HYPOTHESIS TESTING FOR COMPARING TWO RELATED SAMPLES

Paired t-test is a way to test for comparing two related samples, involving small values of n that does not require the variances of the two populations to be equal, but the assumption that the two populations are normal must continue to apply. For a paired t-test, it is necessary that the observations in the two samples be collected in the form of what is called matched pairs i.e., "each observation in the one sample must be paired with an observation in the other sample in such a manner that these observations are somehow "matched" or related, in an attempt to eliminate extraneous factors which are not of interest in test."[5] Such a test is generally considered appropriate in a before-and-after-treatment study. For instance, we may test a group of certain students before and after training in order to know whether the training is effective, in which situation we may use paired t-test. To apply this test, we first work out the difference score for each matched pair, and then find out the average of such differences, $\overline{D}$, along with the sample variance of the difference score. If the values from the two matched samples are denoted as X_i and Y_i and the differences by D_i ($D_i = X_i - Y_i$), then the mean of the differences i.e.,

$$\overline{D} = \frac{\sum D_i}{n}$$

and the variance of the differences or

$$\left(\sigma_{diff.}\right)^2 = \frac{\sum D_i^2 - \left(\overline{D}\right)^2 \cdot n}{n-1}$$

Assuming the said differences to be normally distributed and independent, we can apply the paired t-test for judging the significance of mean of differences and work out the test statistic t as under:

$$t = \frac{\overline{D} - 0}{\sigma_{diff}/\sqrt{n}} \text{ with } (n-1) \text{ degrees of freedom}$$

where $\overline{D}$ = Mean of differences

$\sigma_{diff.}$ = Standard deviation of differences

n = Number of matched pairs

Example1 :

Memory capacity of 9 students was tested before and after training. State at 5 per cent level of significance whether the training was effective from the following scores:

Student	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Before	10	15	9	3	7	12	16	17	4
After	12	17	8	5	6	11	18	20	3

Use paired t-test as well as A-test for your answer.

$H_0: \mu_1 = \mu_2$ which is equivalent to test $H_0: \overline{D} = 0$

$H_a: \mu_1 < \mu_2$ (as we want to conclude that training has been effective)

As we are having matched pairs, we use paired t-test and work out the test statistic t as under:

$$t = \frac{\overline{D} - 0}{\sigma_{diff.}/\sqrt{n}}$$

To find the value of t, we shall first have to work out the mean and standard deviation of differences as shown below:

Student	*Score before training* X_i	*Score after training* Y_i	*Difference* $(D_i = X_i - Y_i)$	*Difference Squared* D_i^2
1	10	12	−2	4
2	15	17	−2	4
3	9	8	1	1
4	3	5	−2	4
5	7	6	1	1
6	12	11	1	1
7	16	18	−2	4
8	17	20	−3	9
9	4	3	1	1
$n = 9$			$\sum D_i = -7$	$\sum D_i^2 = 29$

$\therefore$ Mean of Differences or $\overline{D} = \frac{\sum D_i}{n} = \frac{-7}{9} = -0.778$

and Standard deviation of differences or

$$\sigma_{diff.} = \sqrt{\frac{\sum D_i^2 - (\overline{D})^2 \cdot n}{n-1}}$$

$$= \sqrt{\frac{29 - (-.778)^2 \times 9}{9-1}}$$

$$= \sqrt{2.944} = 1.715$$

Hence, $$t = \frac{-0.778 - 0}{1.715/\sqrt{9}} = \frac{-.778}{0.572} = -1.361$$

Degrees of freedom $= n - 1 = 9 - 1 = 8$.

As H_a is one-sided, we shall apply a one-tailed test (in the left tail because H_a is of less than type) for determining the rejection region at 5 per cent level which comes to as under, using the table of t-distribution for 8 degrees of freedom:

$$R : t < -1.860$$

The observed value of t is -1.361 which is in the acceptance region and thus, we accept H_0 and conclude that the difference in score before and after training is insignificant i.e., it is only due to sampling fluctuations. Hence we can infer that the training was not effective.

Example 2:

The sales data of an item in six shops before and after a special promotional campaign are:

Shops	*A*	*B*	*C*	*D*	*E*	*F*
Before the promotional campaign	53	28	31	48	50	42
After the campaign	58	29	30	55	56	45

Can the campaign be judged to be a success? Test at 5 per cent level of significance. Use paired t-test as well as A-test.

$H_0: \mu_1 = \mu_2$ which is equivalent to test $H_0: \overline{D} = 0$

$H_a: \mu_1 < \mu_2$ (as we want to conclude that campaign has been a success).

