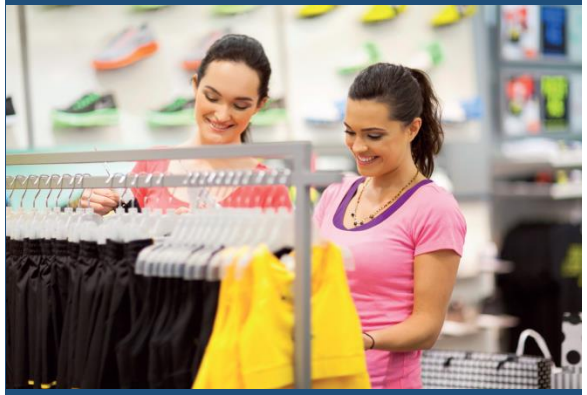


12

Chi-Square Tests

CHI- SQUARE TEST

In sampling testing certain assumptions are made about the population and samples. For example in sampling distribution it is assumed that the samples are drawn from a population following normal distribution. But sometimes it is not possible to make any assumptions about the distribution of the population from where the samples are taken. In such situations we follow a non – parametric test. Chi Square is one such non – parametric test

While collecting data if the observed frequency 'O' and expected frequency 'E' are known then $\chi^2 = \sum_{\text{all cells}} \frac{(f_o - f_e)^2}{f_e}$

Chi-Square as a Statistical Test

- *Chi-square test:* an **inferential statistics** technique designed to test for **significant relationships** between two variables organized in a bivariate table.
- Chi-square requires **no assumptions** about the shape of the population distribution from which a sample is drawn.

Expected frequencies f_e*:* the cell frequencies that would be **expected** in a bivariate table **if** the two tables were **statistically independent**.

Observed frequencies f_o*:* the cell frequencies **actually observed** in a bivariate table.

Meaning of Chi-Square Test:

The Chi-square (χ^2) test represents a useful method of comparing more than 2 population proportion

Experimentally obtained results with those to be expected theoretically on some hypothesis.

Thus Chi-square is a measure of actual divergence of the observed and expected frequencies

Properties of Chi Square Test

1. It is a non – parametric test. Assumptions about the form of the distribution or its parameters are not required
2. It is a distribution free test, which can be used in any type of distribution of population.
3. It is easy to calculate chi – square statistics
4. It analyses the differences between a set of observed frequencies and a set of corresponding expected frequencies.
5. It is a multinomial distribution
6. The variable varies from 0 to ∞
7. It is a one tailed test

Uses of Chi Square Test

1. Useful for the test of goodness of fit
2. Useful for the test of independence of attributes
3. Useful for testing homogeneity
4. Useful for testing given population variance

Conditions for Applying Chi Square Test

1. The total frequencies (N) must be reasonably large than or at least 50
2. Observed frequency of less than 5 is pooled with the preceding or succeeding frequencies so that no observed frequency is less than 5. Then the degree of freedom is based on the resulting number of frequencies.
3. The distribution should not be expressed as proportions or percentages. It should be of original units.

Test of Homogenity	• Was the distribution of political views (liberal, moderate, conservative) different for last three presidential elections in the United States?

What is the chi-square goodness of fit test?

> The Chi-square goodness of fit test is used to compare a randomly collected sample containing a single, categorical variable to a larger population. This test is most commonly used to compare a random sample to the population from which it was potentially collected.

The test begins with the creation of a null and alternative hypothesis. In this case, the hypotheses are as follows:

Null Hypothesis (Ho): The collected data is consistent with the population distribution.

Alternative Hypothesis (Ha): The collected data is not consistent with the population distribution.

What is the chi-square test of independence?

> The Chi-square test for independence looks for an association between two categorical variables within the same population. Unlike the goodness of fit test, the test for independence does not compare a single observed variable to a theoretical population, but rather two variables within a sample set to one another.

The hypotheses for a Chi-square test of independence are as follows:

Null Hypothesis (Ho): There is no association between the two categorical variables in the population of interest.

Alternative Hypothesis (Ha): There is no as an tion between the two categorical variables in the population of interest.

Contingency Tables

Contingency Tables

- Useful in situations involving multiple population proportions
- Used to classify sample observations according to two or more characteristics
- Also called a cross-classification table.

