Chapter 3
Chi-Square Test and Its Application

Learning Objectives

After completing this chapter you should be able to do the following:

- Know the use of chi-square in analyzing nonparametric data.
- Understand the application of chi-square in different research situations.
- Know the advantages of crosstabs analysis.
- Learn to construct the hypothesis in applying chi-square test.
- Explain the situations in which different statistics like contingency coefficient, lambda coefficient, phi coefficient, gamma, Cramer's V, and Kendall tau, for measuring an association between two attributes, can be used.
- Learn the procedure of data feeding in preparing the data file for analysis using SPSS.
- Describe the procedure of testing an equal occurrence hypothesis and testing the significance of an association in different applications by using SPSS.
- Interpret the output of chi-square analysis generated in SPSS.

Introduction

In survey research, mainly two types of hypothesis are tested. One may test goodness of fit for a single attribute or may like to test the significance of association between any two attributes. To test an equal occurrence hypothesis, it is required to tabulate the observed frequency for each variable. The chi-square statistic in "nonparametric" section of SPSS may be used to test the hypothesis of equal occurrence.

The scores need to be arranged in contingency table for studying an association between any two attributes. A contingency table is the arrangement of frequency in rows and column. The process of creating a contingency table from the observed frequency is known as crosstab. The cross tabulation procedure provides tabulation of two variables in two-way table. A frequency distribution provides the distribution of one variable, whereas a contingency table describes the distribution of two or more variables simultaneously (Table 3.1).

J.P. Verma, *Data Analysis in Management with SPSS Software*,
DOI 10.1007/978-81-322-0786-3_3, © Springer India 2013

Table 3.1 Preferences of male and female towards different incentives

		Incentives		
		Gift check (%)	Cash (%)	Gift article (%)
Gender	Male	30	45	25
	Female	10	30	60

The following is an example of a 2 × 3 contingency table. The first variable "gender" has two options, male and female, whereas the second variable "incentives" has three options, gift check, cash, and gift article. Each cell gives the number of individuals who share the combination of traits.

The chi-square can be computed by using the Crosstabs option in "Descriptive Statistics" section of SPSS command. Besides chi-square, the Crosstabs option in SPSS provides the output showing magnitude of association along with summary statistics including percentage of frequency and expected frequency in each cell.

Two-way tabulation in Crosstabs can be used to establish an interdependent relationship between two tables of values but does not identify a causal relation between the values. Cross tabulation technique can be used to analyze the results of a survey study, for example, it may indicate a preference for certain types of jobs based on the respondent's gender.

Advantages of Using Crosstabs

1. Crosstabs analysis is easy to understand and is good for the researchers, who do not want to use more sophisticated statistical techniques.
2. Crosstabs treats all data as nominal. In other words, the data is treated as nominal even if it is measured in interval, ratio, or ordinal form.
3. A table is more explanatory than a single statistics.
4. They are simple to conduct.

Statistics Used in Cross Tabulations

In Crosstabs analysis, usually statistics like chi-square, contingency coefficient, lambda coefficient, phi coefficient, Kendall tau, gamma, or Cramer's V are used. These shall be discussed below:

Chi-Square Statistic

If $X_1, \ldots, X_n$ are independent and identical $N(\mu, \sigma^2)$ random variables, then the statistics $\sum_{i=1}^{n}\left(\frac{X_i-\mu.}{\sigma}\right)^2$ follows the chi-square distribution with $(n-1)$ degrees of freedom and is written as

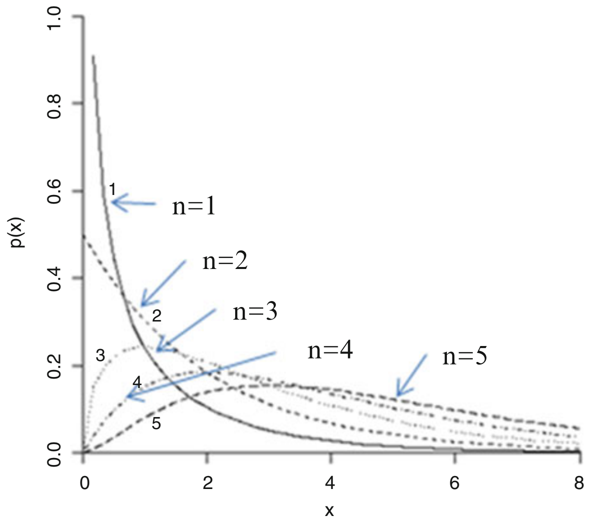


Fig. 3.1 Probability distribution of chi-square for different degrees of freedom

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n-1)$$

The probability density function of the chi-square (χ^2) random variable is

$$f(x)=\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2} \tag{3.1}$$

$$x>0 \text{ and } \mathrm{n}=1,2,3,\ldots$$

The mean and variance of the chi-square statistics are n and $2n$, respectively. The χ^2 distribution is not unique but depends upon degrees of freedom. The family of distribution with varying degrees of freedom is shown in Fig. 3.1.

