Chi-Square Test (X^2-test)

Chi-Square Test (X^2-test)

The Chi-Square test is a statistical procedure used by researchers to examine the differences between categorical variables in the same population. Or A chi-square test is a statistical test used to compare observed results with expected results.

The purpose of this test is to determine if a difference between observed data and expected data is due to chance, or if it is due to a relationship between the variables you are studying. Therefore, a chi-square test is an excellent choice to help us better understand and interpret the relationship between our two categorical variables.

For example, imagine that a research group is interested in whether or not education level and marital status are related for all people in the certain country.

When is the Chi-Square Test Used

Researchers use the Chi-Square test when they find themselves in one of the following situations:

1. They need to estimate how closely an observed distribution matches an expected distribution. This is referred to as a **"goodness-of-fit" test.**
2. They need to estimate whether two random variables are independent.
3. For testing association between factors (groups)
4. testing homogeneity.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

χ^2 = chi squared

O_i = observed value

E_i = expected value

Example 1: a scientist wants to know if education level and marital status are related for all people in some country. He collects data on a simple random sample of n = 300 people, part of which are shown below.

Table of Observed Values

Marital Status by Education | n = 300

	Middle school or lower	High school	Bachelor's	Master's	PhD or higher	Total
Never married	18	36	21	9	6	90
Married	12	36	45	36	21	150
Divorced	6	9	9	3	3	30
Widowed	3	9	9	6	3	30
Total	39	90	84	54	33	300

First step: select hypothesis

Null hypothesis (Ho): there is no relation between marital status and education level.

Alternative hypothesis (Ha): there is significant relation between marital status and education level.

Second step: select significant level **LS= 0.05 or 0.01**

Step three: Find the expected value by the following equation

$$Ei = \frac{\sum rowi * \sum columni}{\text{grand total}}$$

Table of Expected Values

Qualification / Marital Status	Middle School	High School	Bachelor's	Master's	Ph.D
Never Married	$\frac{90 \times 39}{300} = 11.7$	$\frac{90 \times 90}{300} = 27$	25.2	16.2	9.9
Married	19.5	45	42	27	16.5
Divorced	3.9	9	8.4	5.4	3.3
Widowed	3.9	9	8.4	5.4	3.3

Step four calculate the value of chi square

Calculation of χ^2

Observed Values (O)	Expected Values (E)	$(O-E)$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
18	11.7	6.3	39.69	3.39
36	27	9	81	3
21	25.2	-4.2	17.64	0.7
9	16.2	-7.2	51.84	3.2
6	9.9	-3.9	15.21	1.53
12	19.5	-7.5	56.25	2.88
36	45	-9	81	1.8
. . .	. . .	. . .	. . .	. . .
3	3.3	-0.3	0.09	0.02
				$\sum \frac{(O-E)^2}{E}$ $\chi^2 = 23.57$

Step five find df and tabulated chi square value

Degrees of freedom= (columns – 1) (rows-1)
=(5-1) (4-1) =4 × 3 =**12**

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
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

Last step compare calculated chi square with tabulated chi square

$\chi 2$ tabular at LS = 0.05, = 21.03

$\chi 2$ calculated = 23.57

because **$\chi 2$ calculated** is greater than **$\chi 2$ tabular** we reject Null hypothesis, and accept alternative hypothesis.
Alternative hypothesis (Ha): there is significant relation between marital status and education qualification.

Example 2: In an experiment on cotton plants possible to get in the second generation on 625 plants, their flowers are yellow color and 195 plants whose flowers are white, so do the results of experiment follow of isolation and Mendelian ratio, which is 3:1
Null hypothesis (Ho): isolation of flowers color are follow mendel ration 3:1

Alternative hypothesis (Ha): there is significant difference between flower color distribution or

it isn't follow follow mendel ration 3:1

Sum or Totale number of flowers is 625+ 195= 820 flowers (observation)

Expected value = total observation X ratio of observed value, so

Expected value for yellow flowers = ¾ X 820 = 615
Expected value for white flowers = ¼ X 820 = 205

	Observed values (O)	Expected values (E)	(O-E)	$(O-E)^2$	$X^2 = \frac{(O-E)^2}{E}$
Yellow flowers	625	615	10	100	100/615= 0.163
White flowers	195	205	-10	100	100/205= 0.488
Total	**820**	**820**			**0.651**

df = 2-1 = 1
$\chi 2$ tabular at LS = 0.05, = 3.841 and at LS 0.01 = 6.535

because **$\chi 2$ calculated** is less than **$\chi 2$ tabular** we accept Null hypothesis, it means the second generation flowers of cotn plant follow (agree) to mendel ration 3:1.

