UNIT - 5

SMALL SAMPLE :-

t - test :-

If the Sample is Small $n < 30$ the we get t- test for the distribution

The number of degree of freedom = Sample size (-1). ie) $d.f = n-1$ significant value of t at level of significance α. For a single tailed test can be obtained from those of two tailed test looking the value at level of significance 2α

APPLICATION OF t- distribution :-

The t - distribution has a wide number of application in Statistics, some of which are enumerated below.

(i) To test if the Sample mean $(\bar{x})$ differs significantly from the hypothetical value μ. of the population mean.

(ii) To test the significance of the difference between two sample means.

(iii) To test the significance of an observed sample correlation co-efficient and sample regression co-efficient

(iv) To test the significance of ~~ord~~ observed partial and multiple correlation co-efficient.

t- test for sample mean:-

Suppose we want to test:-

(i) If a random sample X_i $(i = 1, 2 \dots n)$ of size 'n' has been drawn from a normal population with a specified mean say μ_0 or.

(ii) If the sample means differs significantly from the hypothetical value μ_0 of the population mean.

Under the null hypothesis H_0:

(i) The sample has been drawn from the population with the mean μ.

(or)

(ii) There is no significant difference between the sample mean $\bar{x}$ and the population mean μ.

The Statistics $t = \dfrac{\bar{x} - \mu_0}{S/\sqrt{n}}$

where,
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and}$$
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$
follows students t-distribution with $(n-1)$ d.f.

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

Remarks:

1) If we take, $d_i = x_i - A$, where A, is any arbitary number, then
$$S^2 = \frac{1}{n-1}\left[\Sigma(x_i - \bar{x})^2\right]$$
$$= \frac{1}{n-1}\left[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right]$$
since variance is ~~ido~~ independent of change of origin,

Also in this case $\bar{x} = A + \dfrac{\Sigma d_i}{n}$

2) we know the Sample Variance
$$s^2 = \frac{1}{n}\sum_{i}(x_i - \bar{x})^2$$

Hence for numerical problems ~~that~~ Statistic
$$t = \frac{\bar{x} - \mu_0}{\sqrt{s^2/n}} = \frac{\bar{x} - \mu_0}{\sqrt{s^2/n-1}} \sim t_{n-1}$$

Test Statistics :-

$$t = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n}}} = \frac{\bar{x} - \mu}{\sqrt{s^2/n-1}} \sim t_{n-1}$$

$$= \frac{(0.742 - 0.700)}{\sqrt{\frac{(0.040)^2}{9}}} \sim t_9$$

$$= \frac{0.0420}{\frac{0.040}{3}} \sim t_9$$

$$= \frac{0.0420}{0.0133} \sim t_9$$

$$t = 3.1579 \sim t_9$$

Conclusion :-

The Critical Value of degree of freedom at 5% level of significance is 2.26 [2.26]

The Computed value of $|t| = 3.16 > 2.26$

$\therefore$ The null hypothesis is rejected.

The mean weekly sales of shop bars in departmental store was 146.3 (μ) bars per store after an advertising campaign the mean weekly sales in 22 (n) stores for a typical week increase to 153.7 ($\bar{x}$). Show that a standard deviation 17.2 was the advertising campaign successful ← (1.72)

Null hypothesis $H_0 : \mu = 146.3$

Alternative hypothesis $H_1 : \mu \neq 146.3$
(one tailed test)

Test statistic :-

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2/n-1}}$$

$$\cancel{\frac{\bar{x}-\mu}{\sqrt{s^2/n}}}$$

Given $\mu = 146.3$, $n = 22$, $\bar{x} = 153.7$, $s = 17.2$

$$t = \frac{153.7 - 146.3}{\sqrt{\frac{(17.2)^2}{22-1}}}$$

$$= \frac{7.4}{3.7533}$$

$$t = 1.9716$$

Conclusion :-

The Tabulated value of t for 21 d.f at 5% level of significance is 1.72

The computed value is $|t| = \cancel{1.72 < 1\%}$
$= 1.97 > 1.72$

The null hypothesis is rejected.

