Unit - II

Test of Significance of Small Samples:-

The test of significance in the previous section hold good only for large samples. i.e., only when the size $n \geq 30$. when the sample is small, i.e., $n < 30$, the sampling distribution of many statistics are not normal, even though the parent population may be normal.

Moreover the assumption of near equality of population parameters and the corresponding sample statistics will not be justified for small samples. Consequently we have to develop entirely different tests of significance that are applicable to small samples.

Student t- Distribution:-

A random variable T is said to be follow Student's t- distribution or simply t - distribution, if it's probability density function is given by

$$\boxed{f(t) = \frac{1}{\sqrt{\nu}\,\beta\left(\frac{\nu}{2}, \frac{1}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}}, \quad -\infty < t < \infty.$$

where ν denotes the no of degrees of freedom of the t- distribution.

The t- distribution was defined by the mathematician Gosset, whose pet name is Student.

Scanned by CamScanner

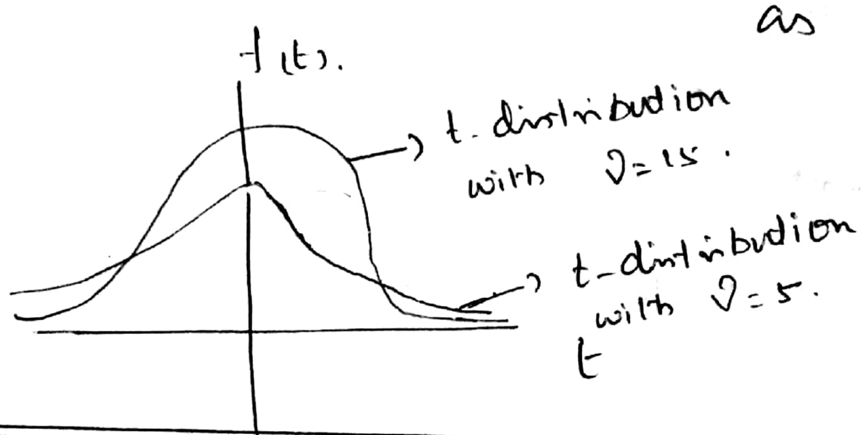
Properties of t-distribution:-

1. The probability curve of the t-distribution is similar to the standard normal curve, and symmetric about $t=0$, bell shaped and asymptotic to the t-axis.

2. For the sufficiently large value of ν, the t-distribution tends to the standard normal distribution.

3. The mean of the t-distribution is zero.

4. The variance of the t-distribution is $\frac{\nu}{\nu-2}$, if $\nu > 2$ and is greater than 1, but it tends to 1 as $\nu \to \infty$.



Uses of t-distribution:-

The t-distribution is used to test the significance of the difference between

1. The mean of the small sample and the mean of the population.

2. The means of two small samples and

3. The co-effi of correlation in the small sample in the population, assumed zero.

Degree's of freedom :.

1. Suppose we wish to find the mean of the sample with observations $x_1, x_2 \dots x_n$. we have to use all the 'n' values taken by the variable with full freedom, for computing $\bar{x}$.

Hence, $\bar{x}$ is said to have 'n' degree's of freedom.

2. Suppose we wish to further compute the S.D's of this sample, using the formula $s^2 = \frac{1}{n}\Sigma(x_i-\bar{x})^2$, since, there is no restriction regarding the value of $\bar{x}$, s is said to have $(n-1)$ degrees of freedom.

If we compute another statistic of the sample based on $\bar{x}$ and s, then that statistic will be assumed to have $(n-2)$ degrees of freedom and so on.

Thus, the no of independent variates used to compute the statistic is known as the no of degrees of freedom of that statistic.

In general the no of degrees of freedom is given by $\nu = n-k$. where n is the no of observations in the sample and k is the no of constraints imposed on them or k is the no of values that have been found out and specified by prior calculations.

test - I. Test of significance of the difference between Sample mean and population Mean:-

If $\bar{x}$ is the mean of sample of size 'n', drawn from a population $N(\mu, \sigma)$.

$$\therefore z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \longrightarrow (1).$$

If the S.D of the Population σ is not known,

then $$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \longrightarrow (2). \quad \{\because \text{ when } n \text{ is large}\}.$$

But n is small, we cannot used s an an estimate of σ.

$$\therefore z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$ can be written as

$$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} \longrightarrow (3).$$

We get $t_{\nu}(\alpha)$ for the LOS α and $\nu = n-1$ from the table value.

If the calculated value of 't' $|t| < t_{\nu}(\alpha)$.

Then the null - hypothesis H_0 is accepted, H_1 is rejected.

Note:- $\boxed{\text{Here D.o.f} = \nu = n-1}$

95 %. Confidence interval of μ is given by

$$\left|\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}}\right| \le t_{0.05} = 0.95.$$

i.e., $$\boxed{\bar{x} - t_{0.05}\frac{s}{\sqrt{n-1}} \le \mu \le \bar{x} + t_{0.05} \times \frac{s}{\sqrt{n-1}}}.$$

Here D.o.f $= \nu = n-1$ for two-tailed test.

A machinist is expected to make engine parts with axle diameter of 1.75 cm. A random sample of 10 parts shows a mean diameter 1.85 cm with an S.D of 0.1 cm. On the basis of this sample, would you say that the work of the machinist is inferior?

Soln:- $\bar{x} = 1.85$, $s = 0.1$, $n = 10$ and $\mu = 1.75$.

H_0 :- $\bar{x} = \mu$,

H_1 :- $\bar{x} \neq \mu$.

Two-tailed test is to be used.

