UNIT-3:-

Test of statistical Hypothesis.

Test of statistical Hypothesis.

It is a two action decision problem after the experimental sample values have been obtained. The Two actions being The acceptance or the rejections of Hypothesis

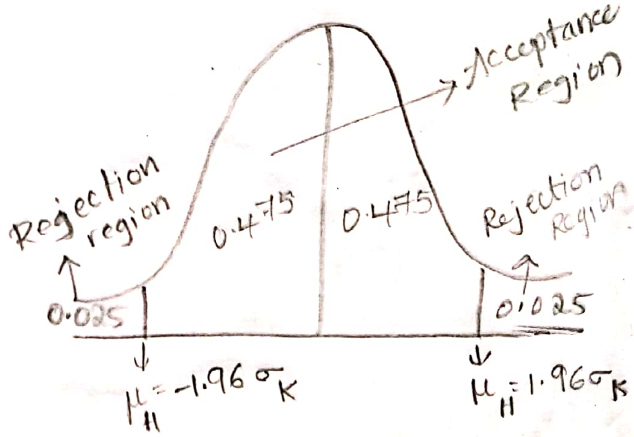
Procedure of Testing Hypothesis.

(1) Setup a Hypothesis.
(2) setup a suitable significance level.
(3) Setting a criterian
(4) Doing computations.
(5) Making Decisions.

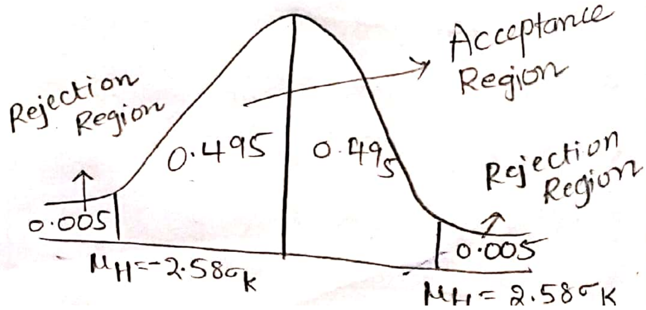
Two Tailed and One Tailed Test of Hypothesis.

A Two Tail Test of Hypothesis will reject null Hypothesis if the sample static is significantly higher than or lower than the hypothesis papulation parameter.

Thus the rejection region is located in both the tails. If we are testing a hypothesis at 5% level of significance, The size of acceptance region on each side of the mean. is 0.475 and the size of acceptance region is 0.025



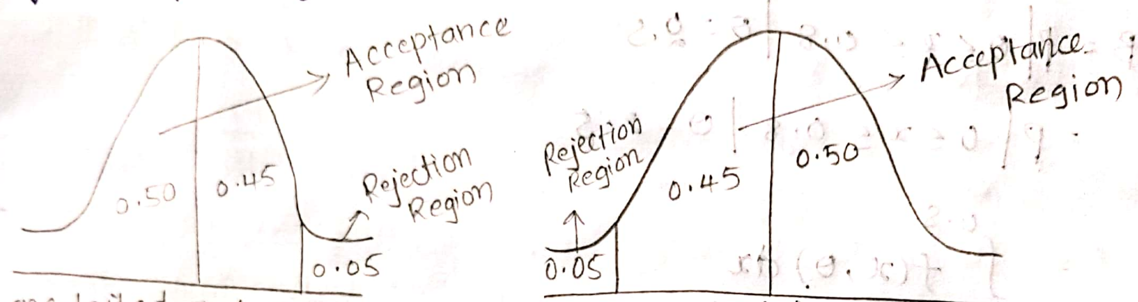
If The Test of Hypothesis at 1% level of significance Then The above curve will be changed as



From The above Two curves, at 5% significance level corresponds to 1.96 standard Error.

At 1% level of significance it corresponds to 2.58 standard Error.

In The case of one-tailed Test, The rejection region will be located in only one tail, either in the left or in the right depending on the alternative hypothesis.