Example 3: a geneticist took a random sample of 300 men to study whether there is association between father and son regarding boldness. He obtained the following results.

	Son	
Father	**Bold**	**Not**
Bold	**85**	**59**
Not	**65**	**91**

Using α = 5% test whether there is association between father and son regarding boldness.

Ho: There is no association between father and son regarding boldness
Ha: There is association between father and son regarding boldness

- First calculate the row and column totals

	Son		
Father	**Bold**	**Not**	**Total (Row)**
Bold	**85**	**59**	**144**
Not	**65**	**91**	**156**
Total (Column)	**150**	**150**	**Grant total= 300**

- Second find expected values $E = \frac{\sum R * \sum C}{\text{grand total}}$

	Son	
Father	**Bold**	**Not**
Bold	$\frac{144 * 150}{300} = 72$	$\frac{144 * 150}{300} = 72$
Not	$\frac{156 * 150}{300} = 78$	$\frac{156 * 150}{300} = 78$

- Third calculate χ2

Observed values (O)	Expected values (E)	(O-E)	$(O-E)^2$	$X^2 = \frac{(O-E)^2}{E}$
85	72	13	169	169/72=2.347
59	72	-13	169	169/72=2.347
65	78	-13	169	169/78=2.166
91	78	13	169	169/78=2.166
Total				**9.028**

$$= \frac{(85-72)^2}{72} + \frac{(59-72)^2}{72} + \frac{(65-78)^2}{78} + \frac{(91-78)^2}{78} = 9.028$$

- Obtain the tabulated value of chi-square

  $df = (r-1)(c-1) = 1 \times 1 = 1$

  $\chi^2_{0.05}(1) = 3.841$ *from table.*
- The decision is to reject H_0 since $\chi^2_{cal} > \chi^2_{0.05}(1)$

Conclusion

- **At 5% level of significance we have evidence to say there is association between father and son regarding boldness, based on this sample data.**

Example 4: Suppose we have a sample of 120 plants, where 32 of them were diseased: (88 non-diseased and 32 diseased). Test the following hypothesis:

Ho: p = 70 %

Ha: p ≠ 70 %

If the hypothesis was true, then p = 32/120 = 0.26, and we will expect that:

0.30 * 120 = 36 diseased

0.70 * 120 = 84 non-diseased

And like Z test, X^2 test depend on the fact that the null hypothesis is true, then:

	diseased	Non-diseased	Total
No. observed	32	88	120
No. expected	36	84	120

$$X^2 = \frac{(O-E)^2}{E}$$

$$= \frac{(32-36)^2}{36} + \frac{(88-84)^2}{84} = \frac{16}{36} + \frac{16}{84} = 0.44 + 0.19 = 0.63$$

It shown that calculated X^2 (0.63) less than tabulated X^2 (3.84), then null hypothesis will accept.

Example 5: In a population, the predicted proportions of white, brown, and pied are 0.36, 0.48, and 0.16, respectively. There were 140 white rabbits, 240 brown rabbits, and 20 pied rabbits in a group of 400 rabbits. Are the proportions in the rabbit sample different from what you'd expect?

Color	Observed (yi)	Expected (E(yi))
White	140	(0.36)*(400)= 144
Brown	240	(0.48)*(400) = 192
Pied	20	(0.16) *(400) = 64
total	400	

$$X^2 = \frac{(O-E)^2}{E}$$

$$= \frac{(140-144)^2}{144} + \frac{(240-192)^2}{192} + \frac{(20-64)^2}{64} = 42.361$$

It shown that calculated X^2 (42.361) more than tabulated X^2 (5.99), then alternative hypothesis will accept.

Example 4.

- **Attack rates among the vaccinated and unvaccinated against measles are given in the Table below.**
- **Prove the protective value of vaccination by $\chi 2$ test.**

Group		Result		
		Attacked	Not-attacked	Total
Vaccinated	(O)	10	90	100
	(E)	18	82	
Unvaccinated	(O)	26	74	100
	(E)	18	82	
Total		36	164	200