Let los be 5 %. $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}} = \frac{1.85-1.75}{\frac{0.1}{\sqrt{10-1}}} = 3.$

$\therefore \nu = n-1 = 9.$

$\therefore$ from the t-table, for $\nu = 9$, $t_{0.05} = 2.26$ and $t_{0.01} = 3.25$

$\therefore |t| > t_{0.05}$ and $|t| < t_{0.01}$

$\therefore$ Ho is rejected and H_1 is accepted at LOS = 5%, but Ho is accepted and H_1 is rejected at LOS at 1%.

That is at 5% LOS the work of the machinist can be assumed to be inferior, but at 1% LOS the work cannot be assumed to be inferior.

---

2. The Mean lifetime of a sample of 25 bulbs is found as 1550 hrs with a S.D of 120 h. The Company manufacturing the bulbs is found as 1550 hrs with a S.D of 120 h. The Company manufacturing the bulbs claims that the average life of their bulbs is 1600 h. Is the claim acceptable at 5% LOS?

Solution:- Here $\bar{x} = 1550$, $S = 120$, $n = 25$ and $\mu = 1600$.

$H_0 : \bar{x} = \mu$ and $H_1 : \bar{x} < \mu$.

$\therefore$ Left-tailed test is to be used. LOS = 5%.

Now, $t = \dfrac{\bar{x} - \mu}{\frac{S}{\sqrt{n-1}}} = \dfrac{-50\sqrt{24}}{120} = -2.04$

D.of $\nu = n-1 = 25 - 1 = 24$.

$\therefore t_{0.05}$ for one-tailed test $= t_{0.1}$ for $\{\nu = 24 = 1.71\}$

$\therefore |t| = 2.04$, $|t_{0.05}| = 1.71$.

$\Rightarrow |t| > |t|_{0.1}$ $\therefore$ Ho is ~~accepted~~ rejected and H_1 is ~~rejected~~ accep at 5% LOS. $\therefore$ The claim of the company can not be accepted at 5% LOS.

The heights of 10 males of a given locality are found to be 175, 168, 155, 170, 152, 170, 175, 160, 160 and 165 cms. Based on this sample, find the 95% confidence limits for the heights of males in that locality?

Solution: To find the $\bar{x}$ (mean) and S.D 's'.

x_i	$d_i = x_i - A$ ($x_i - 165$), A = 165	d_i^2
175	10	100
168	3	9
155	-10	100
170	5	25
152	13	169
170	5	25
175	10	100
160	-5	25
160	-5	25
165	0	0.
$\Sigma x_i = 1650.$	$\Sigma d_i = 0.$	$\Sigma d_i^2 = 578.$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{1650}{10} = 165 \quad \boxed{\bar{x} = 165}.$$

$$s = \sqrt{\frac{\Sigma d_i^2}{n} - \left(\frac{\Sigma d_i}{n}\right)^2} = \sqrt{\frac{578}{10} - 0}$$

$$= \sqrt{57.8} \quad \therefore \quad s^2 = 57.8 \quad \boxed{s = 7.6}$$

from the table,

$$t_{0.05}(\nu = n-1) = t_{0.05}(\nu = 9) = 2.26.$$ (for the t-test.)

$\therefore$ The 95 % Confidence limits for μ are,

$$= \left\{ \bar{x} - 2.26 \frac{s}{\sqrt{n-1}},\ \bar{x} + 2.26 \cdot \frac{s}{\sqrt{n-1}} \right\}$$

$$= \left[165 - \frac{2.26 \times 7.6}{\sqrt{9}},\ 165 + \frac{2.26 \times 7.6}{\sqrt{9}} \right]$$

$\Rightarrow (159.3, 170.7)$

Conclusion:- The heights of the males in the locality are likely to lie within 159.3 cm and 170.7 cm.

---

type - II Test of significance of the difference between means of two small samples drawn from the same normal population.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \longrightarrow (1) \quad \text{where } \sigma = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \longrightarrow (2).$$

Here n_1, n_2 are very small,

s_1, s_2 are the sample S.D's.

Sub (2) in (1).

$$\boxed{t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}}$$ where d.o.f $= n_1 + n_2 - 2$. $\longrightarrow$ (3)

If $n_1 = n_2 = n$ and if the samples are independent i.e., the observations in the two samples are not at all related, then the test statistic is given by

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{n-1}}} \quad \text{with} \quad \nu = 2n-2. \longrightarrow (4).$$

(2) If $n_1 = n_2 = n$ and if the pairs of values of x_1 and x_2 are associated in some way (or correlated) the formula (4) not to be used.

Here $t = \dfrac{\bar{d}}{\frac{s}{\sqrt{n-1}}}$ with $\nu = n-1$

where, $d_i = x_i - y_i$ and

$\boxed{\bar{d} = \bar{x} - \bar{y}}$ and S.D of $\boxed{d = \frac{1}{n}\Sigma (d_i - \bar{d})^2}$

---

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

(7i)

	Size	Mean	S.D
Sample 1	8	1234 h	36 h
Sample 2	7	1036 h	40 h.

Is the difference in the means sufficient to warrant that type 1 bulbs are superior to type 2 bulbs?

Scanned by CamScanner

Solution :- $\bar{x}_1 = 1234$, $s_1 = 36$, $n_1 = 8$,

$\bar{x}_2 = 1036$, $s_2 = 40$, $n_2 = 7$.

Right-tailed test is to be used.

Let LOS be 5%.

$$\therefore t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}} = \frac{198}{\sqrt{\frac{21568}{13}\left(\frac{1}{8}+\frac{1}{7}\right)}} = \frac{198}{21.0887}$$

$\boxed{t = 9.39}$ {$\because$ Calculation value}.

$\nu = n_1+n_2-2 = 13$. [d.o.f].

$\therefore t_{0.05}(\nu = 13)$ for One-tailed test $= t_{10\%} = 1.77$.

