UNIT 4+1 : Hypothesis Testing

1) To make estimates about population parameters (unit 4)

2) To make conclusions / decisions about population parameters

$\mu = 30$ or $\mu \neq 30$

Hypothesis : Assumption / claim about the population parameters.

- NULL hypothesis : (H_0) There is no significant difference between parameter and statistic.

- Alternative Hypothesis : (H_1 or H_a) complement of null hypothesis.

$H_1 : \mu \neq \mu_0 \rightarrow$ Two tail

(i) $H_1 : \mu < \mu_0 \rightarrow$ Left tail

(ii) $H_1 : \mu > \mu_0 \rightarrow$ Right tail

Four tests :-

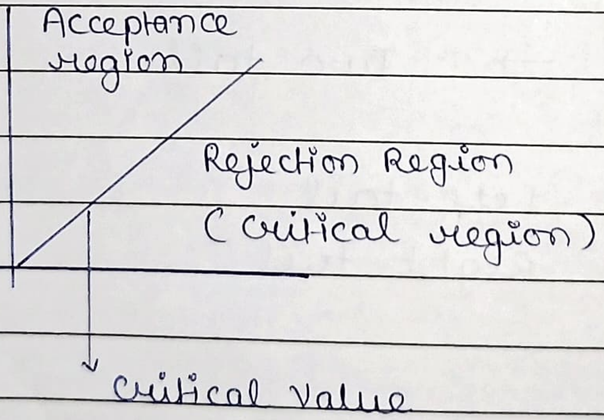
- Student's t - test
- Z - test
- F - test
- χ^2 - test (Chi - square test)

Errors in sampling -

		Decision: Reject H_0	Decision: Accept H_0
True Statement	H_0 is true	Type I error	Correct
	H_0 is false	Correct	Type II error

- $P(\text{Type I error}) = \alpha$
- $P(\text{Type II error}) = \beta$

Critical Region -



Level of significance α

Confidence level $1-\alpha$

Student's t-test for single mean ($n < 30$)

1. Define $H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0 \rightarrow$ Two tailed

or $H_1 : \mu > \mu_0 \rightarrow$ Right tailed

or $H_1 : \mu < \mu_0 \rightarrow$ left tailed

(Right tailed and left tailed: one tailed)

2. Test statistics -

Under H_0,

$t = \dfrac{\bar{x} - \mu}{S/\sqrt{n}}$ follows a t-distribution with $(n-1)$ degree of freedom.

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

$$\bar{x} = \frac{\sum x_i}{n}$$

$$n s^2 = (n-1) S^2$$

$$\frac{S^2}{n} = \frac{s^2}{(n-1)}$$

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

3. Critical Region -

If calculated $|t| >$ tabular t $\rightarrow$ Reject H_0

If calculated $|t| <$ tabulated t $\rightarrow$ Accept H_0.

4. Conclusion.

Q. The mean weekly sales of Soap bars in departmental store is 146.3. After an advertising campaign, the mean weekly sales in 22 stores increase to 153.7 with standard deviation of 17.2. Was the advertising compaign is suffessful test at 5 level of significance.

Sol.

$n = 22$

$\bar{x} = 153.7$

$s = 17.2$

$H_0 : \mu = 146.3$

$H_1 : \mu > 146.3$ (one tailed)

Under H_0,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} \sim t_{21}$$

$$t = \frac{153.7 - 146.3}{17.2/\sqrt{21}} = 1.97$$

from the table -

$t_{0.05}$ for 21 d.f $= 1.721$

Since, calculated $t >$ tabulated t

$\Rightarrow H_0$ is rejected

We conclude the adv. compaign was successful.

Q. A random sample of 16 values from a population of with mean of ~~μ~~ = has mean μ = 41.5 and sum of the squares of deviation from this mean is ~~435~~ 135 Show that a stand the assumption of mean of 43.5 for the population is not reasonable obtain 95% and 99% confidence limit for the same.

Given - $\nu = 15$, $t = 2.131$ when $p = 0.05$

$t = 2.947$, $p = 0.01$

Sol. $n = 16$

~~μ~~ $\bar{x} = 41.5$

$\mu = 43.5$

$\Sigma(x_i - \bar{x})^2 = 135$

$$S^2 = \frac{1}{n-1}\Sigma(x_i-\bar{x})^2 = \frac{135}{15} = 9$$

H_1 : $\mu \neq 43.5$ (two tailed)

H_0 : $\mu = 43.5$

Under H_0, the test statistic is

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} \sim t_{15}$$

$$t = \frac{44.5 - 43.5}{3/4} = -2.667$$

5% level of significance -

$t_{0.05} = 2.131$
$|t| = 2.667$

$\Rightarrow$ H_0 is rejected.

we conclude that the mean of population is 43.5 is not reasonable.