Additive Properties of Chi-Square

If χ_1^2 and $\chi_2^2$are two independent chi-square variates with n_1 and n_2 degrees of freedom, respectively, then $\chi_1^2+\chi_2^2$ is also a chi-square variate with n_1 + n_2 degrees of freedom. This property is used extensively in the questionnaire studies. Consider a study to compare the attitude of male and female consumers about a particular brand of car. The questionnaire may consist of questions under three factors, namely, financial consideration, driving comforts, and facilities. Each of these factors may have several questions. On each of the questions, attitude of male and female users may be compared using chi-square. Further, by using additive properties, the chi-square of each question under a particular factor may be added to compare the attitude of male and female on that factor.

Chi-Square Test

Chi-square test is the most frequently used nonparametric statistical test. It is also known as Pearson chi-square test and provides us the mechanism to test the independence of two categorical variables. The chi-square test is based upon a chi-square distribution just like the way a *t*-test is based upon *t*-distribution or an *F*-test is based upon an *F*-distribution. The results of the Pearson's chi-square test are evaluated by referencing to the chi-square distribution.

The chi-square statistic is denoted as $\chi^2$and is pronounced as kai-square. The properties of chi-square were first investigated by Karl Pearson in 1900 and hence named after Karl Pearson chi-square test.

In using chi-square test, the chi-square (χ^2) statistic is computed as

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_\mathrm{o} - f_\mathrm{e})^2}{f_\mathrm{e}} \tag{3.2}$$

where f_o and f_e are the observed and expected frequencies for each of the possible outcome, respectively.

Steps in the Chi-Square Test

The following steps are used in chi-square test:

1. Compute expected frequency for each of the observed frequency. The procedure for computing expected frequency is different in case of testing the goodness of fit and in testing the independence of attributes. This will be discussed later in the chapter while solving the example.
2. Calculate the value of chi-square statistic χ^2 by using the formula (3.2).
3. Find degrees of freedom of the test. In testing the goodness of fit, the degrees of freedom is equal to $(r - 1)$, where r is the number of categories in the population. On the other hand, in testing the independence of attributes, the degrees of freedom is obtained by $(r - 1) \times (c - 1)$, where r and c are the number of rows and columns, respectively.
4. Find the tabulated value of $\chi^2$with required degrees of freedom and at a given level of significance from Table A.6 in the Appendix.
5. If the calculated $\chi^2$is less than or equal to tabulated χ^2, the null hypothesis is failed to be rejected, and if the calculated $\chi^2$is greater than the tabulated χ^2, the null hypothesis is rejected at the tested level of significance.

Assumptions in Using the Chi-Square

While using chi-square test, following assumptions are made:

1. Sample must be random.
2. Frequencies of each attribute must be numeric and should not be in percentages or ratios.
3. Sample size must be sufficiently large. The chi-square test shall yield inaccurate findings if the sample size is small. In that case, the researcher might end up committing a type II error.
4. The observations must be independent of each other. In other words, the chi-square test cannot be used to test the correlated data. In that situation, McNemar's test is used.
5. Normally, all cell frequencies must be 5 or more. In large contingency tables, 80% of cell frequencies must be 5 or more. If this assumption is not met, the Yates' correction is applied.
6. The expected frequencies should not be too low. Generally, it is acceptable if 20% of the events have expected frequencies less than 5, but in case of chi-square with one degree of freedom, the conclusions may not be reliable if expected frequencies are less than 10. In all such cases, Yates' correction must be applied.

Application of Chi-Square Test

The chi-square test is used for two purposes: first, to test the goodness of fit and, second, to test the independence of two attributes. In both the situations, we intend to determine whether the observed frequencies significantly differ from the theoretical (expected) frequencies. The chi-square tests in these two situations shall be discussed in the following sections:

To Test the Goodness of Fit

In many decision-making situations, a marketing manager may like to know whether the pattern of frequencies that are observed fits well with the expected ones or not. The appropriate test in such situations is the χ^2 test of goodness of fit. Thus, a chi-square test for goodness of fit is used to verify whether an observed frequency distribution differs from a theoretical distribution or not. This test can also be used to check whether the data is from any specific distribution like normal, binomial or Poisson. The chi-square test for goodness of fit can also be used to test an equal occurrence hypothesis. By using this test, one can test whether all brands are equally popular, or whether all the car models are equally preferred. In using the chi-square test for goodness of fit, only one categorical variable is involved.

Consider a situation in which a researcher is interested to know whether all the three specializations like finance, human resource, and marketing are equally popular among MBA students; an equal occurrence hypothesis may be tested by

Table 3.2 Preferences of the college students about different brands of cold drinks

Color	White	Orange	Brown
Frequencies	50	40	30

Table 3.3 Observed and expected frequencies of responses

	Observed frequencies	Expected frequencies
	(f_o)	(f_e)
White	50	40
Orange	40	40
Brown	30	40

computing the chi-square. The "Nonparametric Tests" option in SPSS provides the computation of chi-square (χ^2). In such situations, following set of hypotheses is tested:

H_0: All three specializations are equally popular.
H_1: All three specializations are not equally popular.

By using the procedure discussed above for applying chi-square test, the null hypothesis may be tested. The procedure would clear by looking to the following solved examples:

Example 3.1 A beverages company produces cold drink with three different colors. One hundred and twenty college students were asked about their preferences. The responses are shown in Table 3.2. Do these data show that all the flavors were equally liked by the students? Test your hypothesis at .05 level of significance.