(from the table)

Now, $t > t_{0.1}(\nu = 13)$.

$\therefore$ Ho is rejected and H_1 is accepted.

ie., type 1 bulbs may be regarded superior to type 2. bulbs at 5% LOS.

---

2. The Mean height and the S.D height of 8 randomly chosen soliders are 166.9 and 8.29 cm resply. The values of 6 randomly chosen sailor's ~~soliders~~ are 170.3 and 8.50 cm resply. Based on this data, can we conclude that Solidors are ingeneral shorter than Sailors?

Ex.

Soln:- Here $\bar{x}_1 = 166.9$, $S_1 = 8.29$, $n_1 = 8$.

$\bar{x}_2 = 170.3$, $S_2 = 8.50$, $n_2 = 6$.

Ho :- $\bar{x}_1 = \bar{x}_2$

H_1 :- $\bar{x}_1 < \bar{x}_2$.

Left-tailed test is to be used.

Let LOS be 5%.

$$\therefore t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{-3.4}{\sqrt{\left(\frac{983.29}{12}\right)\left(\frac{1}{8}+\frac{1}{6}\right)}}$$

$|t| = |-0.695| = 0.695$ (calculated value).

$\therefore \nu = n_1 + n_2 - 2 = 12.$

$\therefore t_{0.05} (\nu = 12) = 1.78$ (Table value).

$\therefore |t| < |t_{0.1}|$

$\Rightarrow H_0$ is accepted and H_1 is rejected.

i.e., Based on the given data, we can not conclude that soldiers are, in general shorter than sailors.

The following data relate to the marks obtained by 11 students in 2 tests, one held at the beginning of a year and the other at the end of the year after intensive coaching.

test 1 :-	19	23	16	24	17	18	20	18	21	19	20
test 2 :-	17	24	20	24	20	22	20	20	18	22	19

Do the data indicate that the students have benifited by coaching?

Soln :- The following data relate to the marks obtained by same set of students in 2 tests.

Scanned by CamScanner

Hence, the marks in the 2 tests can be regarded as correlated and so the t-test for paired values should be used.

Let $d = x_1 - x_2$, where x_1, x_2 denote the marks in the 2 tests.

$\therefore$ The values of d are $-2, -1, -4, 0, -3, -4, 0, -2, 3, -3, 1$.

$$\therefore \Sigma d = -11, \quad \Sigma d^2 = 69.$$

$$\bar{d} = \frac{1}{n} \Sigma d = \frac{1}{11}(-11) = -1.$$

$$s^2 = s_d^2 = \frac{1}{n} \Sigma d^2 - (\bar{d})^2 = \frac{1}{11} \times 69 - (-1)^2$$

$$s^2 = 5.27.$$

$$\boxed{s = 2.296}$$

Now, $H_0 :- \bar{d} = 0$, {the students have not benefitted by coaching.}

$H_1 :- \bar{d} < 0$ i.e., $\{\bar{x}_1 < \bar{x}_2\}$.

$\therefore$ One-tailed test is to be used. $\therefore$ LOS = 5%.

$$t = \frac{\bar{d}}{\frac{s}{\sqrt{n-1}}} = \frac{-1}{\frac{2.296}{\sqrt{10}}} = -1.38 \text{ and } \therefore \boxed{t = 1.38}$$

$$\nu = n-1 = 11-1 = 10.$$

$t_{0.05}$ for one-tailed test $= t_{0.1}(\nu = 10) = 1.81$ (table value).

$\therefore |t| < |t_{0.1}|$ $\therefore$ H_0 is accepted and H_1 is rejected

i.e., There is no significance diff between the two tests ... i.e., the students ...

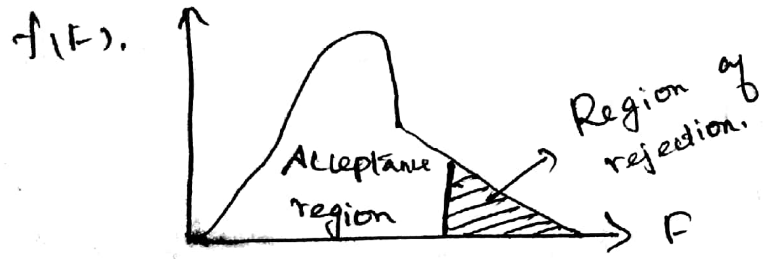
A random variable F is said to follow snedecor's F-distribution (or) simply F-distribution if it's Prob. density fun is given by

$$\boxed{f(F) = \frac{(\nu_1/\nu_2)^{\nu_1/2}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \cdot \frac{F^{\nu_1/2-1}}{\left(1+\frac{\nu_1 F}{\nu_2}\right)^{\nu_1+\nu_2/2}}, F>0.}$$

Here ν_1 and ν_2 are used in f(F) are the degree's of freedom associated with the F-distribution.

Properties of F-distribution:-

1. The curve is of F-distribution is given by



2. The square of the t-variate with n-degrees of freedom follows a F-distribution with 1 and n d.o.f.

3. The Mean of the F distribution is $\frac{\nu_2}{\nu_2 - 2}$. $\nu_2 > 2$.

4. The variance of the F distribution is

$$\frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}, \quad \nu_2 > 4.$$

Scanned by CamScanner

Uses of F-distribution:-

F distribution is used to test the equality of the variance of the populations from which two small samples have been drawn.

F test of significance of the diff between population variances and F-table:-

To test the significance of the difference between population variances, we shall find their estimates σ_1^2 and σ_2^2, based on the sample variances s_1^2 and s_2^2 and then test their equality.

where $\sigma_1^2 = \dfrac{n_1 s_1^2}{n_1 - 1}$, d.o.f $= \nu_1 = n_1 - 1$.