Because of the matched pairs we use paired t-test and work out the test statistic 't' as under:

$$t = \frac{\overline{D} - 0}{\sigma_{diff.}/\sqrt{n}}$$

To find the value of t, we first work out the mean and standard deviation of differences as under:

Shops	*Sales before campaign* X_i	*Sales after campaign* Y_i	*Difference* $(D_i = X_i - Y_i)$	*Difference squared* D_i^2
A	53	58	–5	25
B	28	29	–1	1
C	31	30	1	1
D	48	55	–7	49
E	50	56	–6	36
F	42	45	–3	9
$n = 6$			$\sum D_i = -21$	$\sum D_i^2 = 121$

$\therefore$
$$\overline{D} = \frac{\sum D_i}{n} = -\frac{21}{6} = -3.5$$

$$\sigma_{diff.} = \sqrt{\frac{\sum D_i^2 - (\overline{D})^2 \cdot n}{n-1}} = \sqrt{\frac{121 - (-3.5)^2 \times 6}{6-1}} = 3.08$$

Hence,
$$t = \frac{-3.5 - 0}{3.08/\sqrt{6}} = \frac{-3.5}{1.257} = -2.784$$

Degrees of freedom = $(n - 1) = 6 - 1 = 5$

As H_a is one-sided, we shall apply a one-tailed test (in the left tail because H_a is of less than type) for determining the rejection region at 5 per cent level of significance which come to as under, using table of t-distribution for 5 degrees of freedom:

$$R : t < -2.015$$

The observed value of t is -2.784 which falls in the rejection region and thus, we reject H_0 at 5 per cent level and conclude that sales promotional campaign has been a success.

HYPOTHESIS TESTING OF PROPORTIONS

In case of qualitative phenomena, we have data on the basis of presence or absence of an attribute(s). With such data the sampling distribution may take the form of binomial probability distribution whose mean would be equal to $n \cdot p$ and standard deviation equal to $\sqrt{n \cdot p \cdot q}$, where p represents the probability of success, q represents the probability of failure such that $p + q = 1$ and n, the size of the sample. Instead of taking mean number of successes and standard deviation of the number of successes, we may record the proportion of successes in each sample in which case the mean and standard deviation (or the standard error) of the sampling distribution may be obtained as follows:

Mean proportion of successes $= (n \cdot p)/n = p$

and standard deviation of the proportion of successes $= \sqrt{\dfrac{p \cdot q}{n}}$.

In n is large, the binomial distribution tends to become normal distribution, and as such for proportion testing purposes we make use of the test statistic z as under:

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{p \cdot q}{n}}}$$

where $\hat{p}$ is the sample proportion.

For testing of proportion, we formulate H_0 and H_a and construct rejection region, presuming normal approximation of the binomial distribution, for a predetermined level of significance and then may judge the significance of the observed sample result. The following examples make all this quite clear.

Example 1:

A sample survey indicates that out of 3232 births, 1705 were boys and the rest were girls. Do these figures confirm the hypothesis that the sex ratio is 50 : 50? Test at 5 per cent level of significance.

$$H_0: p = p_{H_0} = \frac{1}{2}$$

$$H_a: p \neq p_{H_0}$$

Hence the probability of boy birth or $p = \dfrac{1}{2}$ and the probability of girl birth is also $\dfrac{1}{2}$.

Considering boy birth as success and the girl birth as failure, we can write as under:

the proportion success or $p = \dfrac{1}{2}$

the proportion of failure or $q = \dfrac{1}{2}$

and $n = 3232$ (given).

The standard error of proportion of success.

$$= \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{\frac{1}{2} \times \frac{1}{2}}{3232}} = 0.0088$$

Observed sample proportion of success, or

$$\hat{p} = 1705/3232 = 0.5275$$

and the test statistic

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{0.5275 - .5000}{.0088} = 3.125$$

As H_a is two-sided in the given question, we shall be applying the two-tailed test for determining the rejection regions at 5 per cent level which come to as under, using normal curve area table:

$$R : |z| > 1.96$$

The observed value of z is 3.125 which comes in the rejection region since $R : |z| > 1.96$ and thus, H_0 is rejected in favour of H_a. Accordingly, we conclude that the given figures do not conform the hypothesis of sex ratio being 50 : 50.

Example 2:

The null hypothesis is that 20 per cent of the passengers go in first class, but management recognizes the possibility that this percentage could be more or less. A random sample of 400 passengers includes 70 passengers holding first class tickets. Can the null hypothesis be rejected at 10 per cent level of significance?

$H_0 : p = 20\%$ or 0.20

and $H_a : p \neq 20\%$

H

Hence, $p = 0.20$ and
$q = 0.80$

Observed sample proportion $(\hat{p}) = 70/400 = 0.175$

and the test statistic $z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{0.175 - .20}{\sqrt{\frac{.20 \times .80}{400}}} = -1.25$

As H_a is two-sided we shall determine the rejection regions applying two-tailed test at 10 per cent level which come to as under, using normal curve area table:

$$R : |z| > 1.645$$

The observed value of z is -1.25 which is in the acceptance region and as such H_0 is accepted. Thus the null hypothesis cannot be rejected at 10 per cent level of significance.

