1/8/2022

# UNIT-V

## SMALL SAMPLE TEST

If the size of the sample is less than 30 i.e, $n < 30$ that samples are called small samples.

In this small samples we have three types of tests

1. t-test
2. F-test
3. $\chi^2$-test.

1. t-test:

In t-test, we have three important tests for small samples.

(i) student's t-test (single mean)

(ii) student's t-test (when S.D of the sample is not given test for differences of means)

(iii) Paired sample t-test.

2. F-test (Equality of variance)

3. $\chi^2$-test: (chi square test)

(i) $\chi^2$-test of goodness of fit

(ii) $\chi^2$-test for independence of attributes.

Degree of freedom (D.f):

The no. of independent variables which make up the statistic is known as the degree of freedom (d.f) and it is denoted by v.

In general, the degree of freedom is equal to the total no. of observations less the no. of independent constraints imposed on the observations.

For example, in a set of data of n observations if k is the no. of independent constraints then $v = n-k$

I. Student's t-test:

1. Test of significance for single mean.

1. Null hypothesis $H_0: \mu = \mu_0$

2. Alternative hypothesis $H_1: \mu \neq \mu_0$ (two tailed)

$H_1: \mu > \mu_0$ (or) $\mu < \mu_0$ (one tailed)

3. The level of significance ($\alpha$)

4. Test of statistic

i. If sampl standard deviation of sample s is given then

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

ii. If standard deviation is not known then

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

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

Here $\Sigma(x_i - \bar{x})^2$ = sum of square of difference from mean

5. Conclusion

Degree of freedom $v = n-1$

Calculate the value of t tabulated at $\alpha$ at degree of freedom v.

If |t| calculated is less than t tabulated then we accept the null hypothesis.

If calculated |t| is greater than t tabulated then we reject the null hypothesis.

Note:

Confidence interval (or) Fiducial limits for $\mu$;

Confidence interval $C.I = \left[\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right]$

$$C.I = \bar{x} \pm E$$

where $E = t_{\alpha/2}\frac{s}{\sqrt{n}}$

1. A mechanist is making engine parts with diameters of 0.7 inches. A random sample of 10 parts shows a mean diameter of 0.742 inches with a S.D 0.04 inches. Compute the statistic you would use to test whether the work is meeting the specification at 0.05 levael of significance.

Given that,

sample size$(n) = 10$

sample mean$(\bar{x}) = 0.742$

sample S.D $(s) = 0.04$

Population mean$(\mu) = 0.7$.

1. Null hypothesis $H_0 : \mu = 0.7$
2. Alternative hypothesis $H_1 : \mu \neq 0.7$ (Two tailed
3. level of significance

$\alpha = 0.05 = 5\%$.

4. Test of statistic

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

$$= \frac{0.742 - 0.7}{\frac{0.04}{\sqrt{10-1}}}$$

$$= \frac{0.042}{\frac{0.04}{3}}$$

$$t = 3.15$$

5. Conclusion

At $\alpha = 5\%$.

degree of freedom $= n-1 = 10-1 = 9$.

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

t tabulated $t_{\alpha/2} = 2.262$

$|t| = 3.15$

$|t| > t_{\alpha/2}$

i.e, $3.15 > 2.262$

$\therefore$ The null hypothesis is rejected.

i.e, $\mu \neq 0.7$.

2. A random sample of 6 steel beams has a mean Compressive strength 58,392 lbs per square inch with a S.D of 648 psi. Use the information, the level significance $\alpha = 0.05$ to test whether the true average compressive strength of the steel for which this sample came is 58000 psi.

Given that,

$n = 6$

$\bar{x} = 58392$

$s = 648$

$\mu = 58000$

$\alpha = 0.05 = 5\%$

1. Null hypothesis $H_0 : \mu = 58000$
2. Alternative hypothesis $H_1 : \mu \neq 58000$ (Two tailed)
3. Level of significance

$\alpha = 0.05 = 5\%$

4. Test of statistic

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

$$= \frac{58392 - 58000}{\frac{648}{\sqrt{6-1}}} = 1.35$$

5. Conclusion

The degree of freedom $v = n-1 = 6-1 = 5$

$\alpha = 0.05$

$\alpha/2 = 0.025$

$t_{\alpha/2} = 2.571.$

$|t| = 1.35$

$|t| < t_{\alpha/2}$

i.e, $1.35 < 2.571$

$\therefore$ The null hypothesis is accepted.

i.e, $\mu = 58000.$

3. The mean lifetime of a sample of 25 light bulbs produced by a company is computed to be 1570 hours with S.D of 120 hours. The company claims that the average life of the bulbs produced by a company is 1600hrs. Using level of significance $\alpha = 0.0$[illegible]. Is the claim acceptable?

Given that,

$n = 25$

$\bar{x} = 1570$

$S = 120$

$\mu = 1600$

$\alpha = 0.05 = 5\%.$

1. Null hypothesis $H_0: \mu = 1600$
2. Alternative hypothesis $H_1: \mu \neq 1600$ (Two tailed
3. level of significance
   $\alpha = 0.05 = 5\%.$
4. Test of statistic

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

$$= \frac{1570 - 1600}{\frac{120}{\sqrt{25-1}}} = -1.22$$

5. Conclusion:

degree of freedom $\nu = n-1 = 25-1 = 24$.

$\alpha = 0.05$

$\alpha/2 = 0.025$.

$t_{\alpha/2} = 2.064$.

$|t| = 1.22$

$|t| < t_{\alpha/2}$

i.e, $1.22 < 2.064$

$\therefore$ Null hypothesis is accepted.

i.e, $\mu = 1600$.

4. A random sample of size 16 values from a normal population showed a mean of 53 and sum of squares of deviations from the mean is equal to 150. Can this sample be regarded as taken from the population 56 as mean. Obtain 95% Confidence limits of the mean population.

Given that,

$n = 16$

$\bar{x} = 53$

$\Sigma(x_i - \bar{x})^2 = 150$

$\mu = 56$

we know that, $s^2 = \dfrac{\Sigma(x_i-\bar{x})^2}{n-1}$

$= \dfrac{150}{16-1}$

$s^2 = 10$

$s = \sqrt{10} = 3.16$.

$\boxed{s = 3.16}$

1. Null hypothesis $H_0: \mu = 56$

2. Alternative hypothesis $H_1: \mu \neq 56$ (Two tailed)

3. level of significance

$$\alpha = 0.05 = 5\%.$$

4. Test of statistic (s is not given in question directly)

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

$$= \frac{53 - 56}{\frac{3.16}{\sqrt{16}}}$$

$$= -3.79$$

5. Conclusion

degree of freedom $v = n - 1 = 16 - 1 = 15$

$|t| = 3.79$

$\alpha = 0.05$

$\alpha/2 = 0.025$.

$t_{\alpha/2} = 2.131$

$|t| > t_{\alpha/2}$

i.e, $3.79 > 2.131$

$\therefore$ Null hypothesis is rejected.

i.e, $\mu \neq 56$.

Confidence limits are $\bar{x} \pm E$

$$E = t_{\alpha/2} \frac{s}{\sqrt{n}}.$$

$$C.I = [\bar{x} - E, \bar{x} + E]$$

$$= \left[53 - \frac{(2.131)(3.16)}{\sqrt{16}}, 53 + \frac{(2.131)(3.16)}{\sqrt{16}}\right]$$

$$= [53 - 1.6834, 53 + 1.6834]$$

$$= [51.3166, 54.6834].$$

5. The following are the times between 6 calls for an ambulance in a city and the patients arrival at the hospital 27, 15, 20, 32, 18, 26 min use this figures to judge the reasonableness of the ambulance services claim that it takes on the average 20 min b/w the call for an ambulance and patients arrival at hospital.