$\sigma_2^2 = \dfrac{n_2 s_2^2}{n_2 - 1}$, d.o.f $= \nu_2 = n_2 - 1$.

$\therefore F = \dfrac{\sigma_1^2}{\sigma_2^2}$, ν_1, ν_2 are the d.o.f

if $\sigma_1^2 = \sigma_2^2$, $F = 1$.

I) Two sample of sizes 9 and 8 gave the sums of squares of deviations from their respective means equal to 160 and 91 resply. Can they be regarded as drawn from the same normal populations?

<u>Solution:-</u>

$n_1 = 9, \quad \Sigma(x_i - \bar{x})^2 = 160 \Rightarrow n_1 s_1^2 = 160.$

$n_2 = 8, \quad \Sigma(y_i - \bar{y})^2 = 91 \Rightarrow n_2 s_2^2 = 91.$

$$\therefore \sigma_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{1}{8} \times 160 = 20.$$

$$\sigma_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{1}{7} \times 91 = 13.$$

$\therefore \sigma_1^2 > \sigma_2^2, \quad \nu_1 = n_1 - 1 = 8$ and $\nu_2 = n_2 - 1 = 7.$

$H_0 :- \sigma_1^2 = \sigma_2^2$ and $H_1 :- \sigma_1^2 \neq \sigma_2^2.$

Let $LOS = 5\%.$

$$\Rightarrow F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{20}{13} = 1.54.$$

$\Rightarrow F_{0.05}(\nu_1 = 8, \nu_2 = 7) = 3.73$ (from the table)

$\therefore F < F_{0.05\%} \Rightarrow H_0$ is accepted.

That is the two samples could have come two normal populations with the same variance we cannot say that, the samples have come same population.

1\. A sample of size 13 gave an estimated population variance of 3.0, while another sample size 15 gave an estimate of 2.5. Could both samples be from populations with the same variance?

Solution:- Here, $n_1 = 13$, $\sigma_1^2 = 3.0$ and $\nu_1 = 12 \{13-1\}$

$n_2 = 15$, $\sigma_2^2 = 2.5$ and $\nu_2 = 14 \{15-1\}$

$H_0 :- \sigma_1^2 = \sigma_2^2$ {: the two samples have been drawn from populations with the same variance.}.

$H_1 :- \sigma_1^2 \neq \sigma_2^2$. {Let LOS = 5%.}:

$$\therefore F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{\frac{n_1 s_1^2}{n_1 - 1}}{\frac{n_2 s_2^2}{n_2 - 1}} = \frac{3.0}{2.5} = 1.2$$

for $\nu_1 = n_1 - 1 = 13 - 1 = 12$, $\nu_2 = 15 - 1 = 14$.

$\therefore F_{0.05\%}$ at $\nu_1 = 12$, $\nu_2 = 14$ = 2.53

(from the table)

$\therefore F < F_{0.05}$.

i.e., H_0 is accepted

$\therefore$ The two-samples could have come from two normal population with the same variance.

...o random samples gave the following data.

Sample no	Size	Mean	variance
1	8	9.6	1.2
8	11	16.5	2.5

Can we conclude that, the samples have been draw from the same normal population?

Solution:- we conclude that, the two samples have been drawn from the same population, we've to check first that the variances of the populations do not differ significantly and then check that the sample means do not differ significantly.

$$\sigma_1^2 = \frac{n_1}{n_1-1} s_1^2 = \frac{8 \times 1.2}{7} = 1.37,$$

$$\sigma_2^2 = \frac{n_2}{n_2-1} s_2^2 = \frac{11 \times 2.5}{10} = 2.75.$$

$$F = \frac{\sigma_2^2}{\sigma_1^2} = 2.007 \quad \text{with } \nu_1 = 10, \ \nu_2 = 7.$$

∴ From the table $F_{0.05}(10, 7) = 3.64$. (Table value)

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

H_1 :- $\sigma_1^2 \neq \sigma_2^2$ then H_0 is accepted

∴ $F < F_{0.05}$

ie., the variances of the populations from which samples are drawn may be regarded as equal.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{9.6 - 16.5}{\sqrt{\left(\frac{9.6 + 27.5}{17}\right)\left(\frac{1}{8} + \frac{1}{11}\right)}} = \frac{-6.9}{0.6864}$$

$$= -10.05$$

and $\nu = n_1 + n_2 - 2 = 17$.

$\therefore$ from the table $t_{0.05}(\nu = 17) = 2.11$

If $H_0 :- \bar{x}_1 = \bar{x}_2$

$H_1 :- \bar{x}_1 \neq \bar{x}_2$

Then H_0 is rejected since $|t| > |t_{0.05}|$.

i.e., the means of two samples differ significantly.

$\therefore$ The two samples could not have been drawn from the same normal population.

4. The nicotine contents in two random samples of tobacco are given below.

Sample 1 :-	21	24	25	26	27	
Sample 2 :-	22	27	28	30	31	36.

can you that the two samples came from the same population?