Solution Here it is required to test the null hypothesis of equal occurrence; hence, expected frequencies corresponding to each of the three observed frequencies shall be obtained by dividing the total of all the observed frequencies by the number of categories. Hence, expected frequency (f_e) for each category shall be (Table 3.3)

$$f_e = \frac{50 + 40 + 30}{3} = 40$$

Here, number of categories or rows (r) = 3 and number of columns (c) = 2.

$$\begin{aligned}\chi^2 &= \sum_{i=1}^{r} \frac{(f_o - f_e)^2}{f_e} \\ &= \frac{(50-40)^2}{40} + \frac{(40-40)^2}{40} + \frac{(30-40)^2}{40} \\ &= \frac{100}{40} + 0 + \frac{100}{40} = 2.5 + 2.5\end{aligned}$$

$$\Rightarrow \qquad \text{Cal. } \chi^2 = 5.0$$

Table 3.4 Observed and expected frequencies

	Observed frequencies	Expected frequencies
	(f_o)	(f_e)
Grade A	90	75
Grade B	65	50
Grade C	60	75
Grade D	85	100

Testing the Significance of Chi-Square

The degrees of freedom = $(r-1) = 3-1 = 2$.

From Table A.6 in the Appendix, the tab $\chi^2_{.05}(2) = 5.991$.

Since Cal. $\chi^2 <$ Tab. $\chi^2_{.05}(2)$, the null hypothesis may not be rejected at .05 level of significance. Thus, it may be concluded that all the three colors of cold drinks are equally liked by the college students.

Example 3.2 An examination was undertaken by 300 students, out of which 90 students had grade A, 65 got grade B, 60 got grade C, and the remaining had grade D. Do these figures commensurate with the final examination result which is in the ratio of 3:2:3:4 for various grades, respectively? Test the hypothesis at 5% level.

Solution The null hypothesis which is required to be tested here is

H_0: The students in the various grades were distributed in the ratio 3:2:3:4

The expected number of students (f_e) for each grade under the assumption that H_0 is true is as follows:

$$\text{Expected number of students getting grade A} = \frac{3}{3+2+3+4} \times 300 = 75$$

$$\text{Expected number of students getting grade B} = \frac{2}{12} \times 300 = 50$$

$$\text{Expected number of students getting grade C} = \frac{3}{12} \times 300 = 75$$

$$\text{Expected number of students getting grade D} = \frac{4}{12} \times 300 = 100$$

Thus, the observed and expected frequencies can be listed as shown in Table 3.4.

$$\chi^2 = \sum_{i=1}^{r} \frac{(f_o - f_e)^2}{f_e}$$

$$= \frac{(90-75)^2}{75} + \frac{(65-50)^2}{50} + \frac{(60-75)^2}{75} + \frac{(85-100)^2}{100}$$

$$= \frac{225}{75} + \frac{225}{50} + \frac{225}{75} + \frac{225}{100} = 3 + 4.5 + 3 + 2.25 = 12.75$$

$\Rightarrow$ Calculated $\chi^2 = 12.75$

Testing the Significance of Chi-Square

Degrees of freedom: $(r - 1)$ $4-1 = 3$.

From Table A.6 in the Appendix, the tab $\chi^2_{.05}(3) = 7.815$.

Since Cal. $\chi^2 >$ Tab. $\chi^2_{.05}(3)$, the null hypothesis may be rejected at .05 level of significance. It may thus be concluded that grades A, B, C, and D are not in proportion to 3:2:3:4.

To Test the Independence of Attributes

The chi-square test of independence is used to know whether paired observations on two attributes, expressed in a contingency table, are independent of each other. There may be varieties of situations where chi-square test of independence may be used. For instance one may test the significance of association between income level & brand preference, family size & television size purchased or educational background & the type of job one does. Thus, chi-square test may be used to test the significance of an association between any two attributes.

Let us assume that the population can be classified into r mutually exclusive classes $A_1, A_2, \ldots, A_r$ on the basis of attribute A, and each of these r classes are further classified into c mutually exclusive classes, like $A_iB_1, A_iB_2, \ldots, A_iB_c$, etc.

If $f_{\text{o}_{ij}}$ is the observed frequency of A_iB_j, that is, $(A_iB_j) = f_{\text{o}_{ij}}$, the above classification can be shown in the following table known as contingency table.

B / A	B_1	B_2	...	B_c	Total
A_1	$f_{\text{o}_{11}}$	$f_{\text{o}_{12}}$	...	$f_{\text{o}_{1c}}$	(A_1)
A_2	$f_{\text{o}_{21}}$	$f_{\text{o}_{22}}$	...	$f_{\text{o}_{2c}}$	(A_2)
.					
.					
A_r	$f_{\text{o}_{r1}}$	$f_{\text{o}_{r2}}$	...	$f_{\text{o}_{rc}}$	(A_r)
Total	(B_1)	(B_2)	...	(B_c)	N

By assuming A and B as independent attributes, the expected frequencies of each cell can be computed as

$$E_{ij} = \frac{(A_i)(B_j)}{N}$$

Thus,

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(f_{\text{o}_{ij}} - f_{\text{e}_{ij}}\right)^2}{f_{\text{e}_{ij}}} \tag{3.3}$$

shall be a χ^2 variate with $(r - 1)(c - 1)$ degrees of freedom.