HYPOTHESIS TESTING FOR DIFFERENCE BETWEEN PROPORTIONS

If two samples are drawn from different populations, one may be interested in knowing whether the difference between the proportion of successes is significant or not. In such a case, we start with the hypothesis that the difference between the proportion of success in sample one $(\hat{p}_1)$ and the proportion of success in sample two $(\hat{p}_2)$ is due to fluctuations of random sampling. In other words, we take the null hypothesis as $H_0: \hat{p}_1 = \hat{p}_2$ and for testing the significance of difference, we work out the test statistic as under:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}}}$$

where $\hat{p}_1$ = proportion of success in sample one

$\hat{p}_2$ = proportion of success in sample two

$\hat{q}_1 = 1 - \hat{p}_1$

$\hat{q}_2 = 1 - \hat{p}_2$

n_1 = size of sample one

n_2 = size of sample two

and

$\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$ = the standard error of difference between two sample proportions.*

Example1:

A drug research experimental unit is testing two drugs newly developed to reduce blood pressure levels. The drugs are administered to two different sets of animals. In group one, 350 of 600 animals tested respond to drug one and in group two, 260 of 500 animals tested respond to drug two. The research unit wants to test whether there is a difference between the efficacy of the said two drugs at 5 per cent level of significance. How will you deal with this problem?

$H_0: \hat{p}_1 = \hat{p}_2$

The alternative hypothesis can be taken as that there is a difference between the drugs i.e., $H_a: \hat{p}_1 \neq \hat{p}_2$ and the given information can be stated as:

$$\hat{p}_1 = 350/600 = 0.583$$

$$\hat{q}_1 = 1 - \hat{p}_1 = 0.417$$

$$n_1 = 600$$

$$\hat{p}_2 = 260/500 = 0.520$$

$$\hat{q}_2 = 1 - \hat{p}_2 = 0.480$$

$$n_2 = 500$$

We can work out the test statistic z thus:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{0.583 - 0.520}{\sqrt{\frac{(.583)(.417)}{600} + \frac{(.520)(.480)}{500}}}$$

$$= 2.093$$

As H_a is two-sided, we shall determine the rejection regions applying two-tailed test at 5% level which comes as under using normal curve area table:

$$R : |z| > 1.96$$

The observed value of z is 2.093 which is in the rejection region and thus, H_0 is rejected in favour of H_a and as such we conclude that the difference between the efficacy of the two drugs is significant.

Example 2:

At a certain date in a large city 400 out of a random sample of 500 men were found to be smokers. After the tax on tobacco had been heavily increased, another random sample of 600 men in the same city included 400 smokers. Was the observed decrease in the proportion of smokers significant? Test at 5 per cent level of significance.

Solution:

on tobacco remains unchanged i.e. $H_0: \hat{p}_1 = \hat{p}_2$ and the alternative hypothesis that proportion of smokers after tax has decreased i.e.,

$$H_a: \hat{p}_1 > \hat{p}_2$$

On the presumption that the given populations are similar as regards the given attribute, we work out the best estimate of proportion of smokers (p_0) in the population as under, using the given information:

$$p_0 = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{500\left(\frac{400}{500}\right) + 600\left(\frac{400}{600}\right)}{500 + 600} = \frac{800}{1100} = \frac{8}{11} = .7273$$

Thus, $q_0 = 1 - p_0 = .2727$

The test statistic z can be worked out as under:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_0q_0}{n_1} + \frac{p_0q_0}{n_2}}} = \frac{\frac{400}{500} - \frac{400}{600}}{\sqrt{\frac{(.7273)(.2727)}{500} + \frac{(.7273)(.2727)}{600}}}$$

$$= \frac{0.133}{0.027} = 4.926$$

As the H_a is one-sided we shall determine the rejection region applying one-tailed test (in the right tail because H_a is of greater than type) at 5 per cent level and the same works out to as under, using normal curve area table:

$$R : z > 1.645$$

The observed value of z is 4.926 which is in the rejection region and so we reject H_0 in favour of H_a and conclude that the proportion of smokers after tax has decreased significantly.

HYPOTHESIS TESTING FOR COMPARING A VARIANCE TO SOME HYPOTHESISED POPULATION VARIANCE

The test we use for comparing a sample variance to some theoretical or hypothesised variance of population is different than z-test or the t-test. The test we use for this purpose is known as chi-square test and the test statistic symbolised as χ^2, known as the chi-square value, is worked out. The chi-square value to test the null hypothesis viz, $H_0: \sigma_s^2 = \sigma_p^2$ worked out as under:

$$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2}(n-1)$$

where σ_s^2 = variance of the sample

σ_p^2 = variance of the population

$(n-1)$ = degree of freedom, n being the number of items in the sample.