Q) A random Sample of $\overset{n}{10}$ boys had the following I.Q's test : 70, 120, 110, 101, 88, 88, 95, 98, 107, 100. Do this Data Support the assumption of a population mean I.Q of $\overset{\mu}{100}$? Find a reasonable range in which most of the mean I.Q values of Samples of 10 boys lie.

Soln:-

Null hypothesis $H_0 : M = 100$.

Alternative hypothesis $H_1 : M \neq 100$ (two tailed test.)

Test Statistic :-

Given $n = 10$, $M = 100.M$

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

$$t = \frac{\bar{x} - M}{\sqrt{S^2/n}}$$, where $\bar{x}$ and S^2 are to be Computed from the sample Values of I.Q.

Calculation:-

x	$x - \bar{x}$	$(x - \bar{x})^2$
70	−27.2	739.84.
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	−9.2	84.64
83	−14.2	201.64
95	−2.2	4.84
98	0.80	0.64
107	9.8	96.04
100	2.8	7.84 = 1823.67(3?)

Hence $n = 10$, $\bar{x} = \frac{\Sigma x}{n} = \frac{972}{10} = 97.2$

and $S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = 203.73$

$$t = \frac{97.2 - 100}{\sqrt{\frac{(203.73)}{10}}}$$

$$= \frac{-2.8}{\sqrt{20.3730}}$$

$$= \frac{-2.8}{4.5136}$$

$$t = -0.6203$$

Conclusion:-

The Critical value tabulated $t_{0.05}$ for 9 d.f for two-tailed test is 2.262

The Computed value $|t| = 0.6203 < 2.262$

The null hypothesis is accepted.

The heights of 10 males of a given locality are found to be 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe that the average height is greater than 64 inches? Test at 5% significance level, assuming that for 9 d.f $P(t > 1.833) = 0.05$.

Null hypothesis $H_0 : \mu = 64$

Alternative hypothesis $H_1 : \mu < 64$ (One tailed test).

Test Statistics :-

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2/n}}$$

Calculation:-

x	$x - \bar{x}$	$(x - \bar{x})^2$
70	4	16
67	1	1
62	-4	16
68	2	4
61	-5	25
68	2	4
70	4	16
64	-2	4
64	-2	4
66	0	0.
$\Sigma x = 660$		$\Sigma(x-\bar{x})^2 = 90$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{660}{10} = 66$$

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{90}{10-1} = \frac{\cancel{90}^{10}}{9} = 10.$$

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2/n}} = \frac{66 - 64}{\sqrt{10/10}} = 2$$

$\boxed{t = 2}$

Conclusion

① yes
② N

Tabulated value of t for 9 d.f 5% level of significant for single tailed test is 2.262

The computated value $|t| = 2 < 2.262$

The null hypothesis is accepted.

t - test for different of means

Suppose we want to test is two independent samples x_i $(i = 1, 2, 3 \ldots n)$ and y_j $(j = 1, 2, \ldots n)$ of size n_1 and n_2 have been drawn from two normal population with mean μ_x and μ_y respectively.

Under the null hypothesis (H_0) that the samples have been drawn from the normal population with mean μ_x and μ_y and under the assumption that the population that the population variance are equal

i.e) $\sigma_x^2 = \sigma_y = \sigma^2$, the statistic

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

where, $\bar{x} = \frac{1}{n_i} \sum_{i=1}^{n_i} x_i$

$$\bar{y} = \frac{1}{n_i} \sum_{i=1}^{n_i} y_i$$

Scanned with CamScanner

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

is an unbaised estimated of the common population variance (σ^2), follows students t- distribution with (n_1+n_2-2) d.f

1) The height of six randomly chosen sailors are in inches : 63, 65, 68, 69, 71, and 72. Those of 10 randomly chosen soldiers are 61, 62, 65, 66, 69, 69, 70, 71, 72 and 73. Discuss the light that these data throw on the suggestion the soldiers are on the average taller than soldier.