The value of the χ^2 variate so obtained can be used to test the independence of two attributes.

Consider a situation where it is required to test the significance of association between Gender (male and female) and Response ("prefer day shift" and "prefer night shift"). In this situation, following hypotheses may be tested:

H_0: Gender and Response toward shift preferences are independent.
H_1: There is an association between the Gender and Response toward shift preferences.

The calculated value of chi-square (χ^2) obtained from the formula (3.3) may be compared with that of its tabulated value for testing the null hypothesis.

Thus, if calculated χ^2 is less than tabulated χ^2 with $(r - 1)$ $(c - 1)$ df at some level of significance, then H_0 may not be rejected otherwise H_0 may be rejected.

Remark If H_0 is rejected, we may interpret that there is a significant association between the gender and their preferences toward shifts. Here, significant association simply means that the response pattern of male and female is different. The readers may note that chi-square statistic is used to test the significance of association, but ultimately one gets the comparison between the levels of one attribute across the levels of other attribute.

Example 3.3 Five hundred families were investigated to test the belief that high-income people usually prefer to visit private hospitals and low-income people often go to government hospitals whenever they fall sick. The results so obtained are shown in Table 3.5.

Test whether income and hospital preferences are independent. Compute the contingency coefficient to find the strength of association. Test your hypothesis at 5% level.

Solution The null hypothesis to be tested is

H_0: Income and hospital preferences are independent.

Before computing the value of chi-square, the expected frequencies for each cell need to be computed with the marginal totals and grand totals given in the observed frequency (f_o) table. The procedure is discussed in Table 3.6.

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(f_{o_{ij}} - f_{e_{ij}}\right)^2}{f_{e_{ij}}}$$

$$= \frac{(125 - 140.4)^2}{140.4} + \frac{(145 - 129.6)^2}{129.6} + \frac{(135 - 119.6)^2}{119.6} + \frac{(95 - 110.4)^2}{110.4}$$

$$= 1.69 + 1.83 + 1.98 + 2.15 = 7.65$$

$$\Rightarrow \text{Calculated } \chi^2 = 7.65$$

Table 3.5 Observed frequencies (f_o) of responses

Hospitals			
Income	Government	Private	Total
High	125	145	270
Low	135	95	230
Total	260	240	500

Table 3.6 Expected frequencies (f_e) of responses

Hospitals			
Income	Government	Private	Total
High	$\frac{270 \times 260}{500} = 140.4$	$\frac{270 \times 240}{500} = 129.6$	270
Low	$\frac{230 \times 260}{500} = 119.6$	$\frac{230 \times 240}{500} = 110.4$	230
Total	260	240	500

Test of Significance

Here, $r = 2$ and $c = 2$, and therefore degree of freedom is $(r - 1) \times (c - 1) = 1$.

From Table A.6 in the Appendix, the tab $\chi^2_{.05}(1) = 3.841$.

Since Cal. $\chi^2 >$ Tab. $\chi^2_{.05}(1)$, the null hypothesis may be rejected at .05 level of significance. It may therefore be concluded that there is an association between the income level and the types of hospital preferred by the people.

Precautions in Using the Chi-Square Test

(a) While using chi-square test, one must ensure that the sample is random, representative, and adequate in size.

(b) Chi-square should not be calculated if the frequencies are in percentage form; in that case, these frequencies must be converted back to absolute numbers before using the test.

(c) If any of the cell frequencies is less than 5, then for each cell, .5 is subtracted from the difference of observed and expected frequency while computing chi-square. This correction is known as Yates' correction. SPSS automatically does this correction while computing chi-square.

(d) The sum of the observed frequencies should be equal to the sum of the expected frequencies.

Testing the Significance of Chi-Square in SPSS

(a) In SPSS, the null hypothesis is not tested on the basis of the comparison between calculated and tabulated chi-square; rather, it uses the concept of *p* value. *p* value is the probability of rejecting the null hypothesis when actually it is true.

(b) Thus, the chi-square is said to be significant at 5% level if the p value is less than .05 and is insignificant if it is more than .05.

Contingency Coefficient

Contingency coefficient (C) provides the magnitude of association between the attributes in the cross tabulation. Its value can range from 0 (no association) to 1 (the theoretical maximum possible association). Chi-square simply tests the significance of an association between any two attributes but does not provide the magnitude of the association. Thus, if the chi-square value becomes significant, one must compute the contingency coefficient (C) to know the extent of association between the attributes. The contingency coefficient C is computed by the following formula:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} \tag{3.4}$$

where N is the sum of all frequencies in the contingency table.

Lambda Coefficient

Lambda coefficient is used to test the strength of an association in the cross tabulation. It is assumed that the variables are measured at the nominal level. Lambda can have the value in the range 0 (no association) to 1 (the theoretical maximum possible association). *Asymmetric lambda* measures the percentage improvement in predicting the dependent variable. *Symmetric lambda* measures the percentage improvement when prediction is done in both directions.