Then by comparing the calculated value of χ^2 with its table value for $(n-1)$ degrees of freedom at a given level of significance, we may either accept H_0 or reject it. If the calculated value of χ^2 is equal to or less than the table value, the null hypothesis is accepted; otherwise the null hypothesis is rejected. This test is based on chi-square distribution which is not symmetrical and all the values happen to be positive; one must simply know the degrees of freedom for using such a distribution.*

TESTING THE EQUALITY OF VARIANCES OF TWO NORMAL POPULATIONS

When we want to test the equality of variances of two normal populations, we make use of F-test based on F-distribution. In such a situation, the null hypothesis happens to be $H_0: \sigma_{p_1}^2 = \sigma_{p_2}^2$, $\sigma_{p_1}^2$ and $\sigma_{p_2}^2$ representing the variances of two normal populations. This hypothesis is tested on the basis of sample data and the test statistic F is found, using $\sigma_{s_1}^2$ and $\sigma_{s_2}^2$ the sample estimates for $\sigma_{p_1}^2$ and $\sigma_{p_2}^2$ respectively, as stated below:

$$F = \frac{\sigma_{s_1}^2}{\sigma_{s_2}^2}$$

where $\sigma_{s_1}^2 = \dfrac{\Sigma\left(X_{1i} - \bar{X}_1\right)^2}{(n_1 - 1)}$ and $\sigma_{s_2}^2 = \dfrac{\Sigma\left(X_{2i} - \bar{X}_2\right)^2}{(n_2 - 1)}$

While calculating F, $\sigma_{s_1}^2$ is treated $> \sigma_{s_2}^2$ which means that the numerator is always the greater variance. Tables for F-distribution** have been prepared by statisticians for different values of F at different levels of significance for different degrees of freedom for the greater and the smaller variances. By comparing the observed value of F with the corresponding table value, we can infer whether the difference between the variances of samples could have arisen due to sampling fluctuations. If the calculated value of F is greater than table value of F at a certain level of significance for $(n_1 - 1)$ and $(n_2 - 2)$ degrees of freedom, we regard the F-ratio as significant. Degrees of freedom for greater variance is represented as v_1 and for smaller variance as v_2. On the other hand, if the calculated value of F is smaller than its table value, we conclude that F-ratio is not significant. If F-ratio is considered non-significant, we accept the null hypothesis, but if F-ratio is considered significant, we then reject H_0 (i.e., we accept H_a).

When we use the F-test, we presume that

(i) the populations are normal;
(ii) samples have been drawn randomly;
(iii) observations are independent; and
(iv) there is no measurement error.

Example 1:

Two random samples drawn from two normal populations are:

Sample 1	20	16	26	27	23	22	18	24	25	19		
Sample 2	27	33	42	35	32	34	38	28	41	43	30	37

Test using variance ratio at 5 per cent and 1 per cent level of significance whether the two populations have the same variances.

Solution:

drawn have the same variances i.e., $H_0 : \sigma_{p_1}^2 = \sigma_{p_2}^2$. From the sample data we work out $\sigma_{s_1}^2$ and $\sigma_{s_2}^2$ as under:

Sample 1			*Sample 2*		
X_{1i}	$(X_{1i} - \bar{X}_1)$	$(X_{1i} - \bar{X}_1)^2$	X_{2i}	$(X_{2i} - \bar{X}_2)$	$(X_{2i} - \bar{X}_2)^2$
20	−2	4	27	−8	64
16	−6	36	33	−2	4
26	4	16	42	7	49
27	5	25	35	0	0
23	1	1	32	−3	9
22	0	0	34	−1	1
18	−4	16	38	3	9
24	2	4	28	−7	49
25	3	9	41	6	36
19	−3	9	43	8	64
			30	−5	25
			37	2	4
$\Sigma X_{1i} = 220$		$\Sigma(X_{1i} - \bar{X}_1)^2 = 120$	$\Sigma X_{2i} = 420$		$\Sigma(X_{2i} - \bar{X}_2)^2 = 314$
$n_1 = 10$			$n_2 = 12$		

$$\bar{X}_1 = \frac{\Sigma X_{1i}}{n_1} = \frac{220}{10} = 22; \qquad \bar{X}_2 = \frac{\Sigma X_{2i}}{n_2} = \frac{420}{12} = 35$$

$\therefore$
$$\sigma_{s_1}^2 = \frac{\Sigma\left(X_{1i} - \bar{X}_1\right)^2}{n_1 - 1} = \frac{120}{10 - 1} = 13.33$$

and
$$\sigma_{s_2}^2 = \frac{\Sigma\left(X_{2i} - \bar{X}_2\right)^2}{n_2 - 1} = \frac{314}{12 - 1} = 28.55$$

Hence,
$$F = \frac{\sigma_{s_2}^2}{\sigma_{s_1}^2} \qquad \left(\because \sigma_{s_2}^2 > \sigma_{s_1}^2\right)$$

$$= \frac{28.55}{13.33} = 2.14$$

Degrees of freedom in sample 1 = $(n_1 - 1) = 10 - 1 = 9$
Degrees of freedom in sample 2 = $(n_2 - 1) = 12 - 1 = 11$
As the variance of sample 2 is greater variance, hence

$$v_1 = 11; v_2 = 9$$

The table value of F at 5 per cent level of significance for $v_1 = 11$ and $v_2 = 9$ is 3.11 and the table value of F at 1 per cent level of significance for $v_1 = 11$ and $v_2 = 9$ is 5.20.

Since the calculated value of $F = 2.14$ which is less than 3.11 and also less than 5.20, the F ratio is insignificant at 5 per cent as well as at 1 per cent level of significance and as such we accept the null hypothesis and conclude that samples have been drawn from two populations having the same variances.