Solution :-

$\bar{x}_1$ = Mean of sample 1 $= \frac{123}{5} = 24.6$

$\bar{x}_2$ = Mean of sample 2 $= \frac{174}{6} = 29.0$

S_1^2 = variance of sample 1 $= \frac{1}{5} \Sigma (x_i - \bar{x}_1)^2$

$= \frac{1}{5} \Sigma (x_i - 24.6)^2$

$= 4.24$

S_2^2 = variance of sample 2 $= \frac{1}{6} \Sigma (x_i - 29.0)^2$

$= 18.$

$\sigma_1^2 = \frac{5}{4} \times 4.24 = 5.30.$, $\nu = n_1 - 1 = 5 - 1 = 4.$

$\sigma_2^2 = \frac{6}{5} \times 18.0 = 21.60$, $\nu = n_2 - 1 = 6 - 1 = 5.$

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

$H_1 :- \sigma_1^2 \neq \sigma_2^2.$

$F = \frac{\sigma_2^2}{\sigma_1^2} = \frac{21.60}{5.30} = 4.07.$

$\therefore F_{0.05}(5,4) = 6.26.$ Since, $F < F_{0.05}$, H_0 is accepted

$\therefore$ the variances of the two populations can be regarded as equal.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{24.6 - 29.0}{\sqrt{\frac{21.2 + 108.0}{5 + 6 - 2}\left(\frac{1}{5} + \frac{1}{6}\right)}}$$

$= \frac{-4.4}{2.2943} = -1.92$, $\nu = 9.$

$\therefore$ from table $F_{0.05}(\nu = 9) = 2.26$

$H_0 :- \bar{x}_1 = \bar{x}_2.$

$H_1 :- \bar{x}_1 \neq \bar{x}_2$ $\therefore$ H_0 is accepted, since $|t| < F_{0.05}$.

...ution :-

$\bar{x}_1$ = Mean of sample $1 = \frac{123}{5} = 24.6$

$\bar{x}_2$ = Mean of sample $2 = \frac{174}{6} = 29.0$

S_1^2 = variance of sample $1 = \frac{1}{5} \Sigma (x_i - \bar{x}_1)^2$

$= \frac{1}{5} \Sigma (x_i - 24.6)^2$

$= 4.24$

S_2^2 = variance of sample $2 = \frac{1}{6} \Sigma (x_i - 29.0)^2$

$= 18.$

$\sigma_1^2 = \frac{5}{4} \times 4.24 = 5.30$, $\nu = n_1 - 1 = 5 - 1 = 4$.

$\sigma_2^2 = \frac{6}{5} \times 18.0 = 21.60$, $\nu = n_2 - 1 = 6 - 1 = 5$.

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

$H_1 :- \sigma_1^2 \neq \sigma_2^2$.

$F = \frac{\sigma_2^2}{\sigma_1^2} = \frac{21.60}{5.30} = 4.07$.

$\therefore F_{0.05}(5,4) = 6.26$. Since, $F < F_{0.05}$, H_0 is accepted,

$\therefore$ the variances of the two populations can be regarded as equal.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{24.6 - 29.0}{\sqrt{\frac{21.2 + 108.0}{5 + 6 - 2}\left(\frac{1}{5} + \frac{1}{6}\right)}}$$

$$= \frac{-4.4}{2.2943} = -1.92, \quad \nu = 9.$$

$\therefore$ from table $F_{0.05}(\nu = 9) = 2.26$

$\therefore \bar{x}_1 = \bar{x}_2$.

$\bar{x}_1 \neq \bar{x}_2$ $\therefore$ H_0 is accepted, since $|t| < F_{0.05}$.

Solution:-

$\bar{x}_1$ = Mean of sample 1 $= \frac{123}{5} = 24.6$

$\bar{x}_2$ = Mean of sample 2 $= \frac{174}{6} = 29.0$

s_1^2 = variance of sample 1 $= \frac{1}{5}\sum(x_i - \bar{x}_1)^2$

$= \frac{1}{5}\sum(x_i - 24.6)^2$

$= 4.24$

s_2^2 = variance of sample 2 $= \frac{1}{6}\sum(x_i - 29.0)^2$

$= 18.$

$\sigma_1^2 = \frac{5}{4} \times 4.24 = 5.30$, $\nu = n_1 - 1 = 5 - 1 = 4.$

$\sigma_2^2 = \frac{6}{5} \times 18.0 = 21.60$, $\nu = n_2 - 1 = 6 - 1 = 5.$

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

$H_1 :- \sigma_1^2 \neq \sigma_2^2$,

$$F = \frac{\sigma_2^2}{\sigma_1^2} = \frac{21.60}{5.30} = 4.07.$$

$\therefore F_{0.05}(5,4) = 6.26$. Since, $F < F_{0.05}$, H_0 is accepted

$\therefore$ the variances of the two populations can be regarded as equal.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{24.6 - 29.0}{\sqrt{\frac{21.2 + 108.0}{5+6-2}\left(\frac{1}{5} + \frac{1}{6}\right)}}$$

$$= \frac{-4.4}{2.2943} = -1.92, \quad \nu = 9.$$

$\therefore$ from table $F_{0.05}(\nu = 9) = 2.26$

$H_0 :- \bar{x}_1 = \bar{x}_2$.

$H_1 :- \bar{x}_1 \neq \bar{x}_2$ $\therefore$ H_0 is accepted, since $|t| < F_{0.05}$.

Scanned by CamScanner

Solution :-

$\bar{x}_1$ = Mean of sample 1 $= \frac{123}{5} = 24.6$

$\bar{x}_2$ = Mean of sample 2 $= \frac{174}{6} = 29.0$

s_1^2 = variance of sample 1 $= \frac{1}{5}\sum(x_i - \bar{x}_1)^2$

$= \frac{1}{5}\sum(x_i - 24.6)^2$

$= 4.24$

s_2^2 = variance of sample 2 $= \frac{1}{6}\sum(x_i - 29.0)^2$

$= 18.$

$\sigma_1^2 = \frac{5}{4} \times 4.24 = 5.30.$, $\nu = n_1 - 1 = 5 - 1 = 4.$

$\sigma_2^2 = \frac{6}{5} \times 18.0 = 21.60$, $\nu = n_2 - 1 = 6 - 1 = 5.$

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

$H_1 :- \sigma_1^2 \neq \sigma_2^2.$

$F = \frac{\sigma_2^2}{\sigma_1^2} = \frac{21.60}{5.30} = 4.07.$

$\therefore F_{0.05}(5,4) = 6.26.$ Since, $F < F_{0.05}$, H_0 is accepted, the variances of the two populations can be regarded as equal.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{24.6 - 29.0}{\sqrt{\frac{21.2 + 108.0}{5 + 6 - 2}\left(\frac{1}{5} + \frac{1}{6}\right)}}$$

$= \frac{-4.4}{2.2943} = -1.92$, $\nu = 9.$

$\therefore$ from table $F_{0.05}(\nu = 9) = 2.26$

$H_0 :- \bar{x}_1 = \bar{x}_2.$

$H_1 :- \bar{x}_1 \neq \bar{x}_2$ $\therefore$ H_0 is accepted, since $|t| < F_{0.05}$.

i.e., The Means of two samples do not diff. significantly.