Phi Coefficient

In a situation when both the variables are binary, phi coefficient is used to measure the degree of association between them. This measure is similar to the correlation coefficient in its interpretation. Two binary variables are considered positively associated if most of the data falls along the diagonal cells. In contrast, two binary variables are considered negatively associated if most of the data falls off the diagonal.

The assumptions of normality and homogeneity can be violated when the categories are extremely uneven, as in the case of proportions close to .90, .95 or

.10, .05. In such cases, the phi coefficient can be significantly attenuated. The assumption of linearity cannot be violated within the context of the phi coefficient of correlation.

Gamma

If both the variables are measured at the ordinal level, *Gamma* is used for testing the strength of association of the cross tabulations. It makes no adjustment for either table size or ties. The value of Gamma can range from −1 (100% negative association, or perfect inversion) to +1 (100% positive association, or perfect agreement). A value of zero indicates the absence of association.

Cramer's V

It measures the strength of association between attributes in cross tabulations. It is a variant of the *phi coefficient* that adjusts for the number of rows and columns. Its value can range from 0 (no association) to 1 (the theoretical maximum possible association).

Kendall Tau

Tau b and Tau c both test the strength of association of the cross tabulations in a situation when both variables are measured at the ordinal level. Both these tests *Tau b* and *Tau c* make adjustments for ties, but *Tau b* is most suitable for square tables whereas *Tau c* is most suitable for rectangular tables. Their values can range from −1 (100% negative association, or perfect inversion) to +1 (100% positive association, or perfect agreement). A value of zero indicates the absence of association.

Situation for Using Chi-Square

Chi-square is one of the most popularly used nonparametric statistical tests used in the questionnaire study. Two different types of hypotheses, that is, testing the goodness of fit and testing the significance of association between two attributes, can be tested using chi-square.

Testing an equal occurrence hypothesis is a special case of goodness of fit. In testing an equal occurrence hypothesis, the observed frequencies on different

Table 3.7 Observed frequencies in a contingency table

	Category of preference	
Socioeconomic status	Prefer	Do not prefer
High	80	15
Low	40	65

levels of a factor are obtained. The total of observed frequencies for all the levels is divided by the number of levels to obtain the expected frequencies for each level. Consider an experiment in which it is intended to test whether all the three locations, that is, Delhi, Mumbai, and Chennai, are equally preferred by the employees of an organization for posting. Out of 250 employees surveyed, 120 preferred Delhi, 80 preferred Mumbai, and 50 preferred Chennai. In this situation, the following null hypothesis may be tested using chi-square:

H_0: All the three locations are equally preferred.

Against the alternative hypothesis:

H_1: All the three locations are not equally preferred.

Here, the chi-square test can be used to test the null hypothesis of equal occurrence.

Another application of chi-square is to test the significance of association between any two attributes. Suppose it is desired to know as to whether preference of consumers for a specific brand of soap depends upon their socioeconomic status where the response of 200 customers is shown in Table 3.7.

The following null hypothesis may be tested by using the chi-square for two samples at 5% level to answer the question.

H_0: Socioeconomic status and soap preferences are independent.

Against the alternative hypothesis:

H_1: There is an association between the socioeconomic status and soap preferences.

If the null hypothesis is rejected, one may draw the conclusion that the preference of soap is significantly associated with the socioeconomic status of an individual. In other words, it may be concluded that the response patterns of the customers in high and low socioeconomic status are different.

The above two different kinds of application of chi-square have been discussed below by means of solved examples using SPSS.

Solved Examples of Chi-square for Testing an Equal Occurrence Hypothesis

Example 3.4 In a study, 90 workers were tested for their job satisfaction. Their job satisfaction level was obtained on the basis of the questionnaire, and the respondents were classified into one of the three categories, namely, low, average, and high. The observed frequencies are shown in Table 3.8. Compute chi-square in testing whether there is any specific trend in their job satisfaction.

Table 3.8 Summary of responses of the workers about their job satisfaction levels

Job satisfaction level		
Low	Average	High
40	30	20

Solution Here, the null hypothesis that is required to be tested is

H_0: All the three job satisfaction levels are equally probable.

Against the alternative hypothesis:

H_1: All the three job satisfaction levels are not equally probable.

The SPSS shall be used to compute the value of chi-square for testing the null hypothesis. Computation of chi-square for single sample using SPSS has been shown in the following steps:

Computation of Chi-Square Using SPSS

(a) ***Preparing Data File***

Before using SPSS command to compute chi-square, a data file needs to be prepared. The following steps will help you prepare the data file:

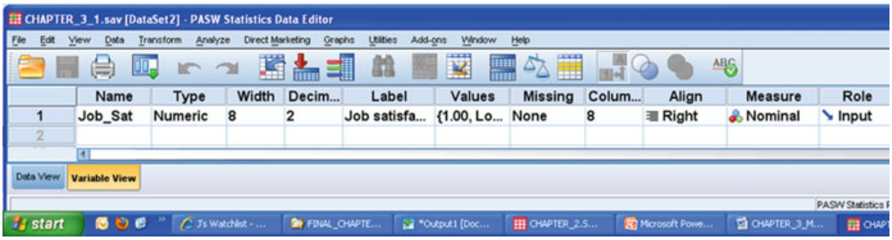
(i) *Starting the SPSS:* Use the following command sequence to start SPSS:

Start → All Programs → IBM SPSS Statistics → IBM SPSS Statistics 20

After checking the option **Type in Data** on the screen you will be taken to the **Variable View** option for defining the variables in the study.

