



Centurion
UNIVERSITY
*Shaping Lives...
Empowering Communities...*

Chi-Square Test



Centurion
UNIVERSITY
*Shaping Lives...
Empowering Communities...*

- 1) Chi-Square Analysis-Basics
- 2) Chi-Square Test-Goodness of Fit
- 3) Chi-Square Test of Independence
- 4) ANOVA-Basics
- 5) ANOVA- One-Way Classification
- 6) ANOVA-Two -Way Classification



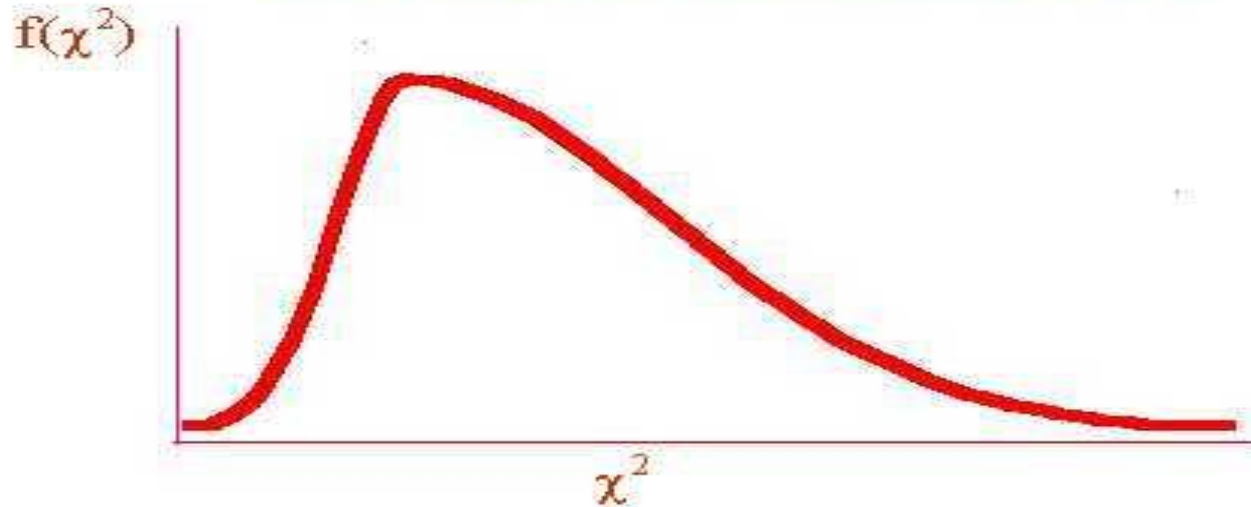
Chi-Square Analysis-Basics

Chi-Square analysis is widely used in research studies for testing hypothesis involving nominal data. **Nominal data are also known by two names- categorical data and attribute data.** The symbol χ^2 statistics is used to designate the chi-square distribution whose value depends on the number of degrees of freedom (d.f.). A chi-square distribution is a skewed distribution particularly with smaller d.f. As the sample size and therefore the d.f. increases, the χ^2 distributions becomes a symmetrical distribution approaching normality. The general shape of the χ^2 distributions for smaller d.f. is given in the next slide



Chi-Square Analysis-Basics

Shape of the Chi-Square Distribution-d.f small



Chi-Square Analysis-Basics

- The χ^2 tests is a **nonparametric** test. Nonparametric means no assumption needs to be made about the form of the original probability distribution from which the samples are drawn.
- It is a classic nonparametric test involving data measurement in nominal scale.
- Please note that all parametric tests make the assumption that the samples are drawn from a specified or assumed population. Thus, nonparametric methods are also called “***distribution free***” methods.

Chi-Square Analysis-Basics

Conditions for Using Chi-Square Test

- The sample observations drawn from a population must be independent and random
- The data must be in **frequency** (counting) form. If the original data are in percentages, they must be converted into frequency.
- No frequency in any cell/category must be less than 5. If the frequency is less than 5 for a category, you have to do some regrouping



Chi-Square Test-Goodness of Fit

χ^2 Test Goodness of Fit : Nominal Data:

This test is used to examine whether a set of observed frequencies comes from a universe that has a particular distribution (e.g. normal distribution).

This can also be used to know whether some observed pattern of frequencies fit well with an expected pattern of frequencies.