∴ The two samples could have been drawn from the same normal population.

Test III Chi-square Distribution:-

If $x_1, x_2 \ldots x_n$ are normally distributed independent random variables, then it is known that $x_1^2 + x_2^2 + \ldots x_n^2$ follows a probability distribution, called Chi-square (χ^2-distribution) with n degrees of freedom. The prob. density fun of the χ^2 distribution is given by

$$\boxed{f(\chi^2) = \frac{1}{2^{\nu/2}\sqrt{\nu/2}} \cdot (\chi^2)^{\nu/2 - 1} \cdot e^{-\chi^2/2}}, \quad 0 < \chi^2 < \infty.$$

where ν is the no of degree's of freedom.

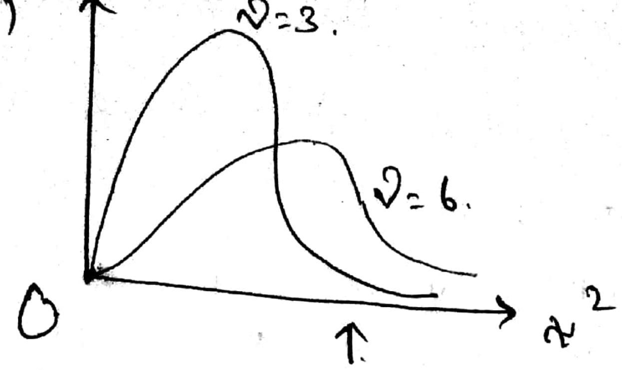
Properties of χ^2 distribution:-

1. A rough sketch of the probability curve of the χ^2 distribution for $\nu = 3$ and $\nu = 6$.

2. As ν becomes smaller and smaller, the curve is skewed more and more to the right. As ν increases, the curve becomes more and more symmetrical.

3. The Mean and variance of the χ^2-distribution are ν and 2ν resply.

4. As ν tends to ∞, the χ^2 distribution becomes a normal



χ^2 curve

Uses of χ^2 - Distribution:-

1. χ^2-distribution is Used to test the goodness of fit. It is used to judge whether a given sample may be reasonably regarded as a simple sample from a certain hypothetical population.

2. It is Used to test the independence of attributes. That is, if a population is known to have two attributes, then χ^2-distribution is used to test whether the two attributes are associated (or) independent, based on a sample drawn from the population.

χ^2 - Test of Goodness of fit:-

On the basis of the hypothesis, assumed about the population, we find the expected frequencies E_i ($i=1,2....n$), corresponding to the Observed frequencies O_i ($i=1,2...n$) such that $\Sigma E_i = \Sigma O_i$

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$, here d.o.f $= \nu = n-1$.

If $\chi^2 > \chi^2_{\alpha}\%$, then we will reject Ho and conclude that the diff is significant.

If $\chi^2 < \chi^2_{\alpha}\%$. then we will accept null hypothesis Ho which means that the given sample is one drawn from the population.

we will conclude that the difference below the observed and expected frequencies is not significant at $\alpha\%$ LOS.

Conditions for the validity of χ^2 - Test:-

1. The no of observations N in the sample must be reasonably large, say ≥ 50.

2. Individual frequencies must not be too small. i.e., $O_i \geq 10$ In case $O_i < 10$. It is combined with the neighboring frequencies, so that the combined frequency is ≥ 10.

3. The number of classes n must be neither too small nor too large. i.e., $4 \leq n \leq 16$.

χ^2 - test of Independence of Attributes:-

$$\left.\begin{array}{l}\text{Expected} \\ \text{frequencies}\end{array}\right\} = E_{ij} = \frac{(\text{Total of observed frequencies in the } i^{th} \text{ row}) \times (\text{total of a observed frequencies in the } j^{th} \text{ column})}{\text{Total of all cell frequencies}}$$

Then we compute,

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \left\{ \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right\}$$

D.o.f $\nu = (m-1) \times (n-1)$.

Note 1: If $\chi^2 < \chi^2_\alpha$, Ho is accepted at $\alpha\%$ LOS.

$\therefore$ A and B are independent.

Note 2: If $\chi^2 > \chi^2_\alpha$, Ho is rejected at $\alpha\%$ LOS.

$\therefore$ A and B are not independent.

The following table shows the distribution of digits in the numbers chosen at random from a telephone directory:-

Digit :-	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 whether the digits may be taken to occur equally frequently in the directory.

<u>Solution:-</u> H_0 :- The digits occur equally frequently, i.e., they follow a uniform distribution.

The total no of digits = 10,000.

If the digits will occur uniformly, then each digit will occur $\frac{10,000}{10} = 1000$ times.

O_i :- 1026 1107 997 966 ...... 853.

E_i :- 10,000 10,000 10,000 10,000 ...... 10,000.

