
$t_\alpha = 3.25$

$\therefore \mu \in (\bar{x} - t_\alpha \hat{\sigma}_{\bar{x}}, \bar{x} + t_\alpha \hat{\sigma}_{\bar{x}})$
$= \mu \in (40 - 3.25 \times 3.33, 40 + 3.25 \times 3.33)$
$\mu \in (29.177, 50.822)$

Hypothesis Testing :

Q. Soap manufacturing company distributes soaps to a large no. of retail stores. Before the campaign mean sale per week per shop was 140/dozen. After the campaign a sample of 26 shops was taken and the mean sale was found to be 147/ dozen with S.D 16. Can you consider the advertisment affective.

ans:

$\mu = 140$ $\quad H_0 : \mu = 140$
$n = 26$ $\quad H_a : \mu > 140$
$\bar{x} = 147$
$s = 16$

$t_{cal} = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{147 - 140}{16/\sqrt{26}} = 2.187$

at 5% LOS, and $26 - 1 = 25$ degree of freedom
$= 1.708$

$\because t_{cal} > t_\alpha$
we reject H_0

Q 9 items of the sample has the following values:
45, 47, 50, 52, 48, 47, 49, 53, 51:
Does the mean of these 9 items differ significantly from the assumed pop mean $= 47.5$

ans: $\bar{x} = \frac{\Sigma x}{n} = 49.11$

$H_0: \mu = 47.5$

$H_a: \mu \neq 47.5$

$x_1 - \bar{x}$	$(x_1-\bar{x})^2$
-4.11	16.89
-2.11	4.45
0.89	0.792
2.89	8.35
-1.11	1.232
-2.11	4.45
-0.11	0.0121
3.89	15.1321
1.89	3.57

$s^2 = \sqrt{\frac{\Sigma(x_1-\bar{x})^2}{n}}$

$= \sqrt{\frac{54.889}{9}}$

$= 6.096$

$\therefore s = 2.469$

t cal $= \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$

$= 1.84$

t_α: 2.306

at 5% LOS and $n-1 = 9-1 = 8$ dof

$t_\alpha = 2.306$

$\because t_{cal} < t_\alpha$

$\therefore$ we accept H_0

<u>Testing diff between the means (2 pop):</u>

Case ①: Independent samples:

→ If the sample size $n_1 + n_2 - 2 < 30$ then, unbiased estimate of common pop S.D is given by:

$$s_p = \sqrt{\frac{\Sigma(x_{i1} - \bar{x}_1)^2 + \Sigma(x_{i2} - \bar{x}_2)^2}{n_1 + n_2 - 2}}$$

→ If unbiased estimate of 2 individual pop S.D deviations are :

$$s_1 = \sqrt{\frac{\Sigma(x_{i1}-\bar{x}_1)^2}{n_1-1}}, \quad s_2 = \sqrt{\frac{\Sigma(x_{i2}-\bar{x}_2)^2}{n_2-1}}$$

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

→ If S.D of the samples are,

$$\cancel{s}_1 = \sqrt{\frac{\Sigma(x_{i1}-\bar{x}_1)^2}{n_1}}, \quad \cancel{s}_2 = \sqrt{\frac{\Sigma(x_{i2}-\bar{x}_2)^2}{n_2}}$$

then $$S_p = \sqrt{\frac{\cancel{s}_1^2 n_1 + \cancel{s}_2^2 n_2}{n_1+n_2-2}}$$

and $$S.E = S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

& ∴ $$t = \frac{\bar{x}_1-\bar{x}_2}{S.E}$$

Q A sample of 8 students of 16 yrs shows mean systolic blood pressure of 118.4 mm of Hg with S.D of 12.17 ml. while a sample of 10 students of 17 yrs shows mean systolic blood pressure of 121 mm of Hg with S.D of 12.88 ml. The investigator feels that BP is related to age. Do you think the data provides enough reasons to support the investigators ~~claim~~ feeling ? At 5% LOS.

ans: $H_0: \mu_1 = \mu_2$ $\quad n_1 = 8 \quad n_2 = 10$

$H_a: \mu_1 \neq \mu_2$ $\quad \bar{x}_1 = 118.4 \quad \bar{x}_2 = 121$

$s_1 = 12.17 \quad s_2 = 12.88$

$$S_p = \sqrt{\frac{s_1^2 n_1 + s_2^2 n_2}{n_1 + n_2 - 2}} = \cancel{34.636}\ 13.33$$

$$S.E = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 6.32$$

$$t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{S.E} = -0.41$$

at 5% LOS & $n_1 + n_2 - 2 = 16$

$t_\alpha = 2.12$

$\because t_{cal} < t_\alpha$

we accept H_0

$\therefore$ BP does not depend upon age. Data does not provide enough info to support investegators feelings.