Soln:

If the heights of sailors and soldiers be represented by the variable X and Y respectively then the null hypothesis is $H_0: \mu_x = \mu_y$

The sailors are not on the average taller than the soldiers.

Alternative hypothesis $H_1: \mu_x > \mu_y$ (right tailed)

under H_0 the test statistic is

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2} = t_{14}$$

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

x	d = x − A	d^2	y	D = y − B	D^2
63	−5	25	61	−5	25
65	−3	9	62	−4	16
68	0	0	65	−1	1
69	1	1	$\boxed{66}$	0	0
71	3	9	69	3	9
72	4	16	69	3	9
			70	4	16
			71	5	25
			72	6	36
			73	7	49
	0	60		18	186

$$\bar{x} = A + \frac{\Sigma d}{n}$$

$$\bar{x} = 68 + 0 = 68$$

and $\Sigma(x - \bar{x}) = \Sigma d^2 - \frac{(\Sigma d^2)}{n_1}$

$$= 60 - 0$$

$$= 60.$$

$$\bar{y} = B + \frac{\Sigma D}{n_2}$$

$$= 66 + \frac{18}{10}$$

$$= 67.8.$$

and $\Sigma(y - \bar{y}) = \Sigma D^2 - \frac{(\Sigma D)^2}{n_2}$

$$= 186 - \frac{324}{10} = 153.6$$

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

$$= \frac{1}{14} \left[(60)^2 + (153.6)^2\right]$$

$$S^2 = 15.2571$$

$$t = \frac{68 - 67.8}{\sqrt{15.2571} \left(\frac{1}{6} + \frac{1}{10}\right)^{1/2}}$$

$$= \frac{0.2}{\sqrt{15.2571 \times 0.2667}}$$

$$= 0.099 \quad 0.0088$$

Tabulated $t_{0.05}$ for 14 d.f for single-tailed test is 1.76.

Conclusion :.

Since calculated t is much less than 1.76 is not at all significant ~~is not a~~ 5% level of significant. Hence null hypothesis may be retained at 5% level of significance and we conclude that the data are inconsistence with the suggestion that the sailors are on the average taller than soldier.

2) A Certains stimulus administred to each of the 12 patients resulted in the following increase of blood pressure. 5, 2, 8, -1, 3, 0, -2, 1, 5, 0, 4 and 6 Can it be concluded that the stimulus well in general, be accompained by an increase in blood pressure.

Soln :

Here we are given the increamenes in blood pressure ie) $d_i = (x_i - y_i)$

Null hypothesis is $H_0 : \mu_x = \mu_y$

ie) There is no significance difference in the blood pressure readings of the patients before and after drug. In the other words the given incremants are just by change (fluctuations of Sampling) are not due to the stimulus.

Scanned with CamScanner

Alternative Hypothesis, $H_1 : \mu_x < \mu_y$

(i) The stimulus results is an increase in blood pressure.

Test Statistic under H_0, the test statistic is

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

d	5	2	8	-1	3	0	-2	1	5	0	4	6	31
d^2	25	4	64	1	9	0	4	1	25	0	16	36	185

$$S^2 = \frac{1}{n-1}\Sigma(d-\bar{d}) = \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right]$$

$$= \frac{1}{11}\left[185 - \frac{(31)^2}{12}\right] = \frac{1}{11}(185 - 80.08)$$

$S^2 = 9.5382$ and $d = \frac{\Sigma d}{n} = \frac{31}{12} = 2.58$.

$$t = \frac{\bar{d}}{s/\sqrt{n}} = \frac{2.58 \times \sqrt{12}}{\sqrt{9.5382}} = \frac{2.58 \times 3.464}{3.09} = 2.89$$

Tabulated $t_{0.05}$ for 11 d.f for right tailed test is 1.80 [This is the value of $t_{0.10}$ for 11 d.f in the table for two tailed test given the appendix]

Conclusion:

Since Calculated $t > t_{0.05}$. H_0 is rejected at 5% level of significance. Hence we conclude that the stimulus will, in general, be compared by an increase in blood pressure.