Given that, $n=6$.

| $x_i$ | $x_i-\bar{x}$ | $(x_i-\bar{x})^2$ |
|---|---|---|
| 27 | 4 | 16 |
| 15 | -8 | 64 |
| 20 | -3 | 9 |
| 32 | 9 | 81 |
| 18 | -5 | 25 |
| 26 | 3 | 9 |
| | 0 | 204. |

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{138}{6} = 23$$

$$s^2 = \frac{\Sigma(x_i-\bar{x})^2}{n-1} = \frac{204}{6-1} = 40.8.$$

$$s = \sqrt{40.8} = 6.38$$

$$\boxed{s=6.38}$$

Given that,

$\mu = 20$.

$n=6$

$\bar{x}=23$

$s=6.38$

1. Null hypothesis $H_0 : \mu = 20$

2. Alternative hypothesis $H_1 : \mu \neq 20$ (Two tailed)

3. level of significance

$\alpha = 0.05 = 5\%$.

4. Test on statistic

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

$$= \frac{23 - 20}{\frac{6.38}{\sqrt{6}}}$$

$$= 1.15$$

5. Conclusion

level of freedom $v = n-1 = 6-1 = 5$

$\alpha = 0.05$

$\alpha/2 = 0.025$

$t_{\alpha/2} = 2.571$

$|t| = 1.15$

$|t| < t_{\alpha/2}$

i.e, $1.15 < 2.571$

$\therefore$ The Null hypothesis is accepted.

$\mu = 20$.

6. A random sample of 10 boys had the following IQ's 70, 120, 110, 101, 88, 83, 95, 98, 107, 100.

i. Do these data support assumption of a population mean IQ of 100.

ii. Find the reasonable range (C.I) in which most of the mean IQ values of a sample of 10 boys.

Given that,

$n = 10$

70, 120, 110, 101, 88, 83, 95, 98, 107, 100.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{972}{10} = 97.2.$$

| $x_i$ | $(x_i-\bar{x})$ | $(x_i-\bar{x})^2$ |
|---|---|---|
| 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 | 0.8 | 0.64 |
| 107 | 9.8 | 96.04 |
| 100 | 2.8 | 7.84 |

$$\Sigma(x_i-\bar{x})^2 = 1833.6$$

$$S^2 = \frac{\Sigma(x_i-\bar{x})^2}{n-1} = \frac{1833.6}{10-1} = 203.73.$$

$$S = \sqrt{203.73} = 14.27.$$

1. Null hypothesis $H_0: \mu = 100$

2. Alternative hypothesis $H_1: \mu \neq 100$ (Two tailed).

3. Level of significance

$\alpha = 0.05 = 5\%$.

4. Test of statistic

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

$$= \frac{97.2-100}{\frac{14.27}{\sqrt{10}}}$$

$$= -0.62.$$

5. Conclusion

level of freedom $v = n-1 = 10-1 = 9$.

$|t| = 0.62$

$\alpha = 0.05$

$\alpha/2 = \frac{0.05}{2} = 0.025.$

$t_{\alpha/2} = 2.262.$

$|t| < t_{\alpha/2}$

i.e, $0.62 < 2.262$

$\therefore$ the null hypothesis is accepted.

i.e, $\mu = 100$.

ii. confidence interval is $\bar{x} \pm E$

$E = t_{\alpha/2} \frac{s}{\sqrt{n}}$

$= \frac{2.262(14.27)}{\sqrt{10}}$

$= 10.207$

$C.I = [\bar{x} - E, \bar{x} + E]$

$= [97.2 - 10.207, 97.2 + 10.207]$

$= [86.993, 107.407]$

7. The following values given the lengths of 12 samples of egyptian cotton taken from a shipment: 48, 46, 49, 46, 52, 45, 43, 47, 47, 46, 45, 50. Test if the mean length of the shipment can be taken as 46. use a 0.05 level of significance.

Given that, $n = 12$

| $x_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ |
|---|---|---|
| 48 | 1 | 1 |
| 46 | -1 | 1 |
| 49 | 2 | 4 |
| 46 | -1 | 1 |
| 52 | 5 | 25 |
| 45 | -2 | 4 |
| 43 | -4 | 16 |
| 47 | 0 | 0 |
| 47 | 0 | 0 |
| 46 | -1 | 1 |
| 45 | -2 | 4 |
| 50 | 3 | 9 |

$\sum(x_i - \bar{x})^2 = 66.$

$\mu = 46$

$\bar{x} = \frac{\sum x_i}{n} = 47$

$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{66}{12-1} = 6$

$s = \sqrt{6} = 2.4495$

1. Null hypothesis $H_0 : \mu = 46$
2. Alternative hypothesis $H_1 : \mu \neq 46$ (Two tailed)
3. level of significance

$\alpha = 0.05 = 5\%$

4. Test of statistic

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

$$= \frac{47 - 46}{\frac{2.4495}{\sqrt{12}}}$$

$t = 1.4142$

5. Conclusion

$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

degree of freedom $(v) = n-1 = 12-1 = 11$

$t_{\alpha/2} = 2.201$

$|t| < t_{\alpha/2}$

$\therefore$ The null hypothesis is accepted.

i.e, $\boxed{\mu = 46}$

Testing procedure significance of difference of means:

1. Null hypothesis $H_0: \mu_1 = \mu_2$

2. Alternative hypothesis $H_1: \mu_1 \neq \mu_2$ (Two tailed)

$H_1: \mu_1 > \mu_2$ (one tailed)

$H_1: \mu_1 < \mu_2$ (one tailed)

3. Level of significance ($\alpha$)

4. Test of statistic -

a) If standard deviations $s_1$ and $s_2$ are known

then $s^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

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

b) If S.D's $s_1$ and $s_2$ are not known

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

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

5. Conclusion

Degree of freedom($v$) $= n_1 + n_2 - 2$

calculate $t$ value at $v$ and $\alpha$ (or) $\alpha/2$.

If $|t| < t_\alpha$, then the null hypothesis is accepted.

If $|t| > t_\alpha$, then the null hypothesis is rejected.

1. Find the maximum difference that we can expect the probability 0.95 between the means of samples of sizes 10 and 12 from a normal population if their standard deviations are found to be 2 and 3 respectively.

Given that,

$n_1 = 10 \qquad n_2 = 12$

$s_1 = 2 \qquad s_2 = 3.$

and given probability $= 0.95$

$1-\alpha = 0.95$

$\alpha = 0.05$

we know that,

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

$$= \frac{10(4) + 12(9)}{10 + 12 - 2}$$

$$\boxed{s^2 = 7.4}$$

i. Null hypothesis $H_0 : \mu_1 = \mu_2$

ii. Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$ (Two tailed level of test significance)

iii. level of significance $(\alpha) = 0.05$

iv. Test of statistic

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

For $\alpha = 0.05$

$\alpha/2 = 0.025.$

degree of freedom $(v) = n_1 + n_2 - 2$

$= 10 + 12 - 2$

$\boxed{v = 20}$

In two tailed, $t_{\alpha/2} = 2.086 = t$

From ①,

$$2.086 = \frac{\bar{x} - \bar{y}}{\sqrt{(7.4)\left(\frac{1}{10}\right) + \frac{1}{12})}}$$

$\therefore \boxed{\bar{x} - \bar{y} = 2.4297}$

$\therefore$ Maximum difference of means $= 2.4297$

2. Two horses A and B were tested according to the time to run ~~to~~ a particular track with the following result.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Horse A | 28 | 30 | 32 | 33 | 33 | 29 | 34 |
| Horse B | 29 | 30 | 30 | 24 | 27 | 29 | |

Test whether the two horses have the same running capacity.