Q 6 Guinea pigs injected with 0.5 mg of medication took an average of 15.4 seconds on average to fall asleep with an unbiased standard deviation = 2.2 seconds. While 6 other guinea pigs injected with 1.5 mg of medication took an average of 11.2 s to fall asleep with an unbiased S.D - 2.6 s. Use 5% LOS to test that diff in the doses have no effect.

ans: $n_1 = 6$ $n_2 = 6$

$\bar{x}_1 = 15.4$ $\bar{x}_2 = 11.2$

$s_1 = 2.2$ $s_2 = 2.6$

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$.

$$S_p = \sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1+n_2-2}} = 2.408$$

$$S.E = S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 1.39$$

$$|t_{cal}| = \left|\frac{\bar{x}_1 - \bar{x}_2}{S.E}\right| = 3.02$$

at 5% LOS and 10 dof, $t_\alpha = 2.228$

$t_{cal} > t_\alpha$

∴ we reject H_0 ∴ the accept H_a

∴ The doses have effect.

Q Heights of 6 sailors in inches are:
63, 65, 68, 69, 71, 72. and the Heights of 10 soldiers:
61, 62, 65, 66, 69, 69, 70, 71, 72, 73. Discuss whether Soldiers on an avg are taller than sailors.

ans: $n_1 = 6$ $n_2 = 10$

$\bar{x}_1 = 68$ $\bar{x}_2 = 67.8$

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

29/3/23

$(x_1-\bar{x}_1)^2$	$(x_2-\bar{x}_2)^2$
25	46.24
9	33.64
0	7.84
1	3.24
9	1.44
16	1.44
	4.84
	10.24
	27.04

$S_p = 3.9$ (2).

$S.E = 3.9\sqrt{\frac{1}{6}+\frac{1}{10}} = 2.01$

$|t_{cal}| = \left|\frac{\bar{x}_1-\bar{x}_2}{SE}\right| = 0.099$

at 5% LOS & $n_1+n_2-2 = 14$ dof

$\therefore t_\alpha = 2.145$

$\because t_{cal} < t_\alpha$ we accept H_0

case ②: $\mu_1 = \mu_2 = \mu$

i.e $\mu = 0$

Here, we are going to test the effectiveness of the methods on the same sample.

Q Certain injections given to 12 patients resulted in change in BP. 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4. Can it be concluded that the injection in general will be accompanied by an increase in BP.

ans: $H_0: \mu = 0$ (no inc)

$H_a: \mu \neq 0$

$\bar{x} = 2.58$

$s^2 = \frac{\sum(x_i-\bar{x})^2}{n} = 8.74$

$t_{cal} = \frac{\bar{x}-\mu}{s/\sqrt{n-1}} = 2.89$

at 5% LOS & 11 dof, $t_\alpha = 2.201$

$\because$ $t_{cal} > t_\alpha$

we reject H_0

$\therefore$ the injection will be accompanied by inc in BP.

Q 10 boys were given a test in stats and the scores were recorded. They were given a months coaching and a 2nd test was given to them. Test if the marks given below give evidence to the fact that students are benefited by coaching.

T_1 :	70	68	56	75	80	90	68	75	56	58
T_2 :	~~70~~ 68	70	52	73	75	78	80	92	54	55

ans: x : $T_2 - T_1$: $\bar{x} = 0.1$

$-2 \quad +2 \quad -4 \quad -2 \quad -5 \quad -12 \quad 12 \quad 17 \quad -2 \quad -3$

(make table)

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n} = 67.90$$

H_0 : $\mu = 0$

H_a : $\mu \neq 0$

$$t_{cal} = \frac{\bar{x} - \mu}{S / \sqrt{n-1}} = 0.036$$

at 5% LOS & 9 dof $t_\alpha = 2.262$

$t_{cal} < t_\alpha$

$\therefore$ accept H_0.