Chi-Square Test-Goodness of Fit

Nominal Data: χ^2 Test Univariate Goodness of Fit

Test Statistic

$$\chi^2 = \sum (O_i - E_i)^2 / E_i$$

Where O_i = Observed Frequency for the i th category

E_i = Expected Frequency for the i th category



Centurion
UNIVERSITY

Example:

Chi-Square Test-Goodness of Fit

In consumer marketing, a common problem the marketing manager encounters is the selection of the appropriate package design. Assume that a marketer wishes to compare five different package designs. He is interested in knowing which is the most preferred one so that the same can be introduced in the market. A random sample of 200 consumers gives the following picture:

Package Design	Preference by Consumers
A	36
B	52
C	40
D	35
E	37
Total	200

Are the consumer preferences for the designs show any significant differences?



Chi-Square Test-Goodness of Fit

Centurion
UNIVERSITY

Solution:
*Shaping Lives
Empowering Communities...*

Null Hypothesis: All package designs are equally preferred.

Alternative Hypothesis: No, they are not equally preferred

Package Design	Observed(O)	Expected(E)	$(O - E)^2$	$\left(\frac{(O - E)^2}{E}\right)$
A	36	40	16	0.400
B	52	40	144	3.600
C	40	40	0	0.000
D	35	40	25	0.625
E	37	40	9	0.225
Total	200	200		4.850

$$\chi^2 = \sum \left(\frac{(O - E)^2}{E} \right) = 4.850$$

χ^2

The critical value for 4 d.f at 5% level of significance is 9.49. Since the calculated value is less than critical value at 5% level, accept the null hypothesis of equal preference. The conclusion is that all packages are equally preferred and difference in preference in the sample survey may have arisen due to chance.



Chi-Square Test-Cross Tab

- The goodness-of-fit test is suitable for situations involving one categorical variable (e.g. package design). If there are two categorical variables, and our interest is to find out whether these two variables are associated with each other, the test of independence is the appropriate technique to use. This test is very popular for analyzing cross-tabulations in which an investigator is keen to find out whether the two categorical variables are having any relationship with each other.



Chi-Square Test-Cross Tab

Centurion
UNIVERSITY
*Shaping Lives...
Empowering Communities...*

Example:

In a market survey conducted to examine whether the choice of a brand is related to the income strata of the consumers, a random sample of 600 consumers reveal the following:

Income Strata (Income Per month)	Brand1	Brand2	Brand3	Total
Less thanRs.10000	132	128	50	310
Rs10000-15000	62	60	28	150
Rs15000-20000	30	30	26	86
Above Rs 20000	16	22	16	54
Total	240	240	120	600

The manger who conducted this survey wants to know whether the brand preference is associated with the income strata.



Chi-Square Test-Cross Tab

Centurion
UNIVERSITY
*Shaping Lives...
Empowering Communities...*

Solution:

The null hypothesis is that there is no association between the brand preference and the income level (These two are independent). The alternative hypothesis is that the brand and income level are associated (dependent).

Let us take a level of significance of 5%.

In order to calculate the χ^2 value, you need to work out the expected frequency in each cell in the contingency table. In our example, there are 4 rows and 3 columns amounting to 12 elements. There will be 12 expected frequencies.



Chi-Square Test-Cross Tab

Observed Frequencies

	Brand1	Brand2	Brand3
Income Strata			
Less than 10000	132	128	50
10000 to 15000	62	60	28
15000 to 20000	30	30	26
Above 20000	16	22	16

Expected Frequencies

	Brand1	Brand2	Brand3
Income Strata			
Less than 10000	124	124	62
10000 to 15000	60	60	30
15000 to 20000	34.4	34.4	17.2
Above 20000	21.6	21.6	10.8



Chi-Square Test-Cross Tab

Compute

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

There are 12 observed frequencies (O) and 12 expected frequencies (E). As in the case of the goodness of fit, calculate this value. In our case, the computed

$$\chi^2 = \sum \left(\frac{(O-E)^2}{E} \right) = 12.76.$$

The upper χ^2 value at 5% level for 6 d.f = 12.59.

The null hypothesis is rejected. The conclusion is that the brand preference and income level are associated.



Centurion
UNIVERSITY
*Shaping Lives...
Empowering Communities...*

]The following statements are regarding Chi-square test. Are they true or false?

1. The sample drawn must be independent and random
2. The samples drawn must be from a normal population
3. The data measurements must be categorical
4. The data must be expressed in percentage form.