Critical values	Level of significance		
	1%	5%	10%
Two Tailed Test	$\lvert z\rvert = 2.58$	$\lvert z\rvert = 1.96$	$\lvert z\rvert = 1.645$
Right Tailed Test	$z = 2.33$	$z = 1.645$	$z = 1.28$
Left tailed Test.	$z = -2.33$	$z = -1.645$	$z = -1.28$

Ex: In The Frequency Function $f(x,\theta) = \frac{1}{\theta}$, $0 \le x \le \theta$

$= 0$, elsewhere.

and that you are Testing The Hypothesis, $H_0 : \theta = 1.5$ against $H_1 : \theta = 2.5$ by means of a single observed value of x what would be the sizes of Type I and type-II errors, if you choose the inverval $0.8 \le x$ as the critical region? Also obtain the power function of the Test.

$f(x,\theta) = \frac{1}{\theta}$

$H_0 = 1.5$, $H_1 = 2.5$

$0.8 \le x$

Now $\alpha = P\left|0.8 \le x \le \theta \right| \theta = 1.5$

$= P\left|0.8 \le x \le 1.5 \right| \theta = 1.5$

$$= \int_{0.8}^{1.5} f(x,\theta)\,dx \Big|_{\theta=1.5}$$

$$= \left[\frac{x}{\theta}\right]_{0.8}^{1.5}$$

$$= \frac{1.5-0.8}{1.5} = \frac{0.7}{1.5} = 0.467$$

$$\beta = P\left|0 \le x \le 0.8 \right| \theta = 2.5$$

$$= P\left|0 \le x \le 0.8 \right| \theta = 2.5$$

$$= \int_{0}^{0.8} f(x,\theta)\,dx \Big|_{\theta=2.5}$$

$$= \left[\frac{x}{\theta}\right]_{0}^{0.8}$$

$$= \frac{0.8-0}{2.5} = \frac{0.8}{2.5} = 0.32.$$

The power function of Test, $1-\beta$

$$= 1-0.32$$

$$= 0.68$$

(2) Let 'p' be the probability that a coin will fall head in a single toss. In order to test the hypothesis $H_0: P = 1/2$, The coin is tossed 6 times and the hypothesis H_0 is rejected if more than 4 heads are obtained. Find the probability of the error of the first kind. If the alternative hypothesis is $H_1: P = 3/4$ Find the probability of the Error of second kind.

If 'x' denotes the number of heads in 'n' tosses of a coin, Then $X - B(n,p)$ so That

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \binom{6}{x} p^x (1-p)^{6-x}$$

Since $n=6$, $\alpha = P[X \geq 5 \mid H_0]$

Scanned with OKEN Scanner

$$= P[X=5 \mid P=\tfrac{1}{2}] + P[X=6 \mid P=\tfrac{1}{2}]$$

$$= \binom{6}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right) + \binom{6}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^0$$

$$= \frac{6}{2^6} + \frac{1}{2^6} = \frac{7}{2^6} = \frac{7}{64}$$

$$\beta = 1 - P[(X=5) \mid P=\tfrac{3}{4} + P(X=6) \mid P=\tfrac{3}{4}]$$

$$= 1 - \left[\binom{6}{5}\left(\frac{3}{4}\right)^5\left(\frac{1}{4}\right) + \binom{6}{6}\left(\frac{3}{4}\right)^6\left(\frac{1}{4}\right)^0\right]$$

$$= 1 - \left[\frac{1458}{4096} + \frac{729}{4096}\right] = 1 - 0.5339$$

$$= 0.4661$$

Large Sample Tests:

If the size of the sample $n > 30$, Then the sample is called large sample. For large samples, There are four important tests available for testing the significance level. They are the Tests of significance for

(i) single mean

(ii) difference of mean.

(iii) single proportion.

(iv) difference of proportion.