(ii) *Defining variables:* There is only one variable *Job Satisfaction Level* that needs to be defined. Since this variable can assume any one of the three values, it is a nominal variable. The procedure of defining the variable in SPSS is as follows:

1. Click **Variable View** to define variables and their properties.
2. Write short name of the variable, that is, *Job_Sat* under the column heading **Name.**
3. For this variable, define the full name, that is, *Job Satisfaction Level* under the column heading **Label**.
4. Under the column heading **Values,** define "1" for low, "2" for medium, and "3" for high.
5. Under the column heading **Measure,** select the option "Nominal" because the variable Job_Sat is a nominal variable.
6. Use default entries in rest of the columns.



CHAPTER_3_1.sav [DataSet2] - PASW Statistics Data Editor

	Name	Type	Width	Decim...	Label	Values	Missing	Colum...	Align	Measure	Role
1	Job_Sat	Numeric	8	2	Job satisfa...	{1.00, Lo...	None	8	Right	Nominal	Input
2											

Fig. 3.2 Defining variable along with its characteristics

After defining the variables in variable view, the screen shall look like Fig. 3.2.

(iii) *Entering data:* Once the variable *Job_Sat* has been defined in the **Variable View**, click **Data View** on the left bottom of the screen to open the format for entering data column wise.

In this example, we have only one variable *Job_Sat* with three levels as Low, Medium, and High. The Low satisfaction level was observed in 40 workers, whereas Medium satisfaction level was observed in 30 workers and High satisfaction level was observed in 20 workers. Since these levels have been defined as 1, 2, and 3, the data shall be entered under one variable *Job_Sat* as shown below:

Data feeding procedure in SPSS under **Data View**

S.N.	Job_Sat	
1	1	
2	1	
3	1	Type “1” 40 times as Low satisfaction level was observed in 40 workers
.	.	
.	.	
.	.	
40	1	
41	2	
42	2	
.	.	Type “2” 30 times as Medium satisfaction level was observed in 30 workers
.	.	
.	.	
70	2	
71	3	
72	3	
73	3	Type “3” 20 times as Low satisfaction level was observed in 20 workers
.	.	
.	.	
.	.	
90	3	

After entering data, the screen shall look like Fig. 3.3. Only the partial data has been shown in the figure as data set is long enough to fit in the window. Save the data file in the desired location before further processing.

(b) **SPSS *Commands for Computing Chi-Square***
After preparing the data file in data view, take the following steps to compute the chi-square:

(i) *Initiating the SPSS commands to compute chi-square for single variable:* In data view, click the following commands in sequence:

Analyze → Nonparametric Tests → Legacy Dialogs → Chi − Square

The screen shall look like Fig. 3.4.
Note: In other versions of SPSS, the command sequence is as follows:

Analyze → Nonparametric Tests → Chi-Square

(ii) *Selecting variable for computing chi-square*: After clicking the "Chi-Square" option, you will be taken to the next screen for selecting the variable for which chi-square needs to be computed. Since there is only one variable *Jobs satisfaction level* in the example, select it from the left panel by using the left click of the mouse and bring it to the right panel by clicking the arrow. The screen shall look like Fig. 3.5.

(iii) *Selecting the option for computation:* After selecting the variable, option needs to be defined for the computation of chi-square. Take the following steps:

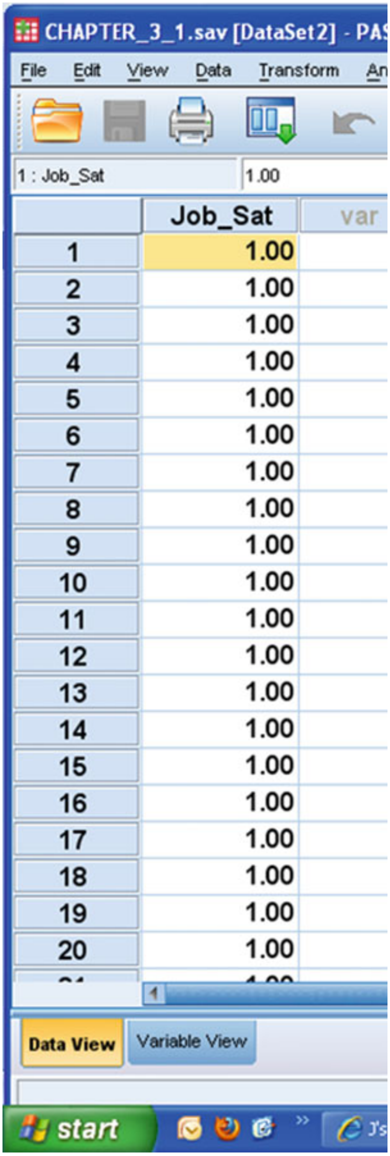
- Click the **Options** in the screen shown in Fig. 3.5. This will take you to the next screen that is shown in Fig. 3.6.
  - Check the option "Descriptive."
  - Use default entries in other options.
  - Click ***Continue***. This will take you back to screen shown in Fig. 3.5
- Press **OK.**

(c) ***Getting the Output***
Pressing **OK** will lead you to the output window. The output panel shall have two results that are shown in Tables 3.9 and 3.10. These outputs can be selected by using right click of the mouse which may be copied in the word file.