Given that,

$n_1 = 7,\ n_2 = 6.$

| $x_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | $y_i$ | $y_i - \bar{y}$ | $(y_i - \bar{y})^2$ |
|---|---|---|---|---|---|
| 28 | -3.29 | 10.8241 | 29 | 0.83 | 0.6889 |
| 30 | -1.29 | 1.6641 | 30 | 1.83 | 3.3489 |
| 32 | 0.71 | 0.5041 | 30 | 1.83 | 3.3489 |
| 33 | 1.71 | 2.9241 | 24 | 4.17 | 17.3889 |
| 33 | 1.71 | 2.9241 | 27 | -1.17 | 1.3689 |
| 29 | -2.29 | 5.2441 | 29 | 0.83 | 0.6889 |
| 34 | 2.71 | 7.3441 | | | 26.8334 |
| | | 31.4287 | | | |

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{219}{7} = 31.29$$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{169}{6} = 28.17$$

We know that,

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

$$s^2 = \frac{31.4287 + 26.8334}{7 + 6 - 2}$$

$$s^2 = 5.2965$$

i. Null hypothesis $H_0 : \mu_1 = \mu_2$

ii. Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$ (Two tailed test).

iii. level of significance $\alpha = 0.05$

iv. Test of statistic

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

$$= \frac{31.29 - 28.19}{\sqrt{5.2965\left(\frac{1}{7} + \frac{1}{6}\right)}}$$

$$t = 2.4211$$

v. Conclusion : $v = n_1 + n_2 - 2 = 11$

$\alpha = 0.05$

$\alpha/2 = 0.025$

$t_{\alpha/2} = 2.201$

$|t| \not\geq t_{\alpha/2}$

$\therefore$ Null hypothesis is ~~accepted~~ rejected. i.e, $\mu_1 \neq \mu_2$

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

| Type 1 | Type 2 |
|---|---|
| $n_1 = 8$ | $n_2 = 7$ |
| $\bar{x} = 1234$ hrs | $\bar{y} = 1036$ hrs |
| $s_1 = 36$ hrs | $s_2 = 40$ hrs. |

Is the difference in the means sufficient to that type 1 is superior to type 2 regarding length of life.

Given that,

$n_1 = 8 \quad n_2 = 7$

$\bar{x} = 1234 \quad \bar{y} = 1036$

$s_1 = 36 \quad s_2 = 40$

We know that,

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

$$= \frac{8(36)^2 + 7(40)^2}{8 + 7 - 2}$$

$$s^2 = \cancel{21568} \quad 1659.0769$$

i. Null hypothesis $H_0: \mu_1 = \mu_2$

ii. Alternative hypothesis $H_1: \mu_1 \not\geq \mu_2$ (Two tailed test).

iii. level of significance $\alpha = 0.05$

iv. Test of statistic

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

$$= \frac{1234 - 1036}{\sqrt{1659.0769\left(\frac{1}{8} + \frac{1}{7}\right)}} = 9.3925$$

v. Conclusion

degree of freedom$(v) = n_1 + n_2 - 2 = 8 + 7 - 2 = 13$

$\alpha = 0.05$, $\alpha/2 = 0.025$ $t_\alpha = 1.771$.

$|t| > t_\alpha$

$\therefore$ The null hypothesis is rejected.

4. A group of 5 patients treated with medicine A weight 42, 39, 48, 60, 41. Second group of 7 patients treated from the same hospital with medicine B weight 38, 42, 56, 64, 68, 69, 62. Do you agree with the claim that medicine B increases weight significantly.

Given that,

$n_1 = 5$, $n_2 = 7$

| $x_i$ | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ | $y_i$ | $y_i - \bar{y}$ | $(y_i - \bar{y})^2$ |
|---|---|---|---|---|---|
| 42 | -4 | 16 | 38 | -19 | 361 |
| 39 | -7 | 49 | 42 | -15 | 225 |
| 48 | 2 | 4 | 56 | -1 | 1 |
| 60 | 14 | 196 | 64 | 7 | 49 |
| 41 | -5 | 25 | 68 | 11 | 121 |
| | | 290 | 69 | 12 | 144 |
| | | | 62 | 5 | 25 |
| | | | | | 926 |

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{230}{5} = 46$.

$\bar{y} = \frac{\Sigma y_i}{n} = \frac{399}{7} = 57$

we know that,

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

$$s^2 = \frac{290 + 926}{5 + 7 - 2} = 121.6.$$

i. Null hypothesis $H_0: \mu_1 = \mu_2$

ii. Alternative hypothesis $H_1: \mu_1 < \mu_2$ (one tailed)

iii. level of significance

$\alpha = 0.05$

iv. Test of statistic

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

$$= \frac{46 - 57}{\sqrt{121.6\left(\frac{1}{5} + \frac{1}{7}\right)}}$$

$t = -1.7036$

$|t| = 1.7036.$

v. Conclusion

degree of freedom $(v) = n_1 + n_2 - 2 = 10$

$\alpha = 0.05.$

$t_\alpha = 1.812.$

$|t| < t_\alpha$

$\therefore$ The null hypothesis is accepted.

i.e, Medicine B increases weight significantly.

5. Measuring specimens of nylon yarn taken from 2 machines, it was found that 8 specimens from first ~~massion~~ machine had a mean 9.67 with S.D of 1.81 while 10 specimens from second machine had a mean of 7.43 with S.D of 1.48. Assuming that the populations are normal, test the hypothesis $H_0$ is $\mu_1 - \mu_2 = 1.5$ against $H_1: \mu_1 - \mu_2 > 1.5$ at 0.05 level of significance.

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

$\bar{x}_1 = 9.67$ $\qquad$ ~~$x_2$~~ $\bar{y} = 7.43.$

$s_1 = 1.81$ $\qquad s_2 = 1.48.$

we know that,

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

$$s^2 = \frac{8(1.81)^2 + 10(1.48)^2}{8+10-2}$$

$$\boxed{s^2 = 3.00705}.$$

i. Null hypothesis is $H_0 : \mu_1 - \mu_2 = 1.5$

ii. Alternative hypothesis $H_1 : \mu_1 - \mu_2 > 1.5$ (one tailed)

iii. level of significance

$\alpha = 0.05$

iv. Test of statistic ($\because \delta = \mu_1 - \mu_2 = 1.5$)

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

$$t = \frac{9.67 - 7.43 - 1.5}{\sqrt{3.00705\left(\frac{1}{8} + \frac{1}{10}\right)}}$$

$t = 0.8996.$

v. Conclusion

degree of freedom $(v) = n_1 + n_2 - 2 = 8 + 10 - 2 = 16.$

$\alpha = 0.05$

$t_\alpha = 1.746.$

Here, $|t| < t_\alpha$,

i.e, $0.8996 < 1.746.$

$\therefore$ The null hypothesis is accepted.

i.e, $\mu_1 - \mu_2 = 1.5.$

Paired t-test:

Let $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ be the pairs of sales data before and after the sales promotion in a business concern. we apply the paired t-test to examine the significance of difference of two situations.

Let $d_i = x_i - y_i$ (or) $y_i - x_i$

Let the null hypothesis $H_0 : \mu_1 = \mu_2$

Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$

level of significance $(\alpha)$

Test of statistic

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

$$\bar{d} = \frac{\Sigma d}{n}$$

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

Conclusion:

Degree of freedom $v = n-1$

calculate $t_\alpha$ t values at $v$ and $\alpha$.

If $|t| < t_\alpha$ then we accept the null hypothesis

If $|t| > t_\alpha$ then we reject the null hypothesis

1. Memory capacity of 10 students ~~ca~~ were tested before and after training. State whether the training was effective or not from the following source.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Before training : | 12 | 14 | 11 | 8 | 7 | 10 | 3 | 0 | 5 | 6 |
| After training : | 15 | 16 | 10 | 7 | 5 | 12 | 10 | 2 | 3 | 8. |

Given,

$n = 10$

| $x_i$ | $y_i$ | $d$ | $(d-\bar{d})$ | $(d-\bar{d})^2$ |
|---|---|---|---|---|
| 12 | 15 | -3 | -1.8 | 3.24 |
| 14 | 16 | -2 | -0.8 | 0.64 |
| 11 | 10 | 1 | 2.2 | 4.84 |
| 8 | 7 | 1 | 2.2 | 4.84 |
| 7 | 5 | 2 | 3.2 | 10.24 |
| 10 | 12 | -2 | -0.8 | 0.64 |
| 3 | 10 | -7 | -5.8 | 33.64 |
| 0 | 2 | -2 | -0.8 | 0.64 |
| 5 | 3 | 2 | 3.2 | 10.24 |
| 6 | 8 | -2 | -0.8 | 0.64 |
| | | -12 | | 69.6. |

$$\bar{d} = \frac{\Sigma d}{n} = \cancel{\frac{-24}{10}} = \cancel{-2.4}$$

$$= \frac{-12}{10} = -1.2$$

$$S^2 = \frac{\Sigma(d-\bar{d})^2}{n-1} = \frac{69.6}{10-1} = 7.73. \quad S = \sqrt{7.73} = 2.78$$

i. Null hypothesis $H_0: \mu_1 = \mu_2$

ii. Alternative hypothesis $H_1: \mu_1 < \mu_2$ (one ~~two~~ tailed)

iii. level of significance

$\alpha = 0.05$

iv. Test of statistic

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

$$t = \frac{-1.2}{\frac{2.78}{\sqrt{10}}}$$

$$t = -1.36$$

$$|t| = 1.36$$

v. Conclusion

degree of freedom $(v) = n-1 = 10-1 = 9$

$\alpha = 0.05$

~~$\alpha/2 = 0.025$~~ $t_\alpha = 1.833$.