At 1% level of significance -

we accept H_0 as $|t| > t_{0.01}$

we conclude that the mean 43.5 of population is reasonable.

Confidence limits of μ

$$\bar{x} \pm \frac{s}{\sqrt{n}} t_\alpha$$

95% confidence limits -

$$41.5 + \frac{3}{4}(2.131)$$

$$39.902 < \mu < 43.098$$

95% confidence interval = (39.902, 43.098)

99% confidence limits of μ

$$41.5 \pm \frac{3}{4}(2.947)$$

$$39.29 < \mu < 43.71$$

99% confidence interval -
$(39.29, 43.71)$

Q. The heights of 10 males of a given locality are 70, 67, 62, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that average height is greater than 64 inches test at 5 level significance

Q. Solution - $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ $\mu = 64$

$\bar{x} = 66$ $s^2 = \dfrac{1}{n-1}\left[\Sigma(x_i - \bar{x})^2\right]$

$t = 2$

Assumptions for student's t - test for single mean -

1) The population from which the sample is drawn is normal.

2) The sample is random

3) The population the standard deviation σ is unknown.

Assumptions for students t test for difference of means -

1) The populations from which the samples have been drawn are normal.

2) The samples are random and independent of each other.

3) The population variances are equal and are unknown.

$$\sigma_x^2 = \sigma_y^2 = \sigma^2 \text{ (say)},$$

where σ^2 is unknown.

Let x_i ($i = 1, 2, 3 \ldots n_1$) and y_i ($j = 1, 2, 3 \ldots n_2$) be two independent samples of sizes n_1 and n_2, respectively. drawn from two normal populations with means μ_x and μ_y resp.

① $H_0 : \mu_x - \mu_y = d$

$H_1 : \mu_x - \mu_y \neq d$ or $\mu_x - \mu_y < d$ or $\mu_x - \mu_y > d$

② Under H_0,

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

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

$$s_1^2 = \frac{1}{n_1}\Sigma(x_i - \bar{x})^2$$

$$s_2^2 = \frac{1}{n_2}\Sigma(y_i - \bar{y})^2$$

than,

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

t follows student's t - distribution with $n_1 + n_2 - 2$ d.f

Calculated $|t| <$ tabulated t $\Rightarrow$ Accept H_0,

Q. Samples of two types of electric bulbs were tested for length of life and the following data was obtained

	Type I	Type II
Sample size	$n_1 = 8$	$n_2 = 7$
Sample means	$\bar{x}_1 = 1234$	$\bar{x}_2 = 1036$
Sample SD's	$s_1 = 36$	$s_2 = 40$

Is the difference in the means sufficient to say that Type I is superior to Type II regarding the length of life.

The value of t at 5% level of significance for $\nu = 13$, $t = 1.771$

$\nu = 14$, $t = 1.761$

$\nu = 15$, $t = 1.753$

Sol.

$H_0 : \mu_1 - \mu_2 = 0$ i.e Type I is not superior to Typ II

$H_1 : \mu_1 > \mu_2$ or $H_1 : \mu_1 - \mu_2 = 0$ ⇒ Type I is superior to type II.

Under H_0,

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2} = t_{13}$$

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

$$= \frac{1}{13}\left[(8)(36)^2 + (7)(40)^2 \right]$$

$$= 1659.08$$

$$\therefore t = \frac{(1234 - 1036)}{\sqrt{165.08\left(\frac{1}{8}+\frac{1}{7}\right)}} = 9.39$$

$t_{0.05}$ for 13 d.f $= 1.771$

Calculated t > tabulated t

=> Reject H_0

we conclude that type I is superior to type II.

Q. The heights of 6 randomly chosen sellers are 63, 65, 68, 69, 71, 72 and the heights of 10 randomly chosen soldiers are 61, 62, 65, 66, 69, 69, 70, 71, 72, 73.

Discuss the suggestion that the sellers are on average are taller than soldiers.