Example 2:

Given $n_1 = 9; n_2 = 8$

$$\Sigma\left(X_{1i} - \bar{X}_1\right)^2 = 184$$

$$\Sigma\left(X_{2i} - \bar{X}_2\right)^2 = 38$$

Apply F-test to judge whether this difference is significant at 5 per cent level.

To test this, we work out the F-ratio as under:

$$F = \frac{\sigma_{s_1}^2}{\sigma_{s_2}^2} = \frac{\Sigma\left(X_{1i} - \bar{X}_1\right)^2/(n_1 - 1)}{\Sigma\left(X_{2i} - \bar{X}_2\right)^2/(n_2 - 1)}$$

$$= \frac{184/8}{38/7} = \frac{23}{5.43} = 4.25$$

$v_1 = 8$ being the number of d.f. for greater variance

$v_2 = 7$ being the number of d.f. for smaller variance.

The table value of F at 5 per cent level for $v_1 = 8$ and $v_2 = 7$ is 3.73. Since the calculated value of F is greater than the table value, the F ratio is significant at 5 per cent level. Accordingly we reject H_0 and conclude that the difference is significant.

QUESTIONS:

1. A coin is tossed 10,000 times and head turns up 5,195 times. Is the coin unbiased?
2. In some dice throwing experiments, A threw dice 41952 times and of these 25145 yielded a 4 or 5 or 6. Is this consistent with the hypothesis that the dice were unbiased?
3. A machine puts out 16 imperfect articles in a sample of 500. After machine is overhauled, it puts out three imperfect articles in a batch of 100. Has the machine improved? Test at 5% level of significance.
4. In two large populations, there are 35% and 30% respectively fair haired people. Is this difference likely to be revealed by simple sample of 1500 and 1000 respectively from the two populations?
5. In a certain association table the following frequencies were obtained: (AB) $= 309, (Ab) = 214, (aB) = 132, (ab) = 119.$
   Can the association between AB as per the above data can be said to have arisen as a fluctuation of simple sampling?
6. A sample of 900 members is found to have a mean of 3.47 cm. Can it be reasonably regarded as a simple sample from a large population with mean 3.23 cm. and standard deviation 2.31 cm.?
7. The means of the two random samples of 1000 and 2000 are 67.5 and 68.0 inches respectively. Can the samples be regarded to have been drawn from the same population of standard deviation 9.5 inches? Test at 5% level of significance.
8. A large corporation uses thousands of light bulbs every year. The brand that has been used in the past has an average life of 1000 hours with a standard deviation of 100 hours. A new brand is offered to the corporation at a price far lower than one they are paying for the old brand. It is decided that they will switch to the new brand unless it is proved with a level of significance of 5% that the new brand has smaller average life than the old brand. A random sample of 100 new brand bulbs is tested yielding an observed sample mean of 985 hours. Assuming that the standard deviation of the new brand is the same as

that of the old brand,

(a) What conclusion should be drawn and what decision should be made?

(b) What is the probability of accepting the new brand if it has the mean life of 950 hours?

9. Ten students are selected at random from a school and their heights are found to be, in inches, 50, 52, 52, 53, 55, 56, 57, 58, 58 and 59. In the light of these data, discuss the suggestion that the mean height of the students of the school is 54 inches. You may use 5% level of significance (Apply t-test as well as A-test).

10. In a test given to two groups of students, the marks obtained were as follows:

First Group	18	20	36	50	49	36	34	49	41
Second Group	29	28	26	35	30	44	46		

Examine the significance of difference between mean marks obtained by students of the above two groups. Test at five per cent level of significance.

11. The heights of six randomly chosen sailors are, in inches, 63, 65, 58, 69, 71 and 72. The heights of 10 randomly chosen soldiers are, in inches, 61, 62, 65, 66, 69, 69, 70, 71, 72 and 73. Do these figures indicate that soldiers are on an average shorter than sailors? Test at 5% level of significance.

12. Ten young recruits were put through a strenuous physical training programme by the army. Their weights (in kg) were recorded before and after with the following results:

Recruit	1	2	3	4	5	6	7	8	9	10
Weight before	127	195	162	170	143	205	168	175	197	136
Weight after	135	200	160	182	147	200	172	186	194	141

Using 5% level of significance, should we conclude that the programme affects the average weight of young recruits (Answer using t-test as well as A-test

13. Answer using F-test whether the following two samples have come from the same population:

Sample 1 17 27 18 25 27 29 27 23 17

Sample 2 16 16 20 16 20 17 15 21

Use 5% level of significance.

14. The following table gives the number of units produced per day by two workers A and B for a number of days:

A 40 30 38 41 38 35

B 39 38 41 33 32 49 49 34

Should these results be accepted as evidence that B is the more stable worker? Use F-test at 5% level.

15. A sample of 600 persons selected at random from a large city gives the result that males are 53%. Is there reason to doubt the hypothesis that males and females are in equal numbers in the city? Use 1% level of significance.