~~$t_{\alpha/2} = 2.262$~~

$\therefore |t| < t_\alpha$

i.e, ~~$1.36 < 2.262$~~ $1.36 < 1.833$

$\therefore$ The null hypothesis is accepted.

2. The blood pressure of 5 women before and after intake of a certain drug are given below

Before: 110 120 125 132 125

After: 120 118 125 136 121

Test whether there is significant change in blood pressure at 1% level of significance.

Given that, $n = 5$

| $x$ | $y$ | $d$ | $d-\bar{d}$ | $(d-\bar{d})^2$ |
|---|---|---|---|---|
| 110 | 120 | -10 | -8.4 | 70.56 |
| 120 | 118 | 2 | 3.6 | 12.96 |
| 125 | 125 | 0 | 1.6 | 2.56 |
| 132 | 136 | -4 | -2.4 | 5.76 |
| 125 | 121 | 4 | 5.6 | 31.36 |
| | | -8 | | 123.2 |

$\bar{d} = \frac{\Sigma d}{n} = \frac{-8}{5} = -1.6$

$s^2 = \frac{\Sigma (d-\bar{d})^2}{n-1} = \frac{123.2}{5-1} = 30.8$

$s = \sqrt{30.8} = 5.549$

i. Null hypothesis $H_0: \mu_1 = \mu_2$

ii. Alternative hypothesis $H_1: \mu_1 < \mu_2$

iii. level of significance

$\alpha = 0.01$

iv. Test of statistic

$$t = \frac{\bar{d}}{\frac{s}{\sqrt{n}}} = \frac{-1.6}{\frac{5.549}{\sqrt{5}}} = -0.6447$$

$|t| = 0.6447$

v. Conclusion

degree of freedom $(v) = n-1 = 4$

$\alpha = 0.01$

$t_\alpha = 3.747$

$|t| < t_\alpha$

i.e, $0.6447 < 3.747$

low

$\therefore$ The null hypothesis is accepted.

F - test (Equality of variance):

Let two independent random samples of sizes $n_1$ and $n_2$ be drawn from two normal populations. To test the hypothesis that the two population variances $\sigma_1^2$ and $\sigma_2^2$ are equal.

1. Null hypothesis $H_0: \sigma_1^2 = \sigma_2^2$
2. Alternative hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$

$\sigma_1^2 < \sigma_2^2$ or $\sigma_1^2 > \sigma_2^2$

3. level of significance $\alpha$ (5% or 1%)
4. Test of statistic

$$F = \frac{S_1^2}{S_2^2} \qquad (\text{if } S_1^2 > S_2^2)$$

(or)

$$F = \frac{S_2^2}{S_1^2} \qquad (\text{if } S_2^2 > S_1^2)$$

$$F = \frac{\text{Greater value}}{\text{smaller value}}$$

Here $S_1^2$ and $S_2^2$ are population variances.

To find $S_1^2 = \dfrac{n_1 s_1^2}{n_1 - 1}$ (or) $\dfrac{\Sigma(x_i - \bar{x})^2}{n-1}$

$S_2^2 = \dfrac{n_2 s_2^2}{n_2 - 1}$ (or) $\dfrac{\Sigma(y_i - \bar{y})^2}{n-1}$

Here $s_1^2$ and $s_2^2$ are sample variances.

5. Conclusion

Degree of freedom $(v_1, v_2) = (n_1 - 1, n_2 - 1)$

To calculate F value at $v_1, v_2$ and

if $|F| < F_\alpha$ we accept null hypothesis

If $|F| > F_\alpha$ we reject null hypothesis

1. Pumpkins were grown under 2 experimental Conditions. Two random samples of 11 & 9 pumpkins, show the S.D of their weight as 0.8 and 0.5 respectively. Assuming that the weight distributions are normal, test hypothesis that the true variances are equal

Given : $n_1 = 11$ $n_2 = 9$.

$s_1 = 0.8$ $s_2 = 0.5$.

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{11(0.8)^2}{11-1} = 0.704$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{9(0.5)^2}{9-1} = 0.28125$$

1. Null hypothesis $H_0: \sigma_1^2 = \sigma_2^2$
2. Alternative hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$.
3. level of significance

$\alpha = 0.05$

4. Test of statistic $F = \frac{S_1^2}{S_2^2}$ $(S_1^2 > S_2^2)$

$$= \frac{0.704}{0.28125} = 2.5031.$$

5. conclusion

Degree of freedom $(v_1, v_2) = (n_1 - 1, n_2 - 1)$

$= (10, 8)$

$\alpha = 0.05$

$F\alpha =$

$|F| < F\alpha$

$\therefore$ Null hypothesis is accepted

$\therefore \sigma_1^2 = \sigma_2^2$.

2. In the sample of 10 observations, the sum of squares of deviations from mean was 130 and th another sample of 12 observations, it was 314. Test whether the difference is significance at 5

Given that,

$n_1 = 10$, $n_2 = 10$

$\Sigma(x_i - \bar{x})^2 = 130$

$\Sigma(y_i - \bar{y})^2 = 314$

$$S_1^2 = \frac{\Sigma(x_i - \bar{x})^2}{n_1 - 1} = \frac{130}{9} = 13.333$$

$$S_2^2 = \frac{\Sigma(y_i - \bar{y})^2}{n_2 - 1} = \frac{314}{12 - 1} = 28.5454.$$

1. Null hypothesis $H_0: \sigma_1^2 = \sigma_2^2$

2. Alternative hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$

3. level of significance

$$\alpha = 0.01$$

4. Test of statistic

$$F = \frac{S_2^2}{S_1^2} \qquad (S_2^2 > S_1^2)$$

$$F = \frac{28.5454}{13.33}$$

$$\boxed{F = 2.1414}$$

5. Conclusion

$V_1 = 10 - 1 = 9$

$V_2 = 12 - 1 = 10.$

$(V_1, V_2) = (9, 10)$

$\alpha = 0.05$

$F_{(9,10)}$

$|F| < F_\alpha$

$\therefore$ Null hypothesis accepted.

3. In one sample of 8 observations from a normal population, the sum of the squares of deviations of the sample values from the sample mean is 84.4 and in other sample of 10 observations it was 102.6. Test at 5% level whether the population have the same variance.

Given that, $n_1 = 8, n_2 = 10$

$$\Sigma(x_i - \bar{x})^2 = 84.4$$

$$\Sigma(y_i - \bar{y})^2 = 102.6$$

$$S_1^2 = \frac{\Sigma(x_i - \bar{x})^2}{n_1 - 1} = \frac{84.4}{7} = 12.0511$$

$$S_2^2 = \frac{\Sigma(y_i - \bar{y})^2}{n_2 - 1} = \frac{102.6}{9} = 11.4.$$

i. Null hypothesis, $H_0: \sigma_1^2 = \sigma_2^2$

ii. Alternative hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$

iii. level of significance

$$\alpha = 0.05$$

iv. Test of statistic

$$F = \frac{S_1^2}{S_2^2}$$

$$= \frac{12.0511}{11.4}$$

$$\boxed{F = 1.0576}$$

Conclusion:

$v_1 = 7,\ v_2 = 9.$

$(v_1, v_2) = (7, 9)$

$\alpha = 0.05$

$F_\alpha = 3.27$

$|F| < F_\alpha$

$\therefore$ Null hypothesis is accepted.