Solution of previous question -

x	y
63	61
65	62
68	65
69	66
71	69
72	69
	70
	71
	72
	73

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$H_0 : \mu_x = \mu_y$ or $\mu_x - \mu_y = 0$

$H_1 : \mu_x > \mu_y$ or $\mu_x - \mu_y > 0$

$$\bar{x} = \frac{408}{6} = 68$$

$$\bar{y} = \frac{678}{10} = 67.8$$

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

$(x_i - \bar{x})^2$	$(y_j - \bar{y})^2$
25	
9	
0	
1	
9	
16	
60	

$\Sigma (y_j - \bar{y})^2 = 151.81$

$$s = \frac{1}{14}[60 + 151.86]$$

$$s^2 = 15.1328$$

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{68 - 67.8 - 0}{\sqrt{15.1328\left(\frac{1}{6} + \frac{1}{10}\right)}}$$

$$t = 0.099$$

$t_{0.05}$ for 14 d.f $= 1.76$

Since, calculated value lies in acceptance region, we accept H_0 at $\alpha = 5\%$.

Paired t-test for difference of means -

If $n_1 = n_2 = n$ and two samples are not independent but paired together.

$$d_i = x_i - y_i$$

$$t = \frac{\bar{d}}{s/\sqrt{n}} \sim t_{n-1},$$

where $s^2 = \frac{1}{n-1} \Sigma (d_i - \bar{d})^2$

Q. In a certain experiment, to compare two types of animal food. the following results of increase in weight were obtain.

Food A	Food B	d_i	$(d_i - \bar{d})^2$
49	52	-3	1
53	55	-2	0
51	52	-1	1
52	53	-1	1
47	50	-3	1
50	54	-4	4
52	54	-2	0
53	53	0	4
		-16	12

Assuming that the same set of animals were used in both foods. can we conclude that food B is better than food A?

Sol. $H_1 : \mu_x < \mu_y$ or $\mu_x - \mu_y < 0$

$H_0 : \mu_x = \mu_y$

Under H_0,

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

$$\bar{d} = \frac{\Sigma d_i}{n} = \frac{-16}{8} = -2$$

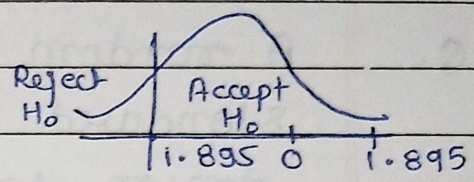
$$s^2 = \frac{1}{7}(12) = 1.714$$

$$t = \frac{-2}{\sqrt{\frac{1.714}{8}}} = -4.34$$

$t_{0.05}$ for 7 df $= 1.895$

H_1 is true

$\therefore$ food B is better than food A.



Z-test for single mean ($n \geq 30$)

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad (\text{If } \sigma \text{ is known})$$

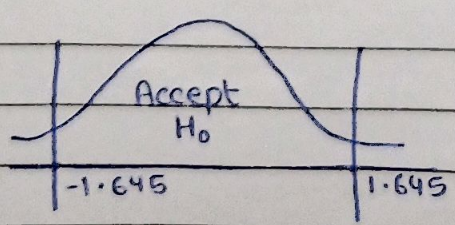
If σ is unknown for large samples,

$$s^2 = \sigma^2$$

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim N(0,1)$$

(If σ is unknown)

Z_α	1%	5%	10%
two tailed	2.58	1.96	1.645
one tailed	2.33	1.645	1.28



Confidence limits / Fiducial limits -

$$\bar{x} - \frac{\sigma}{\sqrt{n}} z_\alpha < \mu < \bar{x} + \frac{\sigma}{\sqrt{n}} z_\alpha$$

Q. A random sample of 900 has a mean 3.4 and standard deviation 2.61 is the sample drawn from a large population of mean 3.25. If the population is normal than 95% and 98% confidence level ~~mean~~ of ~~two~~ true mean.

Solution of previous question -

$n = 900$, $\bar{x} = 3.4$, $s = 2.61$

$H_0 = \mu = 3.25$

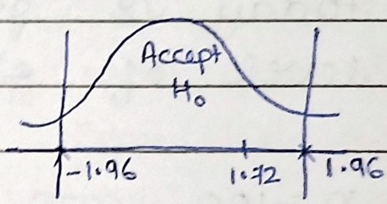
$H_1 = \mu \neq 3.25$ (two tailed)

Under Ho,

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{3.4 - 3.25}{261/\sqrt{900}}$$

$$z = 1.72$$

$$z_{0.05} = 1.96$$



since $z < z_{0.05}$,

we accept Ho at 5% level of significance and we conclude that sample has drawn from large population with $\mu = 3.25$.