16. 12 students were given intensive coaching and 5 tests were conducted in a month. The scores of tests 1 and 5 are given below. Does the score from Test 1 to Test 5 show an improvement? Use 5% level of significance.

No. of students	1	2	3	4	5	6	7	8	9	10	11	12
Marks in 1st Test	50	42	51	26	35	42	60	41	70	55	62	38
Marks in 5th test	62	40	61	35	30	52	68	51	84	63	72	50

Chi-Square Test

The chi-square test is an important test amongst the several tests of significance developed by statisticians. Chi-square, symbolically written as χ^2 (Pronounced as Ki-square), is a statistical measure used in the context of sampling analysis for comparing a variance to a theoretical variance. As a non-parametric* test, it "can be used to determine if categorical data shows dependency or the two classifications are independent. It can also be used to make comparisons between theoretical populations and actual data when categories are used."[1] Thus, the chi-square test is applicable in large number of problems. The test is, in fact, a technique through the use of which it is possible for all researchers to (i) test the goodness of fit; (ii) test the significance of association between two attributes, and (iii) test the homogeneity or the significance of population variance.

CHI-SQUARE AS A TEST FOR COMPARING VARIANCE

The chi-square value is often used to judge the significance of population variance i.e., we can use the test to judge if a random sample has been drawn from a normal population with mean (μ) and with a specified variance (σ_p^2). The test is based on χ^2-distribution. Such a distribution we encounter when we deal with collections of values that involve adding up squares. Variances of samples require us to add a collection of squared quantities and, thus, have distributions that are related to χ^2-distribution. If we take each one of a collection of sample variances, divided them by the known population variance and multiply these quotients by $(n-1)$, where n means the number of items in the sample, we shall obtain a χ^2-distribution. Thus, $\frac{\sigma_s^2}{\sigma_p^2}(n-1) = \frac{\sigma_s^2}{\sigma_p^2}$ (d.f.) would have the same distribution as χ^2-distribution with $(n-1)$ degrees of freedom.

Example 1:

Weight of 10 students is as follows:

S. No.	1	2	3	4	5	6	7	8	9	10
Weight (kg.)	38	40	45	53	47	43	55	48	52	49

Can we say that the variance of the distribution of weight of all students from which the above sample of 10 students was drawn is equal to 20 kgs? Test this at 5 per cent and 1 per cent level of significance.

S. No.	X_i (Weight in kgs.)	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
1	38	–9	81
2	40	–7	49
3	45	–2	04
4	53	+6	36
5	47	+0	00
6	43	–4	16
7	55	+8	64
8	48	+1	01
9	52	+5	25
10	49	+2	04
$n = 10$	$\Sigma X_i = 470$	$\Sigma(X_i - \bar{X})^2 = 280$	

$$\bar{X} = \frac{\Sigma X_i}{n} = \frac{470}{10} = 47 \text{ kgs.}$$

$$\therefore \qquad \sigma_s = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{280}{10-1}} = \sqrt{31.11}$$

or

$$\sigma_s^2 = 31.11.$$

Let the null hypothesis be $H_0 : \sigma_p^2 = \sigma_s^2$. In order to test this hypothesis we work out the χ^2 value as under:

$$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2}(n-1)$$

$$= \frac{31.11}{20}(10-1) = 13.999.$$

Degrees of freedom in the given case is $(n - 1) = (10 - 1) = 9$. At 5 per cent level of significance the table value of $\chi^2 = 16.92$ and at 1 per cent level of significance, it is 21.67 for 9 d.f. and both these values are greater than the calculated value of χ^2 which is 13.999. Hence we accept the null hypothesis and conclude that the variance of the given distribution can be taken as 20 kgs at 5 per cent as also at 1 per cent level of significance. In other words, the sample can be said to have been taken from a population with variance 20 kgs.

Example 2:

A sample of 10 is drawn randomly from a certain population. The sum of the squared deviations from the mean of the given sample is 50. Test the hypothesis that the variance of the population is 5 at 5 per cent level of significance.

$$n = 10$$

$$\Sigma\left(X_i - \overline{X}\right)^2 = 50$$

$$\therefore \qquad \sigma_s^2 = \frac{\Sigma\left(X_i - \overline{X}\right)^2}{n-1} = \frac{50}{9}$$

Take the null hypothesis as $H_0: \sigma_p^2 = \sigma_s^2$. In order to test this hypothesis, we work out the χ^2 value as under:

$$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2}(n-1) = \frac{\frac{50}{9}}{5}(10-1) = \frac{50}{9} \times \frac{1}{5} \times \frac{9}{1} = 10$$

Degrees of freedom = (10 – 1) = 9.

The table value of χ^2 at 5 per cent level for 9 d.f. is 16.92. The calculated value of χ^2 is less than this table value, so we accept the null hypothesis and conclude that the variance of the population is 5 as given in the question.