4. Two random samples for the following result.

| sample | size | sample mean | sum of square of deviations from m- |
|---|---|---|---|
| 1 | 10 | 15 | 90 |
| 2. | 12 | 14 | 108. |

Test whether the samples came from same population (can be done in t-test).

Given that,

$n_1 = 10 \quad \& \quad n_2 = 12$

$s_1 = 15 \quad s_2 = 14$

$\& \ \Sigma(x_i - \bar{x})^2 = 90 \quad \Sigma(y_i - \bar{y})^2 = 100$

$$s_1^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \frac{90}{9} = 10.$$

$$s_2^2 = \frac{\Sigma(y_i - \bar{y})}{n_2 - 1} = \frac{108}{11} = 9.8181.$$

i. Null hypothesis

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

ii. Alternative hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$.

iii. level of significance $\alpha = 0.05$

iv. Test of statistic

$$F = \frac{S_1^2}{S_2^2} = \frac{10}{9.81} = 1.0936$$

Conclusion:

$(v_1, v_2) = (10, 9)$

$\alpha = 0.05$ & $F_\alpha = 2.90$

$|H| < F_\alpha$

∴ the opening times.

∴ The $H_0$ is accepted

i.e, $\sigma_1^2 = \sigma_2^2$.

5 A study is conducted to compare the length of time b/w men & women to assemble a certain product. Past experience indicates that the distributions of times for both men and women is approximately normal but the variance of times for women is less than that for men. A random sample of times for 11 men and 14 women produced the following data.

Test the hypothesis that $\sigma_1^2 = \sigma_2^2$ against the alternative $\sigma_1^2 > \sigma_2^2$. Use 0.05 level of significance.

| | Men | women |
|---|---|---|
| | $n_1 = 11$ | $n_2 = 14$ |
| | $s_1 = 6.1$ | $s_2 = 5.3$ |

Given that,

$n_1 = 11$ $\quad n_2 = 14$

$s_1 = 6.1$ $\quad s_2 = 5.3$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{11(6.1)^2}{11-1} = 40.931$$

$$S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{14(5.3)^2}{14-1} = 30.2507.$$

i. Null hypothesis $H_0: \sigma_1^2 = \sigma_2^2$

ii. Alternative hypothesis $H_1: \sigma_1^2 > \sigma_2^2$

iii. level of significance $\alpha = 0.05$.

iv. Test of statistic

$$F = \frac{S_1^2}{S_2^2} = \frac{40.931}{30.2507} = 1.3530$$

v. Conclusion

Degree of freedom $(v_1, v_2) = (n_1 - 1, n_2 - 1)$

$= (11-1, 14-1)$

$= (10, 13)$

$\alpha = 0.05$

$F_\alpha = 2.61$

$|F| < F_\alpha$

$\therefore$ the null hypothesis $H_0$ is accepted.

$\sigma_1^2 = \sigma_2^2$

6. An instructor has two classes A and B, in a particular subject class A has 16 students, ~~[illegible]~~ while class B has 25 students. On the same examination although there was no significance difference in mean grades class A has S.D of 9 whil class B has S.D of 12 can conclude at 0.01 level of significance that the variability of class B is greater than class A.

Given, $n_1 = 16$ $\quad n_2 = 25$

$S_1 = 9$ $\quad S_2 = 12$

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} \qquad S_2^2 = \frac{n_2 s_2^2}{n_2 - 1}$$

$$= \frac{16(9)^2}{15} \qquad = \frac{25(12)^2}{24}$$

$S_1 = 86.4.$ $\qquad S_2 = 150.$

i. Null hypothesis $\cdot H_0 = \sigma_1^2 = \sigma_2^2$

ii. Alternative hypothesis $H_1 : H_1 = \sigma_2^2 > \sigma_1^2$

iii. level of significance $\alpha = 0.01$

iv. Test of statistic $F = \frac{S_2^2}{S_1^2}$ $[S_2^2 > S_1^2]$.

$$F = \frac{150}{86.4}$$

$$\boxed{F = 1.736}$$

V. Conclusion: Degree of freedom $(V_1, V_2)$

$$= (n_1 - 1, n_2 - 1)$$
$$= (15, 24)$$

$\alpha = 0.05$

$F_\alpha = 2.11$

$|F| < F_\alpha$

i.e, $1.736 < 2.11$

$\therefore$ The null hypothesis $H_0$ is accepted.

i.e, $\sigma_1^2 = \sigma_2^2$.

7. Two independent samples of 8 and 7 items respectively had the following values of the variables.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Sample 1 | 9 | 11 | 13 | 11 | 16 | 10 | 12 | 14 |
| Sample 2 | 11 | 13 | 11 | 14 | 10 | 8 | 10 | |

Do the estimates of the population variance ~~de~~ significance.

$n_1 = 8 \qquad n_2 = 7$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{96}{8} = 12$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{77}{7} = 11$$

$$\therefore \boxed{\bar{x} = 12, \bar{y} = 11}$$

| $x_i$ | $(x_i-\bar{x})$ | $(x_i-\bar{x})$ | $y_i$ | $(y_i-\bar{y})$ | $(y_i-\bar{y})^2$ |
|---|---|---|---|---|---|
| 9 | -3 | 9 | 11 | 0 | 0 |
| 11 | -1 | 1 | 13 | 2 | 4 |
| 13 | 1 | 1 | 11 | 0 | 0 |
| 11 | -1 | 1 | 14 | 3 | 9 |
| 16 | 4 | 16 | 10 | -1 | 1 |
| 10 | -2 | 4 | 8 | -3 | 9 |
| 12 | 0 | 0 | 10 | -1 | 1 |
| 14 | 2 | 4 | | | |

$$S_1^2 = \frac{\Sigma(x_i-\bar{x})^2}{n_1-1} = \frac{36}{8-1} = 5.1428$$

$$S_2^2 = \frac{\Sigma(y_i-\bar{y})^2}{n_2-1} = \frac{24}{7-1} = 4.$$

i. Null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$

ii. Alternative hypothesis $H_1 : \sigma_1^2 \neq \sigma_2^2$

iii. level of significance $\alpha = 0.05$

iv. Test of statistic

$$F = \frac{S_1^2}{S_2^2}$$

$$= \frac{5.1428}{4}$$

$$\boxed{F = 1.285}$$

v. Conclusion

Degree of freedom $(v_1, v_2) = (n_1-1, n_2-1)$

$= (8-1, 7-1)$

$= (7, 6)$

$\alpha = 0.05$

$F_\alpha = 4.21$

$|F| < F_\alpha$

i.e, $1.285 < 4.21$

Null hypothesis $H_0$ is accepted.

$\therefore \sigma_1^2 = \sigma_2^2$.

8. Two samples of tobacco were found to be as follows:

Sample A : 24 27 26 21 25

Sample B : 27 30 28 31 22 36.

Can it be said that the two samples have come from the same normal population.