Sample results organized in contingency table

Hand Preference	Gender		Row Total
	Female	Male	
Left	12	24	36
Right	108	156	264
Column Total	120	180	300

sample size = n = 300

120 Females, 12 were left handed

180 Males, 24 were left handed

Hypotheses

- The **research hypothesis** (H_1) proposes that the two variables are **related** in the population.
- The **null hypothesis** (H_o) states that **no association exists** between the two cross-tabulated variables in the population, and therefore the variables are **statistically independent**.

$$\chi^2 = \sum_{\text{all cells}} \frac{(f_o - f_e)^2}{f_e}$$

where:

f_o = observed frequency in a particular cell

f_e = expected frequency in a particular cell if H_0 is true

$$\text{Expected frequency} = \frac{\text{Row total} * \text{column total}}{\text{Grand total}}$$

CHI –SQUARE TEST – 2 (Test of Independence)

•A sample of 200 persons with a particular disease was selected. Out of these,100 were given a drug and the others were not given any drug. The result are as follows:

	No of persons		
	Drug	No Drug	Row Total
Cured	65	55	120
Not Cured	35	45	80
Column Total	100	100	200

Test , whether the drug is effective or not

The "x^2" in the chi-square test for independence determines how likely the difference between actual and expected frequencies can be explained by random chance.

Find expected frequency and chi-square lue

Objective :
Whether the drug is effective in curing the disease

Expected Frequency When o= 65

$E=\frac{120*100}{200}$ = 60

O	E	(O - E)	$(O-E)^2$	(O-E)^2/E
65	60	5	25	0.417
35	40	-5	25	0.625
55	60	-5	25	0.417
45	40	5	25	0.625
200.000			χ^2 =	2.083

$$E = \frac{RT * CT}{GT}$$

O-observed frequency

E-Expected frequency

H0 : O = E

H1 : Not stated above

Hypothesis Testing

Test Statistics

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$ =2.083

To find table value

LOS is not given — Assume 5 %

d.f = (r-1) (c-1) = (2-1) (2-1) =1.

Table value Using Excel

=CHIINV(0.05,1) — 3.841459

Decision : Accept Ho
Since calculated value χ^2
Is less than table value of χ^2

Conclusion : Drug is not effective

An automobile manufacturing firm is bringing out a new model. In order to map out its advertising campaign, it wants to determine whether the model appeal depends on age group or not. The firm takes a random sample from persons attending a preview of the new model and obtained the results summarized below

Persons who	Under20	20-40	40-50	50 & over	Total (RT)
Liked the Car	146	78	48	28	300
Disliked the car	54	52	32	62	200
Total (CT)	200	130	80	90	500 (GT)

H0 : O = E

H1 : Not as stated above

Test , whether the model appeal and age groups are independent

$$E = \frac{RT * CT}{GT}$$

O	E	O-E	(O-E)^2	(O-E)^2/E
146	120	26	676	5.633
54	80	-26	676	8.450
78	78	0	0	0
52	52	0	0	0
48	48	0	0	0
32	32	0	0	0
28	54	-26	676	12.51
62	36	26	676	18.778

χ^2 = 45.38

Hypothesis Testing

Test Statistics

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$ =45.38

Decision : Reject Ho
Since calculated value of χ^2
Is greater than table value of χ^2

To find table value

LOS is not given — Assume 5 %

d.f = (r-1) (c-1) = (2-1) (4-1) =3.

Where
r = no of rows & c = no of columns

Table value Using Excel

=CHIINV(0.05,3) — 7.814

Conclusion : Model appeal depends on age groups

TEST OF GOODNESS OF FIT

It is used to decide whether there is a significant difference between theory and Experiment

An ad agency after each member of a random sample of 60 viwers to indicate which of the 6 television program he/she prefers.The results are as follows

Program	1	2	3	4	5	6	Total
Number	5	8	10	12	12	13	60

Can we concude from these data that the program are not equally preferred

Program	O	E	O-E	(O-E)^2	(O-E)^2/E
1	5	10	-5	25	2.5
2	8	10	-2	4	0.4
3	10	10	0	0	0
4	12	10	2	4	0.4
5	12	10	2	4	0.4
6	13	10	3	9	0.9
E = 60/6 =10				χ^2	4.6

Ho : Six Programs are equally preferred

H1 : Six Programs are not equally preferred

Hypothesis Testing

To find table value

LOS is not given

Assume 5 %

Test Statistics

$\chi^2 = \sum \frac{(O-E)^2}{E}$ = **4.6**

d.f = n-1 = 6-1 =5.