Interpretation of the Outputs

Table 3.9 shows the observed and expected frequencies of the different levels of job satisfaction. No cell frequency is less than 5, and, therefore, no correction is required while computing chi-square.

Fig. 3.3 Screen showing entered data for the variable Job_Sat in the data view

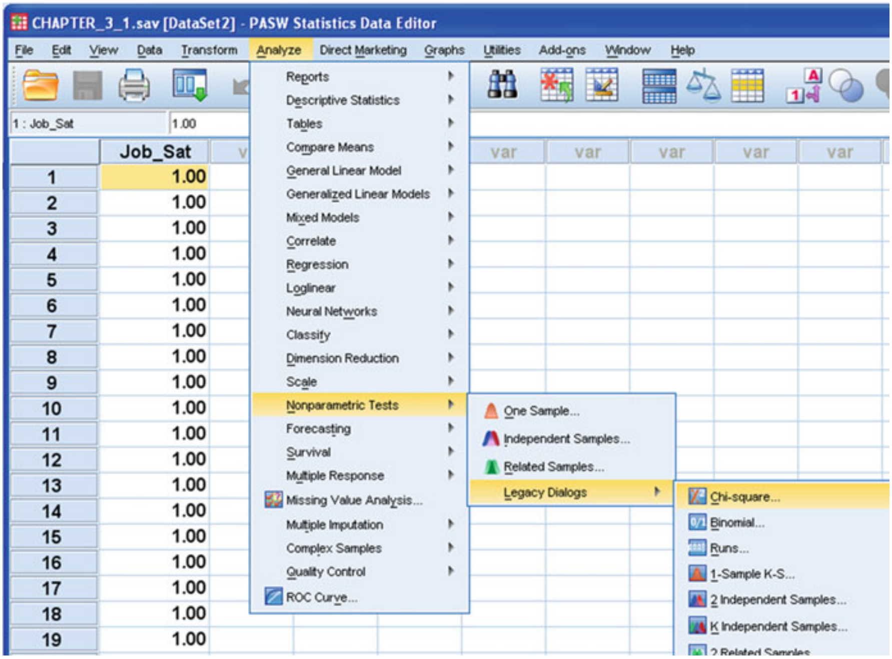


1 : Job_Sat 1.00

	Job_Sat	var
1	1.00	
2	1.00	
3	1.00	
4	1.00	
5	1.00	
6	1.00	
7	1.00	
8	1.00	
9	1.00	
10	1.00	
11	1.00	
12	1.00	
13	1.00	
14	1.00	
15	1.00	
16	1.00	
17	1.00	
18	1.00	
19	1.00	
20	1.00	

The value of chi-square (=6.667) in Table 3.10 is significant at 5% level because its associated *p* value is .036 which is less than .05. Thus, the null hypothesis may be rejected. It may therefore be concluded that all the three job satisfaction levels are not equally probable.

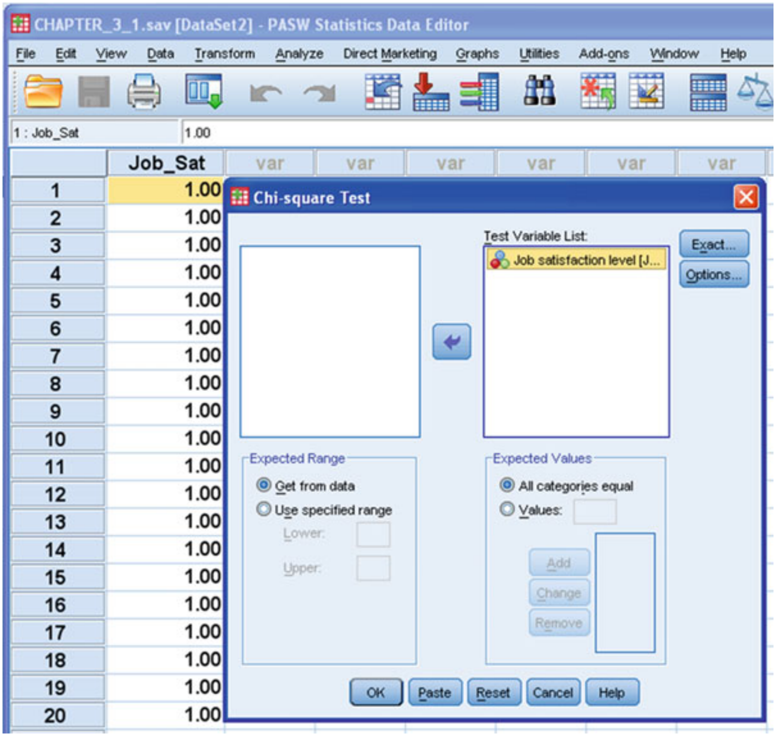
So long the value of p is less than .05, the value of chi-square is significant at 5% level, and if the *p* value becomes more than .05, the chi-square becomes insignificant.