Given that

$n_1 = 5 \quad n_2 = 6.$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{24+27+26+21+25}{5} = 24.6$$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{27+30+28+31+22+36}{6} = 29.$$

| $x_i$ | $(x_i-\bar{x})$ | $(x_i-\bar{x})^2$ | $y_i$ | $(y_i-\bar{y})$ | $(y_i-\bar{y})^2$ |
|---|---|---|---|---|---|
| 24 | -0.6 | 0.36 | 27 | -2 | 4 |
| 27 | 2.4 | 5.76 | 30 | 1 | 1 |
| 26 | 1.4 | 1.96 | 28 | -1 | 1 |
| 21 | -3.6 | 12.96 | 31 | 2 | 4 |
| 25 | 0.4 | 0.16 | 22 | -7 | 49 |
| | | | 36 | 7 | 49 |

$$S_1^2 = \frac{\Sigma(x_i-\bar{x})^2}{n_1-1} = \frac{21.2}{5-1} = 5.3$$

$$S_2^2 = \frac{\Sigma(y_i-\bar{y})^2}{n_1-1} = \frac{108}{6-1} = 21.6$$

$\therefore S_2^2 > S_1^2$

i. Null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$

ii. Alternative hypothesis $H_1 : \sigma_1^2 \neq \sigma_2^2$

iii. level of significance

$\alpha = 0.05$

iv. Test of statistic

$$F = \frac{S_2^2}{S_1^2} = \frac{21.6}{5.3} = 4.015.$$

v. Conclusion

Degree of freedom $(v_1, v_2) = (n_1 - 1, n_2 - 1)$
$= (5-1, 6-1)$
$= (4, 5)$

$\alpha = 0.05$

$F_\alpha = 5.19$

$|F| < F_\alpha$

$\therefore$ Null hypothesis $H_0$ is accepted.

i.e, $\sigma_1^2 = \sigma_2^2$

chi-square test ($\chi^2$)

If a set of events $A_1, A_2, \ldots, A_n$ are observed with frequencies $O_1, O_2, O_3, \ldots O_n$ respectively.

According to probability rule $A_1, A_2, \ldots A_n$ are expected to occur with frequencies $E_1, E_2, \ldots, E_n$ respectively.

Here $O_1, O_2, O_3, \ldots O_n$ are called observed frequencies and $E_1, E_2, \ldots, E_n$ are called expected frequencies.

Here $\chi^2$ is defined as

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

where $E_i = \dfrac{\text{sum of the observations}}{\text{No. of observations}}$

(or) $E_i = N\,P(x_i)$

Note:

1. If the data is given in series of $n$ numbers then degree of freedom $\boxed{v = n-1}$
2. In case of binomial distribution, $v = n-1$
3. In case of poisson distribution, $v = n-2$.
4. In case of normal distribution, $v = n-3$.

In this we have, two important tests

i. chi-square ($\chi^2$) test for goodness of fit.

ii. Chi-square ($\chi^2$) test for independence of attributes.

i. Chi-square ($\chi^2$) test for goodness of fit:

i. Null hypothesis

$H_0$: There is no significance difference between the observed values and the corresponding expected values.

ii. Alternative hypothesis

$H_1$: The above difference is significant.

iii. level of significance $\alpha$

iv. Test of statistic

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

v. Conclusion

Degree of freedom $v = n-1$

To calculate $\chi^2$ at $v$ and $\alpha$.

If $|\chi^2| < |\chi^2_\alpha|$ then we accept the null hypothesis

if $|\chi^2| > \chi^2_\alpha$ then we reject the null hypothesis.

1. The no. of automobile accidents per week in a certain Community are as follows:

12, 8, 20, 2, 14, 10, 15, 6, 9, 4.

Are these frequencies in agreement with the belief that the accident conditions were the same during this 10 week period.

Given that,

12, 8, 20, 2, 14, 10, 15, 6, 9, 4.

i. Null hypothesis

$H_0$: The accident conditions were the same during 10 week period.

ii. Alternative hypothesis

$H_1$: The accident conditions were different during 10 week period.

iii. level of significance

$\alpha = 0.05$.

iv. Test of statistic.

Expected frequency of accidents each week $E_i = \frac{12+8+20+2+14+10+15+6+9+4}{10} = \frac{100}{10} = 10$.

| observed frequency $O_i$ | Expected frequency $E_i$ | $(O_i - E_i)$ | $\frac{(O_i-E_i)^2}{E_i}$ |
|---|---|---|---|
| 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 |
| 6 | 10 | -4 | 1.6 |
| 9 | 10 | -1 | 0.1 |
| 4 | 10 | -6 | 3.6 |
| | | | 26.6 |

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

$$\boxed{\chi^2 = 26.6}$$

v. Conclusion

Degree of freedom $v = n-1 = 10-1 = 9$

$\alpha = 0.05$.

$\chi^2_\alpha = 16.919$.

$|\chi^2| > \chi^2_\alpha$

$\therefore$ Null hypothesis is rejected.

i.e the accident conditions were different during 10 week period.

2. A dice is thrown 264 times with the following results. show that the dice is biased. $\left[\chi^2_{0.05} = 11.07 \text{ for } v=5\right]$.

| | | | | | | |
|---|---|---|---|---|---|---|
| No. appeared on the dice | : 1 | 2 | 3 | 4 | 5 | 6 |
| Frequency | : 40 | 32 | 28 | 58 | 54 | 52 |

i. Null hypothesis

$H_0$: The dice is biased.

ii. Alternative hypothesis

$H_1$: The dice is unbiased.

iii. level of significance

$\alpha = 0.05$

iv. Test of statistic.

$$E_i = \frac{40+32+28+58+54+52}{6} = \frac{264}{6} = 44$$

| $O_i$ | $E_i$ | $O_i - E_i$ | $\frac{(O_i-E_i)^2}{E_i}$ |
|---|---|---|---|
| 40 | 44 | -4 | 0.364 |
| 32 | 44 | -12 | 3.273 |
| 28 | 44 | -16 | 5.818 |
| 58 | 44 | 14 | 4.455 |
| 54 | 44 | 10 | 2.273 |
| 52 | 44 | 8 | 1.455 |
| | | | 17.638 |

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

$$\boxed{\chi^2 = 17.638}$$

v. Conclusion

Degree of freedom $v = n-1 = 6-1 = 5$

$\alpha = 0.05$

$\chi^2_\alpha = 11.07$

$|\chi^2| > \chi^2_\alpha$

i.e, $17.638 > 11.07$

$\therefore$ Null hypothesis is rejected.

i.e, The dice is unbiased.

3. A sample analysis of Examination results of 500 students was made. It was found to be that 220 students had failed, 170 had scored third class, 90 were placed in second class & 20 got first class. Do these figures commensurate with the general examination result which is in the ratio of 4:3:2:1. For the various categories respectively.

i. Null hypothesis

$H_0$: The observed results commenserate with the general examination results.

ii. Alternative hypothesis

$H_1$: The observed results does not commenserate with the general examination results.

iii. level of significance

$\alpha = 0.05$.

iv. Test of statistic

Expected frequencies are in the ratio 4:3:2:1

Total frequency = 500

If we divide total frequency 500 in the ratio 4:3:2:1. we get the expected frequencies

are ~~12/3~~, 200, 150, 100, 50

| $O_i$ | $E_i$ | $O_i - E_i$ | $(O_i - E_i)^2$ | $\frac{(O_i - E_i)^2}{E_i}$ |
|---|---|---|---|---|
| 220 | 200 | 20 | 400 | 2 |
| 170 | 150 | 20 | 400 | 2.67 |
| 90 | 100 | -10 | 100 | 1 |
| 20 | 50. | -30 | 900 | 18 |
| | | | | 23.67. |

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

$$\chi^2 = 23.67.$$

v. Conclusion:

Degree of freedom $v = n-1 = 4-1 = 3$.

$\alpha = 0.05$.

$\chi^2_{tab} = 7.815$.

$|\chi^2| > \chi^2_{tab}$

$\therefore$ Null hypothesis is rejected.

i.e, the observed results doesnot commenserate with the general examination results.