In This case, almost all the distributions eg, binomial, poisson etc., are very closely approximated by normal distribution. Thus if we apply the normal test of the normal probability curve, the standardized variable corresponding to the statistic 't' is

$$Z = \frac{t - E(t)}{S.E.(t)} \sim N(0,1)$$

where $E(t)$ is the mean of t and $S.E(t)$ is the Standard Error of 't'.

Test of significance for the sample mean:

The standard error (S.E.) of mean of a random sample of size 'n' from a population with variance σ^2 is $\sigma/\sqrt{n}$. Thus if we want to test whether the given sample of size 'n' has been drawn from a population sample with mean μ

we set a null hypothesis H_0 i.e., There is no difference between sample mean $\bar{x}$, and population mean 'μ'. Therefore test statistic corresponding to $\bar{x}$ is written as $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.

In the above test statistic, if population variance σ^2 is unknown, we can use its sample variance i.e., $\sigma^2 \Rightarrow s^2$ This is applicable only for large samples. The 95% confidence limits for μ are $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$. similarly. 99% confidence limits for μ are $\bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$ ~~for~~ and for 98% They are

$\bar{x} \pm 2.33 \frac{\sigma}{\sqrt{n}}$

(3) A sample of 100 N.C.C students from university of madras was taken and their average weight was found to be 60 kgs. with a standard deviation of 5kgs. can you infer about the average weight of all the NCC students in the entire university

$n = 100$, $\bar{x} = 60$, $s = 5$

$$\therefore SE = \frac{\sigma}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{5}{\sqrt{100}} = \frac{5}{10} = 0.5.$$

The average weight of all the NCC students in the entire university will be $\bar{x} \pm 3$ S.E. This is because, if the sampling was simple random sampling, the average weight of the students in the entire university cannot deviate from the average weight of the sample by more than thrice the standard Error.

Thus, it is $60 \pm 3(0.5) = 60 \pm 1.5$

$= 58.5$ to 61.5 kgs.

(4) The mean breaking strength of cables supplied by a manufacturer is 1800 with a standard deviation 100. By a new Technique in the manufacturing process, it is claimed that the Breaking strength of the cables have increased. In order to test the claim, a sample of 50 cables is tested. It is found that the mean breaking strength is 1850. can we support the claim at 0.01 level of significance.

Scanned with OKEN Scanner

Null Hypothesis $H_0 : \mu = 1800$.

Alternative Hypothesis $H_1 : \mu > 1850$

Test statistic, $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{(1850 - 1800)}{100} \times \sqrt{50}$

$= 3.536$.

Since $|z| > 2.58$, we conclude that the data provides us an evidance against the null Hypothesis H_0 which may, therefore, be rejected at 1% level of significance.

(5) An Ambulance service claims that it takes on an average 8.9 minutes to reach its destination in emergency calls. To check on this claim, The agency which licences ambulance services has timed them on 50 Emergency calls getting a mean of 9.3 minutes with a standard deviation of 1.6 minutes. What can they conclude at the level of significance $\alpha = 0.05$.

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

$= \frac{9.3 - 8.9}{1.6} \times \sqrt{50}$

$= 1.768$

Since $|z| < 1.96$, The agency can conclude that it can be accepted at the level of significance of 0.05.

(6) A sample of 450 items is taken from a ~~particular~~ papulation whose standard deviation is 20. The mean of the sample is 30. Test whether the sample has come from the papulation with mean 29. Also calculate the 95% confidence limits of the population mean.

$n = 450$, $\sigma = 20$, $\bar{x} = 30$, $\mu = 29$, $\alpha = 0.05$.

Test statistic, $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{30 - 29}{20} \times \sqrt{450}$

$= 1.061$.

Since $|z| < 1.96$, it is accepted that the sample has come from a papulation of mean 29.

95% Confidence limits of the papulation mean are

$\bar{x} \pm 1.96\ \sigma/\sqrt{n} \Rightarrow 30 \pm (1.96 \times 1.061)$

$= 30 \pm 2.079$

$= 27.921$ and 32.079.