$(O_i - E_i)^2$:- $(26)^2, (107)^2, (-3)^2, (-34)^2, (75)^2, (-67)^2, (107)^2, (-28)^2, (-36)^2, (-147)^2$

$$\chi^2 \; \frac{\Sigma (O_i - E_i)^2}{E_i} :- \frac{1}{1000}\left[(26)^2 + (107)^2 + (-3)^2 + (-34)^2 + (75)^2 + (-67)^2 + (107)^2 + (-28)^2 + (-36)^2 + (-147)^2\right]$$

$\chi^2 = 58.542$. {calculation value}.

D.o.f $\nu = n-1 = 10-1 = 9 = 16.919$ { Table value }.

$\chi^2 > \chi^2_{0.05}$ i.e., H_0 is ~~accepted~~ rejected.

... occur uniformly in the directly

2. The following data give the no of air-craft acc[illegible] that occured during the various days of a week.

Day :-	Mon	Tues	Wed	Thurs	Fri	Sat
No of accidents :-	15	19	13	12	16	15

Test whether the accidents are uniformly distributed over the week.

<u>Solution</u>:- H_0:- accidents occur uniformly over the week.

Total no of accidents = 90.

Based on H_0, the expected no of accidents on any day $= \frac{90}{6} = 15$.

O_i :- 15 19 13 12 16 15

E_i :- 15 15 15 15 15 15.

$$\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = \frac{1}{15}(0 + 16 + 4 + 9 + 1 + 0) = 2.$$

$\because \Sigma E_i = \Sigma O_i$ $\nu = 6 - 1 = 5$.

$\therefore \chi^2_{0.05}$ at 5 % = 11.07.

Since, $\chi^2 < \chi^2_{0.05}$, H_0 is accepted.

That is, accidents may be regarded to occur uniformly over the week.

3. The following data show defective articles produced by 4 machine.

Machine :-	A	B	C	D
Production time :-	1	1	2	3
No of defectives :-	12	30	63	98.

Do the figu[illegible] indicate a significant difference in the [illegible] machines?

Scanned by CamScanner

Solution:- H_0:- Production rates of the machines are the same.

$\therefore$ Total no of defectives = 203.

Based on H_0, the expected no of defectives produced by the machines are

E_i :- $\frac{1}{7} \times 203$, $\frac{1}{7} \times 203$, $\frac{2}{7} \times 203$, $\frac{3}{7} \times 203$.

ie, E_i :- 29 29 58 87.

O_i :- 12 30 63 98.

$$\chi^2 = \frac{\Sigma (O_i - E_i)^2}{E_i} = \frac{17^2}{29} + \frac{1^2}{29} + \frac{5^2}{58} + \frac{11^2}{87}$$

$= 11.82$. (calculed value).

Since, $\Sigma (E_i) = E(O_i) = 203$.

$\therefore \nu = n-1 = 4-1 = 3$.

ie, from the table $\chi^2_{0.05} = \nu = 3 = 7.815$.

(Table value).

$\therefore \chi^2 > \chi^2_{0.05}$, H_0 is rejected.

ie, the significant difference in the performance of machines.

Theory predicts that the proportion of beans in 4 groups A, B, C, D should be 9 : 3 : 3 : 1. In an exam among 1600 beans, the no in the 4 groups were 882, 313, 287 and 118. Does the experiment support the theory?

Solution:- H_0 :- The experiment supports the the no of beans in the 4 groups are in the ratio 9 : 3 : 3 : 1

∴ The expected no of beans in the 4 groups are as follows.

E_i :-	$\frac{9}{16} \times 1600$	$\frac{3}{16} \times 1600$	$\frac{3}{16} \times 1600$	$\frac{1}{16} \times 1600$
E_i :-	900	300	300	100
O_i :-	882	313	287	118.

$$\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = \frac{18^2}{900} + \frac{13^2}{300} + \frac{13^2}{300} + \frac{18^2}{100}$$

$$\chi^2 = 4.73. \text{ (Calculated value)}$$

$\Sigma(E_i) = \Sigma(O_i)$ ∴ $\nu = n-1 = 3.$

$$\chi^2_{0.05}(\nu = 3) = 7.82 \text{ (Table value)}.$$

Since $\chi^2 < \chi^2_{0.05}$, H_0 is accepted.

i.e., the experimental data support the theory.

Note:-

example:-

a	b
c	d.

E_i =

a	b	a + b
c	d	c + d
a + c	b + d	a + b + c + d N.

Expected value of $a = (a+b).(a+c) / a+b+c+d$.

" " of $b = (a+b).(b+d) / a+b+c+d$.

" " of $c = (a+c)(c+d) / a+b+c+d$.

" " of $d = (c+d)(b+d) / a+b+c+d$.

The following data is collected on two characters. Based on this, can you say there is no relation between smoking and literacy?

	Smokers	Non-Smokers.
Literates	83	57
Illiterates	45	68.

Solution:- H_0:- Literacy and smoking habit are independent.

	Smokers	Non-smokers.	Total
Literates	83	57	140
Illiterates	45	68	113
Total	128	125	253.

O	E	E (rounded)	$(O-E)^2/E$.
83	$\frac{128 \times 140}{253} = 70.83$	71	$12^2/71 = 2.03$
57	$\frac{140 \times 125}{253} = 69.17$	69	$12^2/69 = 2.09$
45	$\frac{128 \times 113}{253} = 57.17$	57	$12^2/57 = 2.53$
68	$\frac{125 \times 113}{253} = 55.83$	56	$12^2/56 = 2.57$.
			$\chi^2 = 9.22$.

d.o.f $= (m-1)(n-1) = (2-1)(2$

from the χ^2 table, $\chi^2_{0.05}$ $\nu=1 = 3.84$ (Table val

$\chi^2 = 9.22$ (calculated value)

Since, $\chi^2 > \chi^2_{0.05}$, Ho is rejected.

$\therefore$ There is some association between literacy and smoking.