4. A pair of dice are thrown 360 times and the frequency of the each sum is indicated below

| Sum | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 8 | 24 | 35 | 37 | 44 | 65 | 51 | 42 | 26 | 14 | 14. |

would you say that the dice are fair on the bases of the chi-square test at 0.05 level of significance.

i. Null hypothesis

$H_0$: The dice are fair

ii. Alternative hypothesis

$H_1$: the dice are not fair

iii. level of significance

$\alpha = 0.05$

iv. Test of statistic

The probabilities of getting a sum 2, 3, ..., 12.

| Sum | $P(X)$ | $E(X) = N P(X)$ |
|---|---|---|
| 2 | 1/36 | 10 |
| 3 | 2/36 | 20 |
| 4 | 3/36 | 30 |
| 5 | 4/36 | 40 |
| 6 | 5/36 | 50 |
| 7 | 6/36 | 60 |
| 8 | 5/36 | 50 |
| 9 | 4/36 | 40 |
| 10 | 3/36 | 30 |
| 11 | 2/36 | 20 |
| 12 | 1/36 | 10 |

| Sum | $O_i$ | $E_i$ | $(O_i - E_i)$ | $(O_i - E_i)^2$ | $\frac{(O_i - E_i)^2}{E_i}$ |
|---|---|---|---|---|---|
| 2 | 8 | 10 | -2 | 4 | 0.4 |
| 3 | 24 | 20 | 4 | 16 | 0.8 |
| 4 | 35 | 30 | 5 | 25 | 0.83 |
| 5 | 37 | 40 | -3 | 9 | 0.225 |
| 6 | 44 | 50 | -6 | 36 | 0.72 |
| 7 | 65 | 60 | 5 | 25 | 0.417 |
| 8 | 51 | 50 | 1 | 1 | 0.02 |
| 9 | 42 | 40 | 2 | 4 | 0.1 |
| 10 | 26 | 30 | -4 | 16 | 0.53 |
| 11 | 14 | 20 | -6 | 36 | 1.8 |
| 12 | 14 | 10 | 4 | 16 | 1.6 |
| | | | | | 7.442 |
| | | | | | 7.879. |

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

$$\chi^2 = 7.879.$$

v. Conclusion:

Degree of freedom $(v) = n-1 = 11-1 = 10.$

$\alpha = 0.05$

$\chi^2_{tab} = 18.307.$

$|\chi^2| < \chi^2_{tab}$

$\therefore$ Null hypothesis is accepted.

i.e, the dice are fair.

5. 4 Coins were tossed 160 times and the following results were obtained

| No. of heads | : | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| observed frequency | : | 17 | 52 | 54 | 31 | 6 |

Under the assumption that coins are balanced. Find the expected frequencies of 0,1,2,3,4 heads and test the goodness of fit $\alpha = 0.05$

i. Null hypothesis:

$H_0$: The coins are balanced.

ii. Alternative hypothesis:

$H_1$: The coins are not balanced.

iii. level of significance

$\alpha = 0.05$.

iv. Test of statistic.

4 coins are tossed.

HHHH TTTT
HHHT TTTH
HHTH TTHT
HTHH THTT
THHH HTTT
HHTT TTHH
HTTH THHT
HTHT THTH

Expected no. of frequencies of 0,1,2,3,4 are as follows.

| $x$ | : | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| $P(x)$ | : | $1/16$ | $4/16$ | $6/16$ | $4/16$ | $1/16$ |

| $x$ | $O_i$ | $E_i = NP(x)$ | $(O_i - E_i)$ | $\frac{(O_i-E_i)^2}{E_i}$ |
|---|---|---|---|---|
| 0 | 17 | 10 | 7 | 4.9 |
| 1 | 52 | 40 | 12 | 3.6 |
| 2 | 54 | 60 | 14 | 3.267 |
| 3 | 31 | 40 | -9 | 2.025 |
| 4 | 6 | 10 | -4 | 1.6 |
| | | | | 15.392 |

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

$$\chi^2 = 15.392.$$

v. Conclusion.

Degree of freedom $(v) = n-1 = 4-1 = 3$

$\alpha = 0.05$

$\chi^2_{tab} = 7.815$.

$\therefore \chi^2 > \chi^2_{tab}$

i.e, $15.392 > 7.815$

$\therefore$ The null hypothesis is rejected.

i.e, The coins are not balanced.

6. 200 digits were choosen at random from a set of tables. The frequencies of the digits are as shown below.

| Digit: | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Frequency: | 18 | 19 | 23 | 21 | 16 | 25 | 22 | 20 | 21 | 15 |

Use the chi-square test to asses the correctness of the hypothesis that the digits were distributed in equal numbers in the tables from which these were choosen.

i. Null hypothesis

$H_0$: The digits were distributed equally in the tables.

ii. Alternative hypothesis

$H_1$: The digits were not distributed equally in the tables.

iii. level of significance

$\alpha = 0.05$

iv. Test of statistic

$$E_i = \frac{18+19+23+21+16+25+22+20+21+15}{10}$$

$$= \frac{200}{10} = 20.$$

| $O_i$ | $E_i$ | $(O_i - E_i)$ | $\frac{(O_i-E_i)^2}{E_i}$ |
|---|---|---|---|
| 18 | 20 | -2 | 0.2 |
| 19 | 20 | -1 | 0.05 |
| 23 | 20 | 3 | 0.45 |
| 21 | 20 | 1 | 0.05 |
| 16 | 20 | -4 | 0.8 |
| 25 | 20 | 5 | 1.25 |
| 22 | 20 | 2 | 0.2 |
| 20 | 20 | 0 | 0 |
| 21 | 20 | 1 | 0.05 |
| 15 | 20 | -5 | 1.25 |
| | | | 4.3 |

$$\chi^2 = \Sigma\left[\frac{(O_i-E_i)^2}{E_i}\right] = 4.3$$

iv. Conclusion

Degree of freedom

$v = n-1$

$v = 10-1 = 9$

$\alpha = 0.05$

$\chi^2_{tab} = 16.919$

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

∴ Null hypothesis is accepted.

i.e, the digits were not distributed equally in table.

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

| Digit | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 1026 | 1107 | 997 | 966 | 1075 | 933 | 1107 | 972 | 964 | 853 |

Test whether the digits may be taken to occur equally frequently in the directory.

10000

i. Null hypothesis

$H_0$: The digits may be taken to occur equally frequently in the directory

Alternative

ii. ~~Null~~ hypothesis

$H_1$: " " " not occur " "

iii. level of significance

$\alpha = 0.05$

iv. Test of statistic.

$$E_i = \frac{1026+1107+997+966+1075+933+1107+972+964+8}{10}$$

$$= \frac{10000}{10}$$

$E_i = 1000$.

| $O_i$ | $E_i$ | $O_i - E_i$ | $\frac{(O_i-E_i)^2}{E_i}$ |
|---|---|---|---|
| 1026 | 1000 | 26 | 0.676 |
| 1107 | 1000 | 107 | 11.449 |
| 997 | 1000 | -3 | 0.009 |
| 966 | 1000 | -34 | 1.156 |
| 1075 | 1000 | 75 | 5.625 |
| 933 | 1000 | -67 | 4.489 |
| 1107 | 1000 | 107 | 11.449 |
| 972 | 1000 | -28 | 0.784 |
| 964 | 1000 | -36 | 1.296 |
| 853 | 1000 | -147 | 21.609 |
| | | | 58.542 |

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

$$\chi^2 = 58.542$$

v. Conclusion

$\alpha = 0.05$

$\nu = n-1 = 10-1 = 9$

$\chi^2_{tab} = 16.919$

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

$\therefore$ Null hypothesis is rejected

i.e, The digits may be taken to occur equally frequently in the directory.

8. Fit a poisson distribution to the following data and for goodness of fit at level of significance 0.05.

| | | | | | |
|---|---|---|---|---|---|
| x : | 0 | 1 | 2 | 3 | 4 |
| f : | 419 | 352 | 154 | 56 | 19. |

i. Null hypothesis

$H_0$: The poisson can be fitted to the data is good.

ii Alternative hypothesis

$H_1$: " is not good.

iii. level of significance

$\alpha = 0.05$

iv. Test of statistic

Expecting frequencies by using poisson distribution.

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$\lambda = \frac{\Sigma f_i x_i}{N} = \frac{0+352+308+168+76}{1000}$$

$$= \frac{904}{1000}$$

$$= 0.904$$

| $x$ | $O_i$ | $P(x)$ | $E_i = N(P(x))$ | $O_i - E_i$ | $\frac{(O_i-E_i)^2}{E_i}$ |
|---|---|---|---|---|---|
| 0 | 419 | 0.4049 | 404.9 | 14.1 | 0.491 |
| 1 | 352 | 0.3661 | 366.1 | -14.1 | 0.543 |
| 2 | 154 | 0.1655 | 165.5 | -11.5 | 0.799 |
| 3 | 56 | 0.0499 | 49.9 | 6.1 | 0.746 |
| 4 | 19 | 0.0113 | 11.3 | 7.7 | 5.247 |
| | | | | | 7.826. |

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

$$\chi^2 = 7.826.$$

v. Conclusion

degree of freedom$(v) = n-2$

$v = 5-2 = 3.$

$\alpha = 0.05$

$\chi^2_{tab} = 7.815.$

$|\chi^2| > \chi^2_{tab}$

$\therefore$ Null hypothesis is rejected.

i.e, The poisson distribution can be fitted to the data is good.

Chi-square ($\chi^2$) test for independence of attributes:

Literally, an attribute means a quality or characteristic. Examples of attribute are drinking smoking, beauty, honesty, blindness etc.

An attribute may be marked by its presence or absence in a no. of given population. Let the observations be classified according to two attributes and the frequencies $O_i$ in the different categories we shown in a two way table called Contigency table.

** Define 2x2 Contigency table.

Let us consider two attributes A and B. A is divided into two classes and B is divided into two classes. The various cell frequencies can be expressed in the following table known as 2x2 Contigency table.

| | |
|---|---|
| a | b |
| c | d |

| | | |
|---|---|---|
| a | b | a+b |
| c | d | c+d |
| a+c | b+d | N = a+b+c+d |

$$\text{Expected frequency is } E_i = \frac{(\text{Row total})(\text{column total})}{\text{Grand total}}$$

for each cell.

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

Test procedure:

1. Null hypothesis $H_0$

2. Alternative hypothesis $H_1$

3. level of significance $\alpha$

4. Test of statistic

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

5. Conclusion

degree of freedom $(v) = $ (No. of rows $-1$)(No. of columns $-1$)

calculate $\chi^2_{tab}$ value at $\alpha$ significance at $v$ degree of freedom.

If $|\chi^2| < \chi^2_{tab}$ then null hypothesis is accepted.

If $|\chi^2| > \chi^2_{tab}$ then ~~alt~~ null hypothesis is rejected.

The following table gives the classification of 100 workers according to nature of work and gender. Test whether the nature of work is independent of the gender of the work.

| | stable | unstable | total |
|---|---|---|---|
| Males | 40 | 20 | 60 |
| Females | 10 | 30 | 40 |
| Total | 50 | 50 | 100 |

1. Null hypothesis

$H_0$: The nature of work is independent of the gender of the work.

2. Alternative hypothesis

$H_1$: The nature of work is not independent of the gender of the work.

3. Level of significance

$\alpha = 0.05$

4. Test of statistic.

$E(40) = \frac{60(50)}{100} = 30$

$E(20) = \frac{60(50)}{100} = 30.$

$E(10) = \frac{40(50)}{100} = 20$

$E(30) = \frac{40(50)}{100} = 20.$

| | stable | unstable | Total |
|---|---|---|---|
| Males | 30 | 30 | 60 |
| Females | 20 | 20 | 40 |
| Total | 50 | 50 | 100. |

| $O_i$ | $E_i$ | $O_i - E_i$ | $\frac{(O_i-E_i)^2}{E_i}$ |
|---|---|---|---|
| 40 | 30 | 10 | 3.33 |
| 20 | 30 | -10 | 3.33 |
| 10 | 20 | -10 | 5 |
| 30 | 20 | 10 | 5 |
| | | | 16.66 |

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

$$\chi^2 = 16.66$$

5. Conclusion

$v = (\text{No. of rows}-1)(\text{No. of Columns}-1)$

$= (2-1)(2-1)$

$v = 1$

$\alpha = 0.05$

$\chi^2_{tab} = 3.841.$

$|\chi^2| > \chi^2_{tab}$

∴ The null hypothesis is rejected.

i.e, the nature of work is not independent of the gender of the work.

2. On the basis of information given below about the treatment of 200 patients suffering from a disease, state whether the new treatment is comparatively superior to the conventional treatment.

| | favourable | Not favourable | Total |
|---|---|---|---|
| New | 60 | 30 | 90 |
| conventional | 40 | 70 | 110 |

1. Null hypothesis
   $H_0$: The new treatment is comparatively superior to the conventional treatment (or) there is no difference b/w new and conventional.)
2. Alternative hypothesis
   $H_1$: The new treatment is not comparatively superior to the conventional treatment.
3. level of significance
   $\alpha = 0.05$
4. Test of statistic

$$\text{Expected frequency}(E_i) = \frac{(\text{Row total})(\text{column total})}{\text{Grand total}}$$

| | favourable | Not favourable | Total |
|---|---|---|---|
| New | 45 | 45 | 90 |
| conventional | 55 | 55. | 110 |
| Total | 100 | 100 | 200 |

| $O_i$ | $E_i$ | $O_i - E_i$ | $\frac{(O_i - E_i)^2}{E_i}$ |
|---|---|---|---|
| 60 | 45 | 15 | 5 |
| 30 | 45 | -15 | 5 |
| 40 | 55 | -15 | 4.09 |
| 70 | 55 | +15 | 4.09 |
| | | | 18.018 |

$$\chi^2 = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right] = 18.18$$

5. Conclusion

$v = (\text{No. of rows} - 1)(\text{No. of columns} - 1)$

$v = (2-1)(2-1)$

$v = 1$

$\alpha = 0.05$

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

$|\chi^2| > \chi^2_{tab}$

i.e, $18.18 > 3.841$

$\therefore$ Null hypothesis is rejected.

3. From the following data, find whether there is any significant liking in the habit of taking soft drinks among the categories of employees. Use $\chi^2$ distribution test with level of significance 0.05.

| Softdrinks | employees Clerks | Teachers | officers. |
|---|---|---|---|
| pepsi | 10 | 25 | 65 |
| Thums up | 15 | 30 | 65 |
| fanta | 50 | 60 | 30 |

1. Null hypothesis
   $H_0$: There is significant liking in the habit of taking soft drinks.
2. Alternative hypothesis
   $H_1$: There is no significant liking in the habit of taking soft drinks.
3. Level of significance
   $\alpha = 0.05$.
4. Test of statistic

$$E_i = \frac{(\text{Row total})(\text{Column total})}{\text{Grand total}}.$$

| softdrinks | employees clerks | teachers | officers | total |
|---|---|---|---|---|
| pepsi | 10 | 25 | 65 | 100 |
| Thums up | 15 | 30 | 65 | 110 |
| fanta | 50 | 60 | 30 | 140 |
| Total | 75 | 115 | 160 | 350 |

| softdrinks | employees clerks | teachers | officers | total |
|---|---|---|---|---|
| pepsi | $\frac{100(75)}{350} = 21.29$ | 32.86 | 45.71 | 100 |
| Thumsup | 23.57 | 36.14 | 50.28 | 110 |
| fanta | 30 | 46 | 64 | 140 |
| Total | 75 | 115 | 160 | 350. |

| $O_i$ | $E_i$ | $O_i - E_i$ | $\frac{(O_i - E_i)^2}{E_i}$ |
|---|---|---|---|
| 10 | 21.29 | -11.42 ~~-11.29~~ | 6.088 |
| 25 | 32.86 | -7.86 | 1.88 |
| 65 | 45.71 | 19.29 | 8.14 |
| 15 | 23.57 | -8.57 | 3.12 |
| 30 | 36.14 | -6.14 | 1.04 |
| 65 | 50.28 | 14.72 | 4.31 |
| 50 | 30 | 20 | 13.33 |
| 60 | 46 | 14 | 4.26 |
| 30 | 64 | -34 | 18.06 |
| | | | 60.228 |

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

$$\chi^2 = 60.228.$$

V. Conclusion

$v = (3-1)(3-1) = 4.$

$\alpha = 0.05$

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

$|\chi^2| > \chi^2_{tab}$

∴ Null hypothesis is rejected.

i.e, There is no significant liking in the habit of taking soft drinks.