(7) A normal distribution has mean 0.5 and standard deviation 2.5. Find (i) The probability that the mean of a random sample of size 16 from the papulation is positive.
(ii) The probability that the mean of a sample of size 90 from the papulation will be negative.

(i) Test statistic, $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - 0.5}{2.5/4} = \frac{\bar{x} - 0.5}{0.625}$.

$\therefore \bar{x} = 0.5 + 0.625\ z$

The required probability, p that the sample mean is positive is given by

$p = P(\bar{x} > 0) = P[(0.5 + 0.625\ z) > 0]$.

$= P\left(z > \frac{-0.5}{0.625}\right)$

$= P(z > -0.8)$

$= 0.5 + P(0 < z < 0.8)$

$= 0.5 + 0.2881$

$= 0.7881$.

(ii) $z = \frac{\bar{x} - 0.5}{2.5/9.487} = \frac{\bar{x} - 0.5}{0.2635}$.

$\therefore \bar{x} = 0.5 + 0.2635\ z$.

The required probability, p that the sample mean is negative is given by $p = P(\bar{x} < 0) = P[(0.5 + 0.2635\ z) < 0]$

$= P\left(z < \frac{-0.5}{0.2635}\right)$

$= P(z < -1.898)$

$= P(z > 1.898)$

$= 0.5 - P(0 < z < 1.898)$

$= 0.5 - 0.4713$

$= 0.0287$ (from normal Tables).

(8) The average value $\bar{x}$ of a random sample of observations from a certain papulation is normally distributed with mean 20 and standard deviation $\frac{5}{\sqrt{n}}$ How large a sample should be drawn in order to have a probability of at least 0.90 That $\bar{x}$ will lie between 18 and 22.

Sol: $z_1 = \frac{18-20}{\left(\frac{5}{\sqrt{n}}\right)/\sqrt{n}} = \frac{-2n}{5}$

Similarly, $z_2 = \frac{22-20}{\frac{(5/\sqrt{n})}{\sqrt{n}}} = \frac{2n}{5}$

It is required to find $P(z_2 > z > z_1)$

ie., $P\left(\frac{2n}{5} > z > \frac{-2n}{5}\right)$

$= 2P\left(0 < z < \frac{2n}{5}\right) \geq 0.9$

$\Rightarrow P\left(0 < z < \frac{2n}{5}\right) \geq 0.45$

$\frac{2n}{5} = 1.65$

$\therefore n = 4.125.$

Therefore, atleast four samples should be drawn.

(9) If it costs a rupee to draw one number of a sample, how mush would it cost in sampling from a universe with mean 100 and standard deviation 10 to take sufficient number as to ensure that the mean of a sample would in a 5% probability be within 0.01 percent of the True value ? Find the extra cost necessary to double this precision.

Difference between universe mean and sample mean = 0.01 %

For 95% confidence, difference between sample mean and papulation mean should be equal to 1.96 times the standard error ie.,

$$1.96 \times \frac{10}{\sqrt{n}} = 0.01$$

$$\therefore \sqrt{n} = \frac{19.6}{0.01}$$

$$\Rightarrow n = 38,41,600.$$

Double the precision would imply that the difference between sample mean and the papulation mean = 0.005

$$\therefore 1.96 \times \frac{10}{\sqrt{n}} = 0.005$$

$$\Rightarrow n = 1,53,66,400.$$

$\therefore$ Extra Cost = 1,53,66,400 − 38,41,600

= Rs 1,15,24,800.

Student's t-Distribution :

The student's t-distribution is used, when the papulation standard deviation unknown. Let x_i $(i = 1, 2, \dots n)$ be a random sample of size 'n' from a normal papulation with mean μ and variance σ^2. Then the 't-statistic' is defined as

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

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})$

The probability density function of t-distribution with 'γ' degrees of freedom is

$$f(t) = c\left[1 + \frac{t^2}{\gamma}\right]^{-(\gamma+1)/2}, \quad -\infty < t < \infty$$

where 'c' is a constant required to make the area under the curve equal to unity.