95% confidence limits

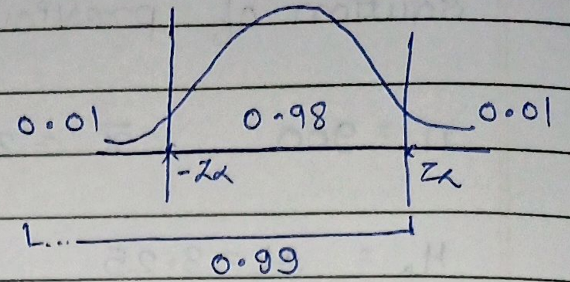
$$\bar{x} - \frac{s}{\sqrt{n}} z_{0.05} < \mu < \bar{x} + \frac{s}{\sqrt{n}} z_{0.05}$$

$$34 - \frac{261}{\sqrt{900}}(1.96) < \mu < 3.4 + \frac{2.61}{\sqrt{900}}(1.96)$$

$$3.23 < \mu < 3.57$$

$\longrightarrow$ (3.23 , 3.57)

98 % confidence limits -



$P(Z < Z_\alpha) = 0.99$

Observe from the table,
Z_α will be $= 2.33$

Q. A random sample of 100 recorded dealth the past year has an average life span of 71.8 years assuming a population standard deviation 8.9 years. Does this seems to appear indicate that the mean life spam today is greater than 70 years use a 5% level of significance.

Sol. $n = 100$ $\qquad H_0 : \mu = 70$

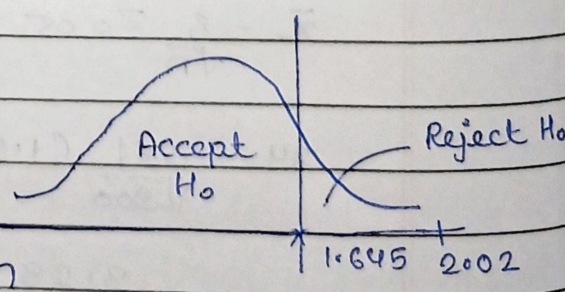
$\bar{x} = 71.8$ $\qquad H_1 : \mu > 70$

$\sigma = 8.9$

Under H_0,

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{71.8 - 70}{8.9/10} = \frac{1.8}{0.89} = 2.02$$

$Z_{0.05} = 1.645$



H_1 is true so,
we conclude that average life span is greater than 70 years.

Z-test for difference of means -

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

if σ_1 and σ_2 are known.

- If σ_1 and σ_2 are unknown -

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Q. The means of 2 samples of 1000 and 2000 are 67.5 and 68 respectively. Can the sample be regarded as drawn from the same population of standard deviation of 2.5

Sol.

$H_0 : \mu_1 = \mu_2$ and $\sigma = 2.5$

$H_1 : \mu_1 \neq \mu_2$

Under H_0,

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{67.5 - 68 - 0}{\sqrt{\frac{(2.5)^2}{1000} + \frac{(2.5)^2}{2000}}} = -5.165$$

we reject H_0, means H_1 is accepted
we conclude that two samples are not drawn from same population.

Q In a survey, 400 women shoppers are chosen at random in super market A, their avg. weekly expenditure is ₹250 with standard deviation of ₹40. For 400 women shopper chosen in market B, the avg. expenditure is ₹220 and standard deviation ₹55 test at 1% level of significane weather the average weekly expenditure of two population are equal.

Sol.

A	B
$n_1 = 400$	$n_2 = 400$
$\bar{x}_1 = 250$	$\bar{x}_2 = 220$
$s_1 = 40$	$s_2 = 55$

F - test for equality of population variances -

$H_0 : \sigma_1^2 = \sigma_2^2$

Under H_0,

$$F = \frac{S_1^2}{S_2^2} \quad \text{if} \quad S_1^2 > S_2^2$$

or

$$F = \frac{S_2^2}{S_1^2} \quad \text{if} \quad S_2^2 > S_1^2$$

$$S_1^2 = \frac{1}{(n_1-1)} \Sigma (x_i - \bar{x})^2$$

$$S_2^2 = \frac{1}{(n_2-1)} \Sigma (y_i - \bar{y})^2$$

$F = \frac{S_1^2}{S_2^2}$ follows a F - distribution

with d.f (ν_1, ν_2), where $\nu_1 = n_1 - 1$

$\nu_2 = n_2 - 1$

Q. Two random samples gave the following results -

	Size	Sample Mean	Sum of sq. of deviation from mean
1.	10	15	90
2.	12	14	108

Test whether the sample comes from same normal population at 5% level of significance.