$\therefore$ There was no effect of coaching.

χ^2 (chi) distribution:

→ Chi-square test is used to test whether there is an association between 2 or more attributes. It is also used to test a characteristic depended on another characteristic. ~~ex:~~ Using chi-square dist in this way, to test the independence of on attribute on another is called test of independence.

→ It is commonly known as ~~good~~ chi-square test of goodness of fit because it enables to a certain how well theoretical distribution like binomial, poisson or normal fit the observed frequencies.

→ It is also used to test whether the proportions $P_1, P_2 \cdots$ in different proportions are equal.

Conditions:

1. (The total no-of observations) N must be greater than 50 (large)

2. Frequency of cell should be greater than 5. If freq is less than or equal to 5 then combine that with the neighbouring freq so that the combined freq is greater than 5 and degree of freedom is reduced accordingly.

3. No. of cla

Yate's correction:

→ In a 2×2 table, degree of freedom is $(2-1)\times(2-1)$. If any of the frequency cell is less than 5, we must use pooling method but this method will result in chi-square in 0 dof which is meaningless.

→ Therefore we use $\chi^2 = \sum \frac{\{|O-E| - 0.5\}^2}{E}$

Type 1: Independence of attributes:

Q Investigate the association between darkness of eyecolour in father and son from the following data.

colour of father's eye

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

ans: H_0: There is no association between darkness of eye colour in father and son.

H_a: There is an association.

$\frac{x}{128} = \frac{138}{1000}$ $\therefore x = 18$

$138 - 18 = 120$

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

$$\chi^2_{cal} = \sum (O-E)^2/E$$

$$= 900\left(\frac{1}{18} + \frac{1}{120} + \frac{1}{110} + \frac{1}{752}\right)$$

$$= 66.87$$

at 5% LOS and $(2-1)(2-1)$ dof,

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

$\because$ $\chi^2_{cal} > \chi^2_{table}$ we reject H_0 & accept H_a

$\therefore$ There is an association between father's and son's eye colour.

Q 2 Batches of 12 animals are given test of inoculation. 1 Batch was inoculated and the other was not. The no of dead and survived animals are given in the table. Can the inoculation be regarded as affecting against the disease at 5% LOS.

ans:

	Dead	Survived	Total
Innoculated	2	10	12
not Inn	8	4	12
Total	10	14	24

→ (always)

Ho : no association

Ha : There is association

O	E	|O−E|−0.5	n^2	.../E
2	5	2.5	6.25	1.25
10	7	2.5	6.25	0.89
8	5	2.5	6.25	1.25
4	7	2.5	6.25	0.89

$$\chi^2_{cal} = \sum \frac{(|O-E|-0.5)^2}{E} = 4.28$$

at 5% LOS , (2−1)(2−1) dof,

$\chi^2_{table} = 3.81$

$\because$ $\chi^2_{cal} > \chi^2_{table}$

we reject Ho.

Type 2 : Goodness of fit :

Q The following table gives the no. of accidents during a week. Find whether the accidents are uniformly dist over the week.

Day: ~~Acc~~	Acc :	~~Total~~
Sun	13	
Mon	15	
Tue	9	
Wed	11	
Thurs	12	
Fri	10	
Sat	14	
Total :	84	

ans: H_0 : Accidents are uniformly dist over the week
H_a : not uniform.

If H_0 is true then there will be 84/7 that is, 12 accidents per week.

$$\therefore \chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{1}{12}\left(1^2+3^2+3^2+1^2+0^2+2^2+2^2\right)$$

$$= \frac{1}{12}(28) = 2.33$$

at 5% LOS & $n-1 = 7-1 = 6$ dof

$\chi^2_{table} = 12.59$

$\because \chi^2_{cal} < \chi^2_{table}$

$\therefore$ we accept H_0

Q Theory Predicts that the proportion of beads in 4 grps A, B, C, D should be 9 : 3 : 3 : 1. In an exp among 1600 beans the number in the 4 grps were 882, 313, 287 & 118. Does the exp result support the theory.

ans: H_0 : The propotion is 9:3:3:1
H_a : Propotion is not given as above.