3) In a Certain experiment to Compare two types of foods A and B the following data. Result of Increase in weight were Observed in

Numbers		1	2	3	4	5	6	7	8	total
Increase in weight in lb	Foods A	49	53	51	52	47	50	52	53	407
	Foods B	52	55	52	53	50	54	54	53	423

(i) Assuming that the two samples of pigs are independent can be conclude that food B is better than food A?

(ii) Also examine the Case when the same set of eight pigs were used in both the foods.

Soln:

Null hypothesis : $H_0 : \mu_x = \mu_Y$

Alternative hypothesis $H_1 : \mu_x < \mu_Y$ (left tailed)

(i) If the two samples of pigs be assumed to be independent than we will apply t- test for difference test for difference of means to test H_0.

Test Statistic :

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

Food A			Food B		
x	d	d^2	y	D	D^2
49	-1	1	52	0	0
53	3	9	55	3	9
51	1	1	52	0	0
52	2	4	53	1	1
47	-3	9	50	-2	4
50	0	0	54	2	4
52	2	4	54	2	4
53	3	9	53	1	1
	7	37		7	23

$$\bar{x} = 50 + \frac{7}{8} = 50.875$$

$$\bar{y} = 52 + \frac{7}{8} = 52.875$$

and $\Sigma(x-\bar{x}) = \Sigma d^2 - \frac{(\Sigma d)^2}{n}$

$$= 37 - \frac{49}{8}$$

$$= 30.875$$

$$\Sigma(y-\bar{y}) = \Sigma D^2 - \frac{(\Sigma D)^2}{n}$$

$$= 23 - \frac{49}{8}$$

$$= 16.875$$

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

Scanned with CamScanner

$$S^2 = \frac{1}{4}(30.875 + 16.878) = 3.41$$

$$t = \frac{50.875 - 52.875}{\sqrt{3.41\left(\frac{1}{8} + \frac{1}{8}\right)}}$$

$$t = -2.17 \Rightarrow |t| = 2.17$$

Tabulated $t_{0.05}$ for $(8 + 8 - 2) = 14$ d.f for one tail test is 1.76.

The computed value of $|t| = 2.17 > 1.76$

$\therefore$ The null hypothesis is rejected.

Scanned with CamScanner

CHI-SQUARE TEST OF GOODNESS OF FIT:

A very powerful test for testing the significance of the discrepancy between theory and experiment was given by Prof. Karl Pearson in 1900 and is known as 'Chi-Square test of goodness of fit". It enables us to find if the deviation of the experiment theory is just by change or is fit really due to the inadequacy of the theory to fit the Observed data. If O_i $(i=1,2 \dots n)$ is set of Observed (experimental) frequencies and E_i $(i=1,2 \dots n)$ is the Corresponding set of expected (theoretical or hypothetical) frequencies then the karl Pearson's. Chi-Square given by

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E_i} \right], \quad \left(\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i \right)$$

follows chi-Square distribution with $(n-1)$ d.f.

1) The following tables gives the numbers of aircraft accidents that occurs during the various day of the week. Find whether the accidents are uniformally distributed over the week.

Days	sun	mon	tues	wed	thus	fri	sat
No. of accidents	14	16	8	12	11	9	14

(Given : the value of chi-square significant at 5, 6, 7 d.f are respectively 11.07, 12.59, 14.07 at the 5% level of significant).

Soln:

Here we set up the null hypothesis that the accidents are uniformally distributed over the week.