$F_{0.05}(9,11) = 2.90$, $F_{0.05}(11,9) = 3.10$, $t_{0.05}(20) = 2.086$

$t_{0.05}(22) = 2.07$

Sol. $H_0 : \mu_1 = \mu_2$, $\sigma_1^2 = \sigma_2^2$

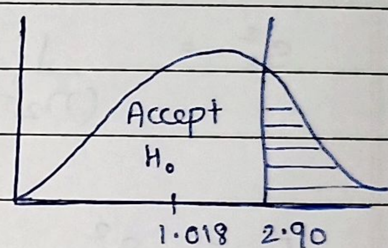
$$s_1^2 = \frac{1}{9}(90) = 10$$

$$s_2^2 = \frac{1}{11}(108) = 9.82$$

Under Ho,

$$F = \frac{s_1^2}{s_2^2} = \frac{10}{9.82} = 1.018$$

$\Rightarrow \quad \sigma_1^2 = \sigma_2^2$



$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$t = 0.742$

Accept Ho, $\mu_1 = \mu_2$

→ Both the sample are from same population.

Q. Pumpkins were grown under two experimental conditions. Two random samples of 11 and 9 show the sample deviation of 0.8 and 0.5 respectively. Assuming that the distributions are normal. Test the hypothesis that the ~~two~~ true variances are equal against the alternative that they are not at 10 % level, assume that probability $P(F_{10,8} \geq 3.25) = 0.05$

$P(F_{8,10} \geq 3.07) = 0.05$

Sol. $H_0 : \sigma_1^2 = \sigma_2^2$ $\qquad n_1 = 11 \qquad s_1 = 0.8$

$H_1 : \sigma_1^2 \neq \sigma_2^2$ $\qquad n_2 = 9 \qquad s_2 = 0.5$

$$S_1^2 = \frac{1}{(n_1 - 1)} \Sigma (x_i - \bar{x})^2$$

$$= \frac{1}{10} n_1 s_1^2$$

$$= \frac{1}{10} (11)(0.8)^2 = 0.704$$

$$S_2^2 = \frac{1}{(n_2 - 1)} n_2 s_2^2$$

$$= \frac{1}{8} (9)(0.5)^2 = 0.28125$$

Under Ho,

$$F = \frac{S_1^2}{S_2^2} = \frac{0.704}{0.28125} = 2.5$$

$$F_\alpha (v_1, v_2) = \frac{1}{F_{1-\alpha}(v_2, v_1)}$$

$$F_{0.05}(8,10) = 3.07$$

$$\frac{1}{F_{0.05}(8,10)} = \frac{1}{3.07} = 0.33$$

100

90 $\quad$ 10%

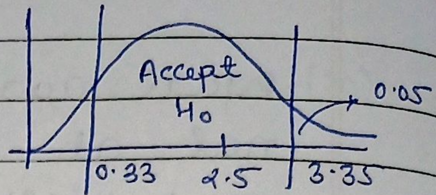
5% $\quad$ 95%

Expected frequency

classmate
Date
Page

Formula -

$E = Np$



H_0 Accepted, we conclude that true variances are equal.

Chi - square (χ^2) test for goodness of fit - (Non- parametric test / Distribution free)

If f_i ($i = 1, 2 \dots n$) is a set of observed or experimental frequencies and e_i ($i = 1, 2 \dots n$) is corresponding set of expected or theoretical frequencies then χ^2 will be -

f_i ($i = 1, 2 \dots n$) $\longrightarrow$ observed / experimental

e_i ($i = 1, 2 \dots n$) $\longrightarrow$ expected / theoretical

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - e_i)^2}{e_i}, \quad \Sigma f_i = \Sigma e_i$$

follows a chi - square distribution with $n-1$ degree of freedom.

Conditions for Validity -

(i) Sample observations should be independent

(ii) Sum of observed frequency is same as sum

$$\Sigma f_i = \Sigma e_i$$

(iii) The Total frequency (N) should be greater than 50 and no e_i should be less than 5.

Q. A sample analysis of examination results of 200 students was made. 68 student secured 3rd division, 2nd division – 68, failed 46 and rest are place in 1st division. Are these figures agree with general examination result which is in the ratio 4 : 3 : 2 : 1 for various categories respectively. Given the $\chi^2_{0.05}$ for 3 d.f = 7.815

Sol.