CHI-SQUARE AS A NON-PARAMETRIC TEST

Chi-square is an important non-parametric test and as such no rigid assumptions are necessary in respect of the type of population. We require only the degrees of freedom (implicitly of course the size of the sample) for using this test. As a non-parametric test, chi-square can be used (i) as a test of goodness of fit and (ii) as a test of independence.

As a test of goodness of fit, χ^2 test enables us to see how well does the assumed theoretical distribution (such as Binomial distribution, Poisson distribution or Normal distribution) fit to the observed data. When some theoretical distribution is fitted to the given data, we are always interested in knowing as to how well this distribution fits with the observed data. The chi-square test can give answer to this. If the calculated value of χ^2 is less than the table value at a certain level of significance, the fit is considered to be a good one which means that the divergence between the observed and expected frequencies is attributable to fluctuations of sampling. But if the calculated value of χ^2 is greater than its table value, the fit is not considered to be a good one.

As a test of independence, χ^2 test enables us to explain whether or not two attributes are associated. For instance, we may be interested in knowing whether a new medicine is effective in controlling fever or not, χ^2 test will helps us in deciding this issue. In such a situation, we proceed with the null hypothesis that the two attributes (viz., new medicine and control of fever) are independent which means that new medicine is not effective in controlling fever. On this basis we first calculate the expected frequencies and then work out the value of χ^2. If the calculated value of χ^2 is less than the table value at a certain level of significance for given degrees of freedom, we conclude that null hypothesis stands which means that the two attributes are independent or not associated (i.e., the new medicine is not effective in controlling the fever). But if the calculated value of χ^2 is greater than its table value, our inference then would be that null hypothesis does not hold good which means the two attributes are associated and the association is not because of some chance factor but it exists in reality (i.e., the new medicine is effective in controlling the fever and as such may be prescribed). It may, however, be stated here that χ^2 is not a measure of the degree of relationship or the form of relationship between two attributes, but is simply a technique of judging the significance of such association or relationship between two attributes.

In order that we may apply the chi-square test either as a test of goodness of fit or as a test to judge the significance of association between attributes, it is necessary that the observed as well as theoretical or expected frequencies must be grouped in the same way and the theoretical distribution must be adjusted to give the same total frequency as we find in case of observed distribution. χ^2 is then calculated as follows:

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

where

O_{ij} = observed frequency of the cell in *i*th row and *j*th column.

E_{ij} = expected frequency of the cell in *i*th row and *j*th column.

If two distributions (observed and theoretical) are exactly alike, $\chi^2 = 0$; but generally due to sampling errors, χ^2 is not equal to zero and as such we must know the sampling distribution of χ^2 so that we may find the probability of an observed χ^2 being given by a random sample from the hypothetical universe. Instead of working out the probabilities, we can use ready table which gives probabilities for given values of χ^2. Whether or not a calculated value of χ^2 is significant can be ascertained by looking at the tabulated values of χ^2 for given degrees of freedom at a certain level of significance. If the calculated value of χ^2 is equal to or exceeds the table value, the difference between the observed and expected frequencies is taken as significant, but if the table value is more than the calculated value of χ^2, then the difference is considered as insignificant i.e., considered to have arisen as a result of chance and as such can be ignored.

As already stated, degrees of freedom* play an important part in using the chi-square distribution and the test based on it, one must correctly determine the degrees of freedom. If there are 10 frequency classes and there is one independent constraint, then there are (10 – 1) = 9 degrees of freedom. Thus, if '*n*' is the number of groups and one constraint is placed by making the totals of observed and expected frequencies equal, the d.f. would be equal to (n – 1). In the case of a contingency table (i.e., a table with 2 columns and 2 rows or a table with two columns and more than two rows or a table with two rows but more than two columns or a table with more than two rows and more than two columns), the d.f. is worked out as follows:

$$\text{d.f.} = (c - 1)(r - 1)$$

where 'c' means the number of columns and 'r' means the number of rows.

Example 1:

A die is thrown 132 times with following results:

Number turned up	1	2	3	4	5	6
Frequency	16	20	25	14	29	28

Is the die unbiased?

Solution: Let us take the hypothesis that the die is unbiased. If that is so, the probability of obtaining any one of the six numbers is 1/6 and as such the expected frequency of any one number coming upward is $132 \times 1/6 = 22$. Now we can write the observed frequencies along with expected frequencies and work out the value of χ^2 as follows:

No. turned up	*Observed frequency* O_i	*Expected frequency* E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
1	16	22	–6	36	36/22
2	20	22	–2	4	4/22
3	25	22	3	9	9/22
4	14	22	–8	64	64/22
5	29	22	7	49	49/22
6	28	22	6	36	36/22

$$\therefore \qquad \Sigma [(O_i - E_i)^2/E_i] = 9.$$

Hence, the calculated value of $\chi^2 = 9$.

$\because$ Degrees of freedom in the given problem is

$(n - 1) = (6 - 1) = 5$.

The table value* of χ^2 for 5 degrees of freedom at 5 per cent level of significance is 11.071. Comparing calculated and table values of χ^2, we find that calculated value is less than the table value and as such could have arisen due to fluctuations of sampling. The result, thus, supports the hypothesis and it can be concluded that the die is unbiased.