Under the null hypothesis there expected frequency of the accidents on each of the days would be

Days	Sun	Mon	Tues	Wed	Thurs	Fri	Sat
No. of accidents	12	12	12	12	12	12	12

Total 84

O_i	E_i	$(O-E)^2$	$\frac{(O-E)^2}{E}$
4	12	4	0.3333
16	12	16	1.3333
8	12	16	1.3333
12	12	0	0
11	12	1	0.0833
9	12	9	0.7500
14	12	4	0.3333
	84		4.1665

$\chi^2 = 4.1665$

$\chi^2 = 4.17$

The number of degree of freedom

= Number of Observation - Number of independent Constraints.

= 7-1

= 6

The tabulated $\chi^2_{0.05}$ for 6 d.f = 12.59

Since the calculated χ^2 is much less than the tabulated value is its highly insignificance and we accept the null hypothesis. Hence we conclude that the accidents are uniformly distributed over the week.

12.59

3) The theory predicts the proportions of beans in the four groups A, B, C and 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 result support the theory?

Soln:

Null hypothesis: We set up the null hypothesis that the theory fits well into the experiment.

ie) The experimental result support the theory

Under the null hypothesis. The expected (theoritical) frequency can be computed as follows

Total numbers of bears ⇒ 882 + 313 + 287 + 118

$= 1600$.

These are to be divided in the ratio = 9 : 3 : 3 : 1

$E(882) = 9/16 \times 1600 = 900$

$E(313) = 3/16 \times 1600 = 300$

$E(287) = 3/16 \times 1600 = 300$

$E(118) = 1/16 \times 1600 = 100$.

O_i	E_i	$(O-E)^2$	$\frac{(O-E)^2}{E}$
882	900	324	0.3600
313	300	169	0.5633
287	300	169	0.5633
118	100	324	3.2400
			4.7266

$\chi^2 = 4.7266$

d.f $= 4-1 = 3$ and tabulated $\chi^2_{0.05}$ for 3 d.f $= 7.815$

Since the calculated value of χ^2 is less than the tabulated value it is not significant. Hence the null hypothesis may be accepted at 5% level of significant and we may conclude that there is good correspondence between theory and experiment.

Scanned with CamScanner

8. The following figures show the distribution of digits in numbers chosen at random from a telephone directory.

Pg 15 27

Digits	0	1	2	3	4	5	6	7	8	9	total
frequency	1026	1107	997	966	1075	933	1107	972	964	853	10,000

Test whare the digits may be taken to occur equally frequenctly in the directory.

Soln:-

Here we set up the null hypothesis that the digits occurr equally frequency in the directly.

Under the null hypothesis the expected frequency for each of the digits 0, 1, 2 ... 9 is $\frac{10000}{10} = 1000$. The value of χ^2 is computed as follows.

Digits	Observed Frequency	Expected Frequency	$(O-E)^2$	$(O-E)^2/E$
0	1026	1000	676	0.676
1	1107	1000	11449	11.449
2	997	1000	9	0.009
3	966	1000	1156	1.156

4	1075	1000	5625	5.625
5	933	1000	4489	4.489
6	1107	1000	11449	11.449
7	972	1000	784	0.784
8	964	1000	1296	1.296
9	853	1000	21609	21.609
Total	10000	10000		58.542

$\chi^2 = 58.542$

The number of degree of freedom $= 10-1 = 9$ (since we are given 10 frequencies subjected to only one linear constraints) $\Sigma O = \Sigma E = 10000$

The tabulated χ^2 0.05 for d.d.f 16.919.

Since the calculated χ^2 is much greatess than the tabulated value it is highly. Significant and we rejected the null hypothesis. Thus we conclude that the digits are not uniformly distributed in the directory.

4) A survey of 320 families with 5 children each revealed the following distribution

15.29.

No. of boys	5	4	3	2	1	0
No. of girls	0	1	2	3	4	5
No. of families	14	56	110	88	40	12

Is the result consistent with the hypothesis that male and female birth are equally probable?