Table value Using Excel

=CHIINV(0.05,5)

11.07

Decision : Accept Ho .Since calculated value of χ^2 Is less than table value of χ^2

Conclusion : All the 6 Programs are equally preferred

Test of Goodness of Fit

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

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

Is the result consistent with the hypothesis that male and female births are equally probable or Test whether the data fits into Binomial distribution

Assume that the male and female births are equally probable ,the probability of male birth p = 0.5.The expected no of families can be calculated by the use of Binomial distribution, The probability of male x male birth in a family of five is given by **P(x) = 5cx** $p^x q^{n-x}$

The expected frequency is N*P(x)

	=BINOMDIST(M10,5,0.5,FALSE)	=320*N1
x	P(X)	E
0	0.03125	10
1	0.15625	50
2	0.3125	100
3	0.3125	100
4	0.15625	50
5	0.03125	10

O	E	O-E	(O-E)^2	(O-E)^2/E
14	10	4	16	1.6
56	50	6	36	0.72
110	100	10	100	1
88	100	-12	144	1.44
40	50	-10	100	2
12	10	2	4	0.4
Here no of observation n= 6 Degrees of freedom : n-1 =5			χ^2 =	7.16

Ho : Male female births are equally probable

H1 : Male female births are not equally probable

Hypothesis Testing

To find table value

LOS is not given

Assume 5 %

d.f = n-1 =6-1=5

Test Statistics

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$ **= 7.16**

Table value Using Excel

=CHIINV(0.05,5) 11.07

Decision : Accept Ho Since calculated value of χ^2 Is less than table value of χ^2

Conclusion : Male and Female births are equally probable

How should the chi-square test be interpreted?

The "x^2" in the goodness of fit test determines how well the sample data matches the characteristics of the larger population. If the sample data does not match the expected properties of the population, this sample is not used for drawing conclusions related to the larger population.

The "x^2" in the chi-square test for independence determines how likely the difference between actual and expected frequencies can be explained by random chance. This difference may or may not have been caused by a sampling error.

What is the p-value in a chi-square test?

The p-value, calculated in a chi-square test, represents an area in the tail of a probability distribution curve. A p-value is a number between zero and one. It is expressed in decimals.

For example, a p-value of 0.0254 implies a 2.54% probability that the results could have happened by chance. The smaller the p-value, the more important (significant) the results are. A significant result is one where the null hypothesis is rejected.

Testing the Population Variance

Weights in kilograms of 10 shipments are given below
38,40,45,53,47,43,55,48,52,49
Can we say that the variance of the distribution of weight of all shipments from which the above sample of 10 shipments are drawn is equal to 20 square Kg

Test Statistics :

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

We know

$$s^2 = \frac{\Sigma(x-\dot{x})\hat{}2}{n-1}$$

(or)

$$s^2(n-1) = \Sigma(x-\dot{x})\hat{}2$$

Test Statistics :

$$\chi^2 = \frac{\Sigma(x-\dot{x})\hat{}2}{\sigma^2}$$

	x	x-xbar	(x-xbar)^2
	38	-9	81
	40	-7	49
	45	-2	4
	53	6	36
	47	0	0
	43	-4	16
	55	8	64
	48	1	1
	52	5	25
	49	2	4
Total	470	0	280
Average	47		

H0 : $\sigma^2 = 20$

H1 : $\sigma^2 \neq 20$

Test Statistics :

$$\chi^2 = \frac{\Sigma(x-\dot{x})\hat{}2}{\sigma^2} = \frac{280}{20} = 14$$

To find table value

LOS is not given | Assume 5 %

d.f = n-1 =10-1=9

Table value Using Excel | =CHIINV(0.05,9) | 16.919

Decision : Accept Ho Since calculated value of χ^2 Is less than table value of χ^2

Conclusion : Variance of the distribution of the weights of all shipments in the population is 20 kgs