Example 2:

Find the value of χ^2 for the following information:

Class	*A*	*B*	*C*	*D*	*E*
Observed frequency	8	29	44	15	4
Theoretical (or expected) frequency	7	24	38	24	7

Solution: Since some of the frequencies less than 10, we shall first re-group the given data as follows and then will work out the value of χ^2 :

Class	*Observed frequency* O_i	*Expected frequency* E_i	$O_i - E_i$	$(O_i - E_i)^2/E_i$
A and *B*	(8 + 29) = 37	(7 + 24) = 31	6	36/31
C	44	38	6	36/38
D and *E*	(15 + 4) = 19	(24 + 7) = 31	–12	144/31

$$\therefore \qquad \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 6.76 \text{ app.}$$

Example 3:

Genetic theory states that children having one parent of blood type *A* and the other of blood type *B* will always be of one of three types, *A*, *AB*, *B* and that the proportion of three types will on an average be as 1 : 2 : 1. A report states that out of 300 children having one *A* parent and *B* parent, 30 per cent were found to be types *A*, 45 per cent per cent type *AB* and remainder type *B*. Test the hypothesis by χ^2 test.

Solution: The observed frequencies of type *A*, *AB* and *B* is given in the question are 90, 135 and 75 respectively.

The expected frequencies of type *A*, *AB* and *B* (as per the genetic theory) should have been 75, 150 and 75 respectively.

We now calculate the value of χ^2 as follows:

Type	*Observed frequency* O_i	*Expected frequency* E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
A	90	75	15	225	225/75 = 3
AB	135	150	–15	225	225/150 = 1.5
B	75	75	0	0	0/75 = 0

$$\therefore \qquad \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 3 + 1.5 + 0 = 4.5$$

$$\because \qquad \text{d.f.} = (n - 1) = (3 - 1) = 2.$$

Table value of χ^2 for 2 d.f. at 5 per cent level of significance is 5.991.

The calculated value of χ^2 is 4.5 which is less than the table value and hence can be ascribed to have taken place because of chance. This supports the theoretical hypothesis of the genetic theory that on an average type *A*, *AB* and *B* stand in the proportion of 1 : 2 : 1.

Example 4:

The following information is obtained concerning an investigation of 50 ordinary shops of small size:

	Shops		*Total*
	In towns	*In villages*	
Run by men	17	18	35
Run by women	3	12	15
Total	20	30	50

Can it be inferred that shops run by women are relatively more in villages than in towns? Use χ^2 test.

Solution: Take the hypothesis that there is no difference so far as shops run by men and women in towns and villages. With this hypothesis the expectation of shops run by men in towns would be:

$$\text{Expectation of } (AB) = \frac{(A) \times (B)}{N}$$

where A = shops run by men

B = shops in towns

$(A) = 35$; $(B) = 20$ and $N = 50$

Thus, expectation of $(AB) = \dfrac{35 \times 20}{50} = 14$

Hence, table of expected frequencies would be

	Shops in towns	*Shops in villages*	*Total*
Run by men	14 (AB)	21 (Ab)	35
Run by women	6 (aB)	9 (ab)	15
Total	20	30	50

Calculation of χ^2 value:

Groups	*Observed frequency* O_{ij}	*Expected frequency* E_{ij}	$(O_{ij} - E_{ij})$	$(O_{ij} - E_{ij})^2/E_{ij}$
(AB)	17	14	3	9/14 = 0.64
(Ab)	18	21	–3	9/21 = 0.43
(aB)	3	6	–3	9/6 = 1.50
(ab)	12	9	3	9/9 = 1.00

$$\therefore \qquad \chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = 3.57$$

As one cell frequency is only 3 in the given 2 × 2 table, we also work out χ^2 value applying Yates' correction and this is as under:

$$\chi^2 \text{(corrected)} = \frac{[|17 - 14| - 0.5]^2}{14} + \frac{[|18 - 21| - 0.5]^2}{21} + \frac{[|3 - 6| - 0.5]^2}{6} + \frac{[|12 - 9| - 0.5]^2}{9}$$

$$= \frac{(2.5)^2}{14} + \frac{(2.5)^2}{21} + \frac{(2.5)^2}{6} + \frac{(2.5)^2}{9}$$

$$= 0.446 + 0.298 + 1.040 + 0.694$$

$$= 2.478$$

$\because$ Degrees of freedom = $(c - 1)(r - 1) = (2 - 1)(2 - 1) = 1$

Table value of χ^2 for one degree of freedom at 5 per cent level of significance is 3.841. The calculated value of χ^2 by both methods (i.e., before correction and after Yates' correction) is less than its table value. Hence the hypothesis stands. We can conclude that there is no difference between shops run by men and women in villages and towns.

REFERENCES

1. Boundless statistics by lumen learning website - https://courses.lumenlearning.com/boundless-statistics/chapter/hypothesis-testing-two-samples/
2. Hypothesis Testing: A Visual Introduction To Statistical Significance Kindle Edition by Scott Hartshorn