Soln:

Let us set up the null hypothesis that the data are consistent with the hypothesis of equally.

$$\psi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E} \right] \sim \psi^2_{n-1}$$

Given $N = 320$, $n = 5$

$p = 1/2$, $q = 1/2$

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

$= 5C_x (1/2)^x (1/2)^{5-x}$

$= 5C_x (1/2)^{\cancel{x}+5-\cancel{x}}$

$= 5C_x (1/2^5)$

$P(x) = 5C_x (1/32)$.

$f(x) = N \cdot P(x)$

$f(0) = 320 \times 5C_0 (1/32)$

$= 320 \times 1/32$

$= 10$

$f(1) = 320 \times 5C_1 (1/32) = 50$

$f(2) = 320 \times 5C_2 (1/32) = 100$

$f(3) = 320 \times 5C_3 (1/32) = 100$

$f(4) = 320 \times 5C_4 (1/32) = 50$; $f(5) = 320 \times 5C_5 (1/32) = 10$

O_i	E_i	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
4	10	16	1.6
56	50	36	0.12
110	100	100	.1
88	100	144	1.44
40	50	100	2
12	10	4	0.4
		400	7.16

$$\psi^2 = \sum_{i=1}^{n} \left[\frac{(O_i - E_i)^2}{E} \right] \sim \psi^2_{n-1}$$

$$\psi^2 = 7.16 \sim \psi_4$$

The critical value of ψ^2 for 4 d.o.f at 5% level of significance is 9.488

The computed value of $|\psi^2| = 7.16 < 9.488$

∴ The null hypothesis is ~~reje~~ accepted.

5. (A) Fit a poisson distribution to the following data and test the goodness of fit.

x	0	1	2	3	4	5	6
f	275	72	30	7	5	2	1

15-30

Soln:-

Mean of the given distribution is

$$\bar{x} = \frac{\sum fx}{N} = \frac{189}{392} = 0.482$$

In order to fit a poisson distribution to the given data we take the mean (Parameter) m of the poisson distribution equal to the mean of the given distribution

ie) We take $m = \bar{x} = 0.482$

The frequency of r successes is given by the poisson law as

$$f(r) = NP_r = \frac{392 \times e^{-0.482} \cdot (0.482)^r}{r!}$$

Now,

$$NP_r = 392 \times e^{+0.482}$$

$$f(0) = 392 \times e^{-0.482}$$

$= \cancel{392 \times \text{Antilog}}\ [-0.482 \log e)$

$= 392 \times \text{Antilog}\ [-0.482 \times \log 2.7183)$

$= 392 \times \text{Antilog}\ [-0.482 \times 0.4343]$

$= 392 \times \text{Antilog}\ (\bar{1}.7907)$

$= 392 \times 0.6175$

$f(0) = 242.1$

$f(1) = m \times f(0)$

$= 0.482 \times 242.1 = 116.69$

$f(2) = m/2 \times f(1) = 0.241 \times 116.69 = 28.12$

$f(3) = m/3 \times f(2) = 0.1607 \times 28.12 = 4.518$

$f(4) = m/4 \times f(3) = 0.1205 \times 4.518 = 0.544$

$f(5) = m/5 \times f(4) = 0.0964 \times 0.544 = 0.052$

$f(6) = m/6 \times f(5) = 0.0802 \times 0.052 = 0.004$

Hence the theoritical poisson frequencies correct to one decimal place are as given below.

X	0	1	2	3	4	5	6	total
Expected frequency	242.1	116.7	28.1	4.5	0.5	0.1	0	392

CALCULATION OF CHI - SQUARE.

Observed frequency	Expected frequency	(O-E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
275	242.1	32.9	1082.41	4.471
72	116.7	44.7	1998.09	17.121
30	28.1	1.9	3.61	0.128
7, 5, 2 } 14	4.5, 0.5, 0.1 } 5.1	9.9	98.01	19.217
1	0			
392	392.0			40.937

$\chi^2 = 40.937$

Degree of freedom $7-1-1-3 = 2$

Tabulated value of χ^2 for 2 d.f at 5% level of significance is 5.99

Conclusion:

Since Calculated Value of χ^2 (40.987) is much greater than 5.99 it is highly significant Hence we conclude the poisson distribution a good fit to given data.