Observed frequency	f_i	e_i	$(f_i - e_i)^2$	–
failed	46	$200 \times \frac{4}{10} = 80$	1156	$\frac{1156}{e_i}$
3rd devision	68	$\frac{3}{10} \times 200 = 60$	64	$\frac{64}{68}$
2nd division	62	40	484	$\frac{484}{40}$
1st division	24	20	16	$\frac{16}{20}$
	200	200		

Expected frequency –

$$N = 200$$

$$E = 200 \times p$$

$$\Sigma e_i = \Sigma f_i$$

$$\chi^2 = \Sigma \left(\frac{(f_i - e_i)^2}{e_i} \right) = 28.417$$

$$\chi^2 = 28.417$$

Since $\chi^2 > \chi^2_{0.05}$ for 3 d.f, we reject H_0

We conclude that general result does not match

Q. A survey of 800 families with 4 children each revealed the following distribution

No. of boys	No. of girls	No. of families (f_i)
0	4	32
1	3	178
2	2	290
3	1	236
4	0	64
		800

Is this result consistent with the hypothesis that male and female births are equally probable

Sol. $p = \frac{1}{2} = q$

Binomial distribution -

$f(x) = {}^4C_x \, p^x q^{4-x}, \quad x = 0, 1, 2, 3, 4$

$f(x) = {}^4C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{4-x}, \quad x = 0, 1, 2, 3,$

$= {}^4C_x \left(\frac{1}{2}\right)^4, \quad x = 0, 1, 2, 3$

$F = 800 \times {}^4C_x \left(\frac{1}{2}\right), \quad x = 0, 1, 2, 3, 4$

$F = 50 \times {}^4C_x, \quad x = 0, 1, 2, 3, 4$

e_i values	$(f_i - e_i)^2$	$(f_i-e_i)^2/e_i$
50	324	6.48
200	484	2.42
300	100	0.33
200	1296	6.48
50	189	3.92
800		19.63

$\chi^2 = 19.63$

$d.f = 4$

$\chi^2_{0.05}$ for 4 d.f $= 9.488$

→ Reject H_0, accept H_1

Q. When the first proof 392 pages of a book were read. the distribution of printing mistakes were found to be as follows -

No. of mistakes in a page (x)	No. of pages (f)	xf	e_i
0	275	0	242.1
1	72	72	116.69
2	30	60	28.12
3	7	21	4.518
4	5	20	0.544
5	2	10	0.052
6	1	6	0.004
	392	189	

15   5.118

Fit a poisson destribution to the above that and test the godness of test.

Sol. We will assume

In order to fit poisson distribution to the given data, we take the mean of PD equal equal to the mean of the given distribution.

$$\bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{189}{392} = 0.482$$

The frequency of r mistakes per page is -

$$f(r) = \frac{e^{-0.482}(0.482)^r}{r!}, \quad r = 0, 1, 2 \ldots 6$$

Expected frequencies -

$$E = N f(r) = 392 \times \frac{e^{-0.482}(0.482)^r}{r!}, \quad r = 0, 1 \ldots 6$$

$$\chi^2 = \frac{\Sigma (f_i - e_i)^2}{e_i} = 40.732$$

$$d.f = 4 - 1 - \underset{\text{because we are making } \mu = \bar{x}}{1} = 2$$

χ^2 test for independence of attributes -

Q. Two sample of boats to votes for two candidates A and B are taken and given in the following table -

Votes (f_i)

Area	A	B	Total
Rural	620	380	1000
Urban	550	450	1000
Total	1170	830	2000

Examine whether the nature of the area is related to voting preference in the election given $\chi^2_{0.05}$ for 1 d.f = 3.841 and for 4 d.f = 9.488

Sol. The Null hypothesis is voting preference is not related to the nature of the area.

$$E(620) = \frac{1000 \times 1170}{2000}$$

(e_i)

	A	B	Total
Rural	585	415	1000
Urban	585	415	1000
	1170	830	2000

$$\chi^2 = \sum \frac{(fi - ei)^2}{ei}$$

$$= \frac{(620-585)^2}{585} + \frac{(380-415)^2}{415} +$$

$$\frac{(550-585)^2}{585} + \frac{(450-415)^2}{415}$$

$$\chi^2 = 10.04$$

$r \times s$

$$d.f = (r-1)(s-1)$$
$$= (2-1)(2-1) = 1$$

$\chi^2_{0.05}$

Since calculated value is greater than the tabulated value, we reject null hypothesis at 5% level of significance and conclude that nature of area is related to voting preference.