The Chi-square Distribution

Definition: If $x_1, x_2, ..., x_n$ be a random sample from a normal population, $N(\mu, \sigma^2)$, then

$$\chi^2 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}{\sigma^2}$$

follows χ^2-distribution with $(n-1)$ degrees of freedom.

Applications of χ^2-test

The test based on χ^2-distribution is known as χ^2-test. It is one of the simplest and most widely used non-parametric test in statistical analysis.

Some of the applications of χ^2-test are as follows:

1. χ^2-test for 'goodness of fit'.
2. χ^2-test for independence of attribute.

Goodness-of fit Test

- χ^2-test of goodness of fit is a very powerful test for testing whether observed frequencies of an experiment follow a certain pattern or theoretical distribution.
- The test is called a goodness-of-fit test because the hypothesis tested is how **good** the observed frequencies **fit** a given pattern.

Hypotheses

H_0: The observed frequencies of an experiment follow a particular pattern. (Or, the observed frequencies fit a given pattern).

H_1: The observed frequencies do not fit the given pattern.

The Test Statistic:

The test statistic for χ^2-test is

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

Where O = Observed frequency
E = Expected frequency

The Decision Criterion using Critical Value:

Let χ^2 = the calculated value of test statistic
and χ^2_α = the critical value (obtained from table at α level of significance for $\nu = n-1$ degrees of freedom)

Conclusion: Reject H_0, if $\chi^2 \geq \chi^2_\alpha$
Accept H_0, if $\chi^2 < \chi^2_\alpha$

<u>The Decision Criterion using *p* - value:</u>

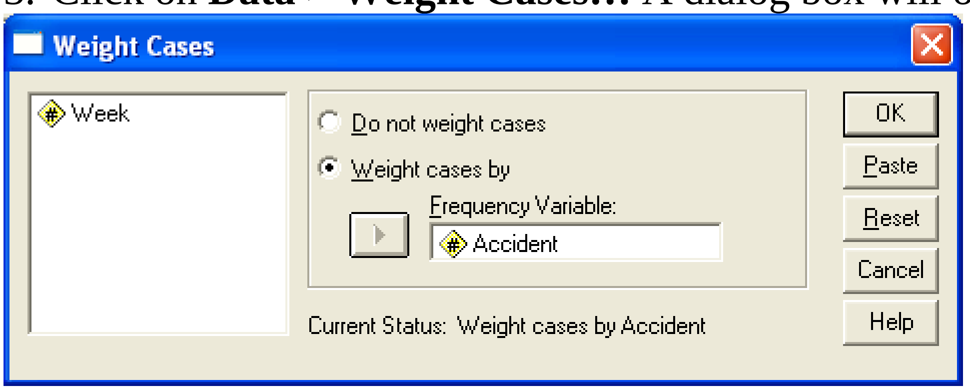
Reject H_0, if p-value $\leq \alpha$

Accept H_0, if p-value $> \alpha$

Example 1: The number of automobile accidents per week in a city are as follows: 12, 8, 20, 2, 14, 10, 15, 6, 9, 4. Are these frequencies in agreement with the belief that accident conditions were same during this 10 weeks period?

Solution: (Using SPSS)

1. Enter the two sets of data into separate columns of the Data Editor.
2. Name the column by clicking the **Variable View** tab at the bottom of Data Editor. Type **Week** and **Accident** in 1^{st} and 2^{nd} rows of **Name** column. Return to Data Editor by clicking **Data View.**
3. Click on **Data > Weight Cases…** A dialog box will open.



4. Select **Weight cases by,** double click on **Accident** to move the variable in the box and then click **OK**.
5. Select **Analyze > Nonparametric Test > Chi-Square…** A dialog box will open.
6. Move the categorical variable **Week** to the box **Test Variable List.**
7. Click **All categories equal > OK**.

The result will be displayed in the 'Output Viewer' as shown below.

Chi-Square Test

Frequencies

Week

	Observed N	Expected N	Residual
1	12	10.0	2.0
2	8	10.0	-2.0
3	20	10.0	10.0
4	2	10.0	-8.0
5	14	10.0	4.0
6	10	10.0	.0
7	15	10.0	5.0
8	6	10.0	-4.0
9	9	10.0	-1.0
10	4	10.0	-6.0
Total	100		

Test Statistics

	Week
Chi-Square(a)	26.600
df	9
Asymp. Sig.	.002

a 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 10.0.

Conclusion: Since the *P*-value = 0.002 (see **Sig**) is less than $\alpha = 0.05$, we reject H_0. Hence, we may conclude that the accident conditions are not same.

Chi-square test for Independence of Attributes

χ^2-test for independence of attributes is used when we want to test:

- whether two or more attributes are independent,
- or, whether there is any association between two or more attributes.

For example: if we want to test

(i) whether smoking is the cause of cancer,
(ii) or, whether a drug is effective in controlling a disease,

we use χ^2-test for independence of attributes.

Hypotheses:

(i) H_0: The attributes are independent (i.e. there is no association or correlation between the attributes)
(ii) H_1: The attributes are not independent (i.e. there is association or correlation between the attributes)

The Test Statistic:

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

Where O = Observed frequency
E = Expected frequency

Calculation of Expected Frequency(E):

$$E = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

Degrees of freedom:

$$\nu = (r-1)(c-1)$$

Where r = No. of rows, c = No. of columns.

Example 1: In an experiment on immunization of cattle from tuberculosis the following results were obtained:

Vaccination	Tuberculosis	
	Affected	Unaffected
Inoculated	12	28
Not inoculated	13	7

Examine whether the vaccine is effective in controlling the disease at 5% level of significance.

Solution: (Using SPSS)

1. Enter all the frequency data into first column and name the column **Count**.
2. In second column (grouping variable indicating whether the score is from affected or unaffected by tuberculosis), enter code 1, if the value is affected and enter 2, if the value is unaffected. Name the column **Tuberculosis.**
3. In third column, enter codes to identify the status of vaccination. Enter the code 1 for Inoculated and code 2 for Not-inoculated. Name the column **Vaccination.**

The Data Editor window looks as follows:

Count	Tuberculosis	Vaccination
12	1	1
28	2	1
13	1	2
7	2	2

4. Click on **Data > Weight Cases…** A dialog box will open.
5. Select **Weight cases by,** double click on **Count** to move frequency data in the box and then click **OK**.

6. Choose **Analyze > Descriptive statistics > Crosstab….**
7. In **For rows:** Enter the variable **Vaccination** containing the categories that define the rows of the table.
8. In For **columns:** Enter the variable **Tuberculosis** containing the categories that define the columns of the table.
9. Under **Exact:** You could select the exact test if the expected cell frequency is less than 5.
10. Under **Statistics:** Select Chi-square value.
11. Under **Cell**: You can select expected frequencies, percentage, etc.
12. Click **OK**.

The Output Viewer will display the results of the analysis. A part of the results is shown below:

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	**6.720**(b)	1	**.010**		
Continuity Correction(a)	5.357	1	.021		
Likelihood Ratio	6.736	1	.009		
Fisher's Exact Test				.013	.010
Linear-by-Linear Association	6.608	1	.010		
N of Valid Cases	60				

a Computed only for a 2x2 table
b 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.33.

Conclusion: Since the P-value = 0.010 (see **Sig**) is less than $\alpha = 0.05$, we reject H_0. Hence, we may conclude that the vaccine is effective in controlling the disease.