Q1) Independent of Attributes:

Let us Consider two attributes A and B, A divied into r classes $A_1, A_2 \ldots A_r$ and B divided into s classes $B_1, B_2 \ldots B_s$ such a classification in which attributes are divided into more than 2 classes is known as manifold classification the various cell frequencies can be expressed in the following table known as $r \times s$ manifold contingency table where (A_i) is the number of person possessing the attributes A_i, $(i = 1, 2 \ldots r)$ (B_j) is the number of person possessing the attributes B_j $(j = 1, 2, \ldots s)$ and (A_iB_j) is the number of person possessing Both the attributes A_i and B_j $[i = 1, 2 \ldots r;$ $j = 1, 2 \ldots s]$ Also, $\sum_{i=1}^{r} (A_i) = \sum_{j=1}^{s} (B_j) = N$ is the total frequency.

$r \times s$ Contingency table :-

A / B	A_1	A_2	....	A_i	....	A_r	total
B_1	(A_1B_1)	(A_2B_1)	....	(A_iB_1)	...	(A_rB_1)	(B_1)
B_2	(A_1B_2)	(A_2B_2)	....	(A_iB_2)	...	(A_rB_2)	(B_2)
$\vdots$							
B_j	(A_1B_j)	(A_2B_j)	...	(A_iB_j)	..	(A_rB_j)	(B_j)
$\vdots$							
B_s	(A_1B_s)	(A_2B_s)	...	(A_iB_s)	..	(A_rB_s)	(B_s)
total	(A_1)	(A_2)	...	(A_i)	...	(A_r)	N.

The problem is to test if two attributes A and B under considerator are independent or not.

Under the Null hypothesis that the attributes are independent, the theortical cell frequencies are calculated as follows.

$P(A_i)$ probability that a person possesses the attributes A_i

$$= \frac{(A_i)}{N} \; ; i = 1, \dots r.$$

$P(B_j)$ probability that a person possesses the attributes B_j

$$= (B_j)/N \; ; j = 1, \dots r.$$

$P[A_i B_j]$ = Probability that a person possesses the attributes A_i and B_j.

$= P(A_i) P(B_j)$

(By compound probability theorem, since the attributes A_i and B_j are independent under the Null hypothesis)

$\therefore P[A_i B_j] = \frac{(A_i)}{N} \cdot \frac{(B_j)}{N}$; $i = 1, 2 \ldots r$; $j = 1, 2 \ldots s$.

$\therefore (A_i B_j)_0$ = Expected number of person possessing both the attributes A_i and B_j

$= N P[A_i B_j] = \frac{(A_i)(B_j)}{N}$

$(A_i B_j)_0 = \frac{(A_i)(B_j)}{N}$ $i = 1, 2, 3 \ldots r$; $s = 1, 2 \ldots s$

By using this formula we can find out expected frequency for each of the cell frequencies $(A_i B_j)$; $(i = 1, 2 \ldots r$; $j = 1, 2 \ldots s)$., under the Null hypothesis of independence of attributes.

The exact test for the independence of attributes is very complicated but a fair degree of approximation is given for large sample (large N) by the χ^2-test of goodness of fit.

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \left[\frac{\{(A_i B_j) - (A_i B_j)_0\}^2}{(A_i B_j)_0} \right]$$

which is distributed as a χ^2-variate with $(r-1)(s-1)$ d.f.

1) Two sample polls of votes for two candidates A and B for a public office are taken, one from among the residents of rural areas. The results are given in the table examine, whether the nature of the area is related to voting performed in the election.

15.37

Votes for area	A	B	Total
Rural	620	380	1000
Urban	550	450	1000
total	1170	820	2000

$\frac{1000 \times 820}{200}$

$E_i = 415$

Soln:-

Under the null hypothesis that the nature of the ~~under~~ area is independent of the voting preference in the election. We get the observed frequency as follows.