1 : Job_Sat 1.00

	Job_Sat
1	1.00
2	1.00
3	1.00
4	1.00
5	1.00
6	1.00
7	1.00
8	1.00
9	1.00
10	1.00
11	1.00
12	1.00
13	1.00
14	1.00
15	1.00
16	1.00
17	1.00
18	1.00
19	1.00

Fig. 3.4 Screen showing the SPSS commands for computing chi-square


CHAPTER_3_1.sav [DataSet2] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
var
var
var
var
var
var
Chi-square Test
Exact...
Options...
Add
Change
Remove
OK
Paste
Reset
Cancel
Help

1 : Job_Sat 1.00

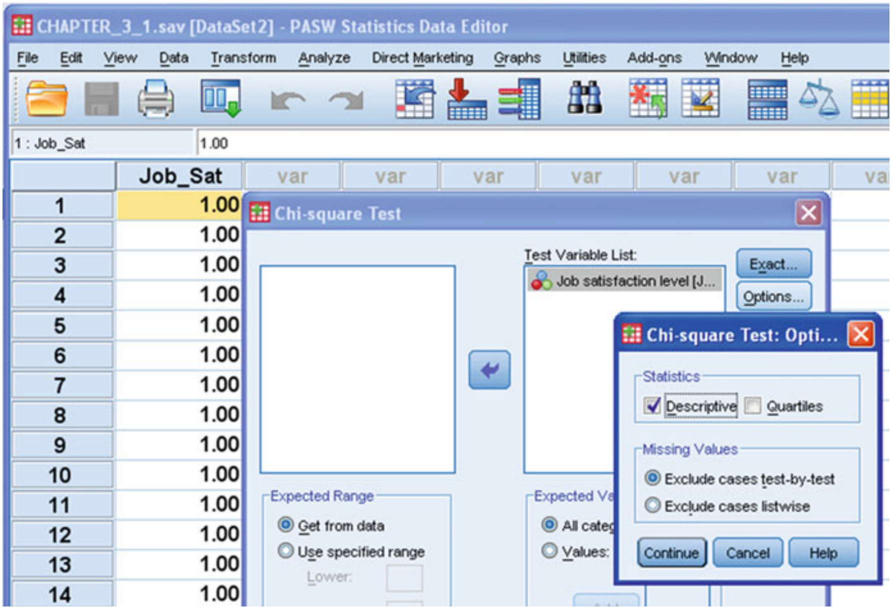
	Job_Sat
1	1.00
2	1.00
3	1.00
4	1.00
5	1.00
6	1.00
7	1.00
8	1.00
9	1.00
10	1.00
11	1.00
12	1.00
13	1.00
14	1.00
15	1.00
16	1.00
17	1.00
18	1.00
19	1.00
20	1.00

Test Variable List: Job satisfaction level [J...

Expected Range
● Get from data
○ Use specified range
Lower:
Upper:

Expected Values
● All categories equal
○ Values:

Fig. 3.5 Screen showing selection of variable for chi-square



1 : Job_Sat 1.00

	Job_Sat
1	1.00
2	1.00
3	1.00
4	1.00
5	1.00
6	1.00
7	1.00
8	1.00
9	1.00
10	1.00
11	1.00
12	1.00
13	1.00
14	1.00

Test Variable List: Job satisfaction level [J...

Expected Range: ● Get from data ○ Use specified range

Expected Va: ● All categ ○ Values:

Statistics: ☑ Descriptive ☐ Quartiles

Missing Values: ● Exclude cases test-by-test ○ Exclude cases listwise

Fig. 3.6 Screen showing option for chi-square computation

Table 3.9 Observed and expected frequencies for different levels of job satisfaction

	Frequencies		
	Observed *N*	Expected *N*	Residual
Low	40	30.0	10.0
Medium	30	30.0	.0
High	20	30.0	−10.0
Total	90		

Table 3.10 Chi-square for the data on job satisfaction level

	Job satisfaction level
Chi-square	6.667[a]
df	2
Asymp sig.	.036

[a] 0 cells (0%) have expected frequencies less than 5. The minimum expected cell frequency is 30

Solved Example of Chi-square for Testing the Significance of Association Between Two Attributes

Example 3.5 Out of 200 MBA students, 40 were given an academic counseling throughout the semester, whereas other 40 did not receive this counseling. On the basis of their marks in the final examination, their performance was categorized as improved, unchanged, and deteriorated. Based on the results shown in Table 3.11, can it be concluded that the academic counseling is effective at 5% level?

Table 3.11 Frequencies of the MBA students in a contingency table

	Performance		
Treatment	Improved	Unchanged	Deteriorated
Counseling group	22	8	10
Control group	4	5	31

Solution In order to check whether academic counseling is effective, we shall test the significance of association between treatment and performance. If the association between these two attributes is significant, then it may be interpreted that the pattern of performance in the counseling and control groups is not same. In that case, it might be concluded that the counseling is effective since the number of improved cases is higher in counseling group than that of control group.
Thus, it is important to compute the chi-square first in order to test the null hypothesis.

H_0: There is no association between treatment and performance.

Against the