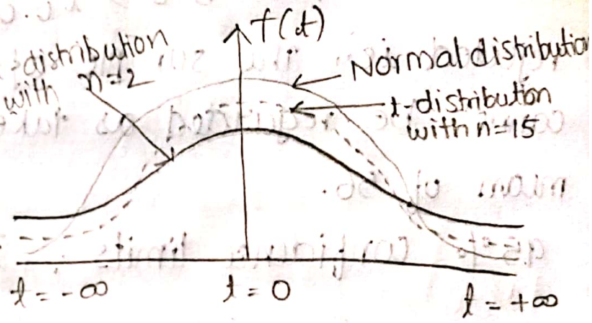
Properties of t-distribution :

(1) The variable 't' in student's t-distribution ranges from minus infinity to plus infinity.

(2) Like the standard normal distribution, the t-distribution is symmetrical and has a mean equal to zero.

(3) If we take $\nu = 1$, we get

$$f(t) = \frac{c}{1+t^2}, \quad -\infty < t < \infty$$



which is the p.d.f of standard cauchy distribution. Hence when $\nu = 1$ student's t-distribution reduces to cauchy distribution.

(4) The variance of the t-distribution is greater than one. But it approaches one as ν (and hence n) becomes large. In other words, the variance of t-distribution approaches the variance of standard normal distribution as the sample size 'n' increases. When $\nu = \infty$, t-distribution and normal distribution are exactly equal.

ex:

(1) A random sample of size 16 has 53 as mean. The sum of the squares of the deviations taken from mean is 150. Can this sample be regarded as taken from the population having 56 as mean. obtain 95% and 99% confidence limits of the mean of the population.

Given $n = 16$, $\bar{x} = 53$, $\mu = 56$ and $\Sigma (x_i - \bar{x})^2 = 150$.

Null Hypothesis H_0: The sample can be regarded as taken from the population having mean of 56.

The Test statistic is

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

$$s^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2 = \frac{150}{15} = 10.$$

$$\therefore \; t = \frac{53 - 56}{\sqrt{10}/\sqrt{16}} = -3.8$$

For $\nu = n-1 = 15$ degrees of freedom and at 1% level i.e., $\alpha = 0.01$, $t_{0.01} = 2.947$

Since $|t| = 3.8 > t_{0.01}$, the null hypothesis is rejected i.e., The sample of size 16 with mean 53 cannot be regarded as taken from the population having mean of 56.

95% confidence limits : $\bar{x} \pm \frac{S}{\sqrt{n}} t_{0.05}$

$$= 53 \pm \frac{\sqrt{10}}{\sqrt{16}} t_{0.05} = 53 \pm \frac{3.162}{4} \times 2.131$$

$$= 53 \pm 1.684$$

$$= 51.32 \text{ and } 54.68.$$

and 99% confidence limits are : $\bar{x} \pm \frac{S}{\sqrt{n}} t_{0.01}$

$$= 53 \pm \frac{3.162}{4} \times 2.947 = 50.67 \text{ and } 55.33.$$

(2) The nine items of a sample had the following values 45, 47, 50, 52, 48, 47, 49, 53 and 51. Does the mean of nine items differ significantly from the assumed population mean of 47.57? Given that for $\gamma = 8$, $P(t=1.8) = 0.945$ and $P(t=1.9) = 0.953$.

Given $\mu = 47.5$

Null Hypothesis $H_0 : \mu = 47.5$

Alternative Hypothesis $H_1 : \mu \neq 47.5$

calculation of sample mean and standard deviation.