No. of exp beans in 4 grps will be,

$\frac{9}{16} \times 1600 = 900$ $\quad$ $\frac{3}{16} \times 1600 = 300$

$\frac{3}{16} \times 1600 = 300$ $\quad$ $\frac{1}{16} \times 1600 = 100$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 4.72$$

χ^2 table $= 7.815$

$$x^2 = \sum \frac{(O-E)^2}{E} = 4.72$$

x^2 table $= 7.815$

Since x^2 table $>$ x^2 cal,
we accept H_0

Q Figures given below are

(a) observed frequencies of distribution
(b) Frequencies of a normal distribution having the same mean, S.D and the total freq as in (a).
⇒ Apply chi-square test for goodness of fit.

freq of (a):

	O	E	$(O-E)^2$	$(O-E)^2/E$
13	1, 12	2, 15 (17)	16	0.941
	66	66	0	0
	220	210	100	0.476
	495	484	121	0.25
	792	799	49	0.061
	924	943	361	0.382
	792	799	49	0.061
	495	484	121	0.25
	220	210	100	0.476
	66	66	0	0
13	12, 1	15, 2 (17)	16	0.941

$$\therefore x^2_{cal} = \sum \frac{(O-E)^2}{E} = 3.84$$

at 5% Los and $13-2-3 = 8$ dof (3 constraints)
x^2 table $= 15.51$

$\because \chi^2_{cal} < \chi^2_{table}$, Ho is accepted.

Q Following mistakes per page were observed in a book:

x_i	No. of mist./page:	0	1	2	3	4
f_i	No. of pages:	211	90	19	5	0

Fit a poisson distribution

ans: H_0: mistakes follow the poisson's dist (fit is good)
Ha: fit is not good.

$$m = \frac{\sum f_i x_i}{\sum f_i} = 0.44$$

$$\therefore P(X=x) = \frac{e^{-m} m^x}{x!} = \frac{e^{-0.44}(0.44)^x}{x!}$$

$$\therefore \text{Exp freq} = NP(X=x)$$
$$= 325 \times \frac{e^{-0.44}(0.44)^x}{x!} \quad ; x = 0, 1, 2, 3, 4.$$

O	E	$(O-E)^2/E$
211	209	0.019
90	92	0.043
19	20	~~0.05~~
5	3	~~1.33~~ 0
0	1	~~4~~
24	24	
325	325	0.062

$\therefore \chi^2_{cal} = 0.062$

at 5% los & 5-2-2 = 1 dof
$\chi^2_{table} = 3.84$

$\because \chi^2_{cal} < \chi^2_{table}$,
Ho is accepted.

BARFI Page No.
Date

Hypothesis concerning several propotions :

Q Sample of 3 shipments . A, B, C of defectives items gave following results.

	A	B	C	Total
Def	5 (x) = 6	8 (y) = 7	9 (9)	22
non-def	35 (34)	42 (43)	51 (51)	128
Total	40	50	60	150

Test whether propotion of defective items is same in 3 shipments at 5% LOS.

ans: $H_0: P_1 = P_2 = P_3$
$H_a: P_1 \neq P_2 \neq P_3$ } always

$\frac{x}{40} = \frac{22}{150}$ $\therefore x = 6$

O	E	$(O-E)^2/E$
5	6	0.166
8	7	0.142
9	9	0
35	34	0.029
42	43	0.023
51	51	0
		0.36

$\chi^2_{cal} = 0.36$

at 5% los, $(2-1)(3-1) = 2$ dof

$\chi^2_{table} = 5.99$

$\because \chi^2_{cal} < \chi^2_{table}$,

we accept H_0.

Q 5 dice were thrown 192 times and no. of times 4, 5 & 6 obtained are as follows:

no. of dice showing 4, 5, 6 :	5	4	3	2	1	0
Freq	6	46	70	48	20	2

Calculate chi-square

ans:

use Binomial to calculate exp freq = $^nC_x p^x q^{n-x}$

$p = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1/2 = 0.5$

O	E	$(O-E)^2/E$
6	6	0
46	30	8.53
70	60	1.66
48	60	2.4
22 { 20, 2	36 { 30, 6	5.44
192	192	18.03

$\chi^2 cal = 18.03$

~~at 5% los and 6.1~~