---

6. A total no of 3759 individuals were interviewed in a public opinion survey on a political proposal. Of them, 1872 were men and the rest women. A total of 2257 individuals were in favour of the proposal and 917 were opposed to it. A total of 243 men were undecided and 442 women were opposed to the proposal. Do you justify or contradict the hypothesis that there is no association between sex and attitude?

Soln:- A careful analysis of the prob results in the following contingency.

	favoured	Opposed	Undecided	Total.
Men	1154	475	243	1872
Women	1103	442	342	1887
Total	2257	917	585	3759.

Ho :- Sex and attitude are independent.

i.e., there is no association between sex and

O	E (rounded)	$(O-E)^2/E$
1154	$\frac{1872 \times 2257}{3759} = 1124$	$30^2/1124 = 0.80$
475	$\frac{1872 \times 917}{3759} = 457$	$18^2/457 = 0.71$
243	$\frac{1872 \times 585}{3759} = 291$	$48^2/291 = 7.92$
1103	$\frac{1887 \times 2257}{3759} = 1133$	$30^2/1133 = 0.79$
442	$\frac{1887 \times 917}{3759} = 460$	$18^2/460 = 0.70$
342	$\frac{1887 \times 585}{3759} = 294.$	$48^2/294 = 7.84$
	d.o.f $= \nu = (3-1)(2-1) = 2$	$\chi^2 = 18.76$

$\therefore$ from the χ^2 table, $\chi^2_{0.05}(\nu = 2) = 5.99$

Since, $\chi^2 > \chi^2_{0.05}$, Ho is rejected.

$\therefore$ Sex and attitude are not independent, i.e., there is some association between sex and attitude.

Fit a binomial distribution for the following data and also test the goodness of fit.

x :-	0	1	2	3	4	5	6	Total
f :-	5	18	28	12	7	6	4	80.

To find the binomial distribution $N(q+p)^n$, which fits given data, we require p.

Scanned by CamScanner

Solution:- We know that the mean of the binomial distribution is np, from which we find p.

To find the Mean:-

x	0	1	2	3	4	5	6	Total.
f	5	18	28	12	7	6	4	80
$f(x)$	0	18	56	36	28	30	24	192.

$$\bar{x} = \frac{\Sigma f(x)}{\Sigma(f)} = \frac{192}{80} = 2.4.$$

$\therefore np = 2.4$

$\therefore 6p = 2.4 \Rightarrow p = 0.4$ and $q = 0.6$.

i.e., the expected frequencies are given by the successive term in the expansion of $80(0.6+0.4)^6$

$\therefore E_i$:-	3.73	14.93	24.88	22.12	11.06	2.95	0.33
Whole of E_i :-	4	15	25	22	11	3	0.
O_i :-	5	18	28	12	7	6	4.

The first class is combined with the second and the last 2 classes are combined with the last but second class in order to make the expected frequency in each class greater than or equal to 10.

After regrouping, we've

E_i :- 19 25 22 14

O_i :- 23 28 12 17.

$$\chi^2 = \sum \frac{(O_i - E_i)}{E_i} = \frac{4^2}{19} + \frac{3^2}{25} + \frac{10^2}{22} + \frac{3^2}{14}$$

$= 6.39$ (calculated value)

we've used the given sample to find $\sum(O_i) = \sum(E_i)$ and p through it's Mean.

$\therefore \nu = n - k = 4 - 2 = 2$ and $\chi^2_{0.05}(\nu = 2) = 5.99$ (Table value)}.

Since $\chi^2 > \chi^2_{0.05}$, Ho is approximately a binomial distribution is rejected.

i.e., the binomial fit for the given distribution is not satisfactory.

Fit a poisson Distribution for the following distribution and also test the goodness of fit.

x :-	0	1	2	3	4	5	Total
f :-	142	156	69	27	5	1	400.

Solution:- Poisson Distribution is

$$P(X = r) = \frac{e^{-\lambda} \cdot \lambda^r}{r!}, \quad r = 0, 1, 2 \ldots, \quad \lambda \to \text{Mean}$$

To find Mean (λ):-

x :-	0	1	2	3	4	5	Total.
f :-	142	156	69	27	5	1	400.
$f(x)$:-	0	156	138	81	20	5	400.

$$\bar{x} = \frac{\Sigma f(x)}{\Sigma f} = \frac{400}{400} = 1 \quad \therefore \boxed{\lambda = 1}.$$

$\therefore$ Expected frequency is given by,

$$N \cdot \frac{e^{-\lambda} \cdot \lambda^r}{r!} = \frac{400 \times e^{-1}}{r!}, \quad r = 0, 1, 2 \ldots \infty.$$

$\therefore$ E_i :- 147.15 147.15 73.58 24.53 6.13 1.23

ie, The values of E_i are very small, for $i = 6, 7, \ldots$ and hence neglected

E_i :- (whole no)	147	147	74	25	6	1.
O_i :-	142	156	69	27	5	1

The last 3 classes are combined into one.

$\therefore$ The expected frequency in that class may be ≥ 10. The regrouping is.

O_i :-	142	156	69	33
E_i :-	147	147	74	32.

$$\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = \frac{5^2}{147} + \frac{9^2}{147} + \frac{5^2}{74} + \frac{1^2}{32}.$$

21.09.

d.o.f = n - k = 4 - 2 = 2.

$\therefore \chi^2_{0.05}$ of Table for ($\nu = 2$) = 5.99 (Table value)

Calculate value $\chi^2 = 1.09$.

Since, $\chi^2 < \chi^2_{0.05}$, which assumes that the given distribution is nearly poisson is accepted.

ie., The poisson fit for the given distribution is satisfactory.

Example [illegible] 6 and 9.

Fitting binomial and poisson

Using chi-square Test

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

$v = 4-1 = 3$