X	$X - \bar{X}$	$(X - \bar{X})^2$
45	-4.1	16.81
47	-2.1	4.41
50	0.9	0.81
52	2.9	8.41
48	-1.1	1.21
47	-2.1	4.41
49	-0.1	0.01

Scanned with OKEN Scanner

53	3.9	15.21
	1.9	3.61
51		
442	0.1	54.89

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

$\bar{x} = 44.2$, $n = 9$, $\mu = 47.5$; $s^2 = \frac{54.89}{8}$

$= 6.86.$

$$|t| = \frac{|44.2-47.5|}{\sqrt{6.86}/\sqrt{9}} = \frac{3.3}{0.873}$$

$= 3.78.$

Since $|t| > t_{0.05} = 2.31$, null hypothesis is rejected and the mean of nine items differ significantly from the assumed papulation mean of 47.5.

* Test of Significance for Difference between Two means (of Independent samples)

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

(3) The means of Two random samples of size 9 and 7 are 196.42 and 198.82 respectively. The sum of the squares of the deviation from the mean are 26.94 and 18.73 respectively. can the samples be considered to have been drawn from the same normal papulation.

Given $n_1 = 9$, $n_2 = 7$

$\bar{x} = 196.42$, $\bar{y} = 198.82$.

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

Null Hypothesis H_0. Samples can be considered to have been drawn from the same normal papulation i.e., $\mu_1 = \mu_2$.

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

$$= \frac{26.94+18.73}{14}$$

$$= 3.262 \text{ and } \quad s = 1.806.$$

$$|t_{cal}| = \frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{|196.42-198.82|}{1.806\left(\sqrt{\frac{1}{9}+\frac{1}{7}}\right)}$$

$$= \frac{2.4}{0.91} = 2.637.$$

For $\gamma=14$, $t_{0.05} = 2.145$. since $t_{cal} > t_{0.05}$ The null Hypothesis is rejected and the samples cannot be considered to have been drawn from the same normal papulation.

* Test of significance for difference between two means (Dependent samples or paired observations).

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

Standard deviation, $s = \sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}} = \sqrt{\frac{\Sigma d^2 - n(\bar{d})^2}{n-1}}$

(4) Find the least value of r in a sample of 18 pairs of observations from a bi-variate normal papulation, significant at 5% level of significance.

Given $n=18$ and for $\gamma=16$, $t_{0.05} = 2.12$.

Test statistic, $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

In order That The calculated value of t is significant at 5% level of significance, we should have

Scanned with OKEN Scanner

$$\left|\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\right| > t_{0.05}$$

$$\left|\frac{r\sqrt{16}}{\sqrt{1-r^2}}\right| > 2.12$$

$$16r^2 > (2.12)^2(1-r^2).$$

$$r^2 > 0.2192$$

Hence $|r| > 0.4682$.

* Chi - square Distribution:

The probability function of χ^2 distribution is given by

$$f(\chi^2) = c\,(\chi^2)^{(\gamma/2-1)} \cdot e^{-\chi^2/2}$$

* Chi-Square Test for Goodness Of Fit.

Test statistic $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$

(5) The number of accidents per week in a city are as follous. 12, 8, 20, 2, 14, 10, 15, 6, 9 and 4. Are These Frequencies in arrangement with the belief that accident conditions were the Same during This 10 weeks period.

let us Take The Hypothesis Ho to be the accident conditions were the same during 10 weeks period.

O	E	O-E	$(O-E)^2/E$
12	10	+2	0.4
8	10	-2	0.4
20	10	+10	10
2	10	-8	6.4
14	10	+4	1.6
10	10	0	0
15	10	+5	2.5

Scanned with OKEN Scanner

6	10	-4	1.6
9	10	-1	0.1
4	10	-6	3.6
Total	100		26.6

The Expected Frequency, $E = \frac{\Sigma O_i}{n} = \frac{100}{10} = 10.$

Test statistic, $\chi^2_{cal} = \frac{\Sigma (O_i - E_i)^2}{E_i} = 26.6$

From the Tables, for $n=10$, degrees of freedom, $\gamma = n-1 = 9$ and $\chi^2_{0.05} = 16.919$

since $\chi^2_{cal} > \chi^2_{0.05}$, Hypothesis is rejected. i.e., The accident condition were not the same during 10 weeks period.

Scanned with OKEN Scanner