$585 = \frac{1170 \times 1000}{200}$

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

$$= \frac{(620-585)^2}{585} + \frac{(380-415)^2}{415} + \frac{(550-585)^2}{585} + \frac{(450-415)^2}{415}$$

$\chi^2 = 2.0940 + 2.9518 + 2.0940 + 2.9518$

$\chi^2 = 10.0916$

Tabulated $\chi^2_{0.05}$ for $(2-1)(2-1)$ $[(r-1)(c-1)] = 1$ d.f is 3.841

The computed value of $\chi^2 = 10.0916 > 3.841$

$\therefore$ The Null hypothesis is rejected.

Scanned with CamScanner

8) $\begin{matrix} a & b \\ c & d \end{matrix}$ Prove that Chi-square test of independence gives

$$\chi^2 = \frac{N(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}, \quad N = a+b+c+d \ldots$$

Soln :-

Under the hypothesis of independence of attributes

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	N

$$E(a) = \frac{(a+b)(a+c)}{N}$$

$$E(b) = \frac{(a+b)(b+d)}{N}$$

$$E(c) = \frac{(c+d)(a+c)}{N}$$

$$E(d) = \frac{(b+d)(c+d)}{N}$$

$$\chi^2 = \frac{[a-E(a)]^2}{E(a)} + \frac{[b-E(b)]^2}{E(b)} + \frac{[c-E(c)]^2}{E(c)} + \frac{[d-E(d)]^2}{E(d)} \longrightarrow (1).$$

$$a - E(a) = a - \frac{(a+b)(a+c)}{N}$$

$$= \frac{a(a+b+c+d) - (a^2+ac+ab+bc)}{N}$$

$$a - E(a) = \frac{ad-bc}{N} = d - E(d).$$

Similarly, we will get,

$$b - E(b) = -\frac{ad-bc}{N} = c - E(c).$$

Sub in ① we get.

$$\chi^2 = \frac{(ad-bc)^2}{N^2}\left[\frac{1}{E(a)} + \frac{1}{E(b)} + \frac{1}{E(c)} + \frac{1}{E(d)}\right]$$

$$= \frac{(ad-bc)^2}{N}\left[\left\{\frac{1}{(a+b)(a+c)} + \frac{1}{(a+b)(b+d)}\right\} + \left\{\frac{1}{(a+c)(c+d)} + \frac{1}{(b+d)(c+d)}\right\}\right]$$

$$= \frac{(ad-bc)^2}{N}\left[\frac{b+d+a+c}{(a+b)(a+c)(b+d)} + \frac{b+d+a+c}{(a+c)(c+d)(b+d)}\right]$$

$$= (ad-bc)^2\left[\frac{c+b+a+d}{(a+b)(a+c)(c+d)(b+d)}\right]$$

$$= \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}.$$

3) Find if A and B are independent (or) positively associated (or) negative associate in the following (A) = 400, (AB) = 294, (α) = 570
(αB) = 380.

Soln:-

Given (A) = 420, (AB) = 294, (α) = 570, (αB) = 380.

$N = (A) + (\alpha)$

$= 420 + 570$

$\boxed{N = 990}$

$N = (A) + (\alpha)$

$B = (AB) + (\alpha B)$

$(AB_0) = \frac{(A)(B)}{N}$

$\delta = (AB) - (AB)_0$.

$(B) = (AB) + (\alpha B)$

$= 294 + 380$

$\boxed{(B) = 674}$

$$(AB)_0 = \frac{(A)(B)}{N}$$

$$= \frac{(420)(674)}{990}$$

$$\boxed{(AB)_0 = 285.94}$$

$$\delta = (AB) - (AB)_0$$

$$= 294 - 285.94$$

$$\boxed{\delta = 8.06}$$

$$\delta = 8.06 > 0.$$

$\therefore$ A and B are positively associated.

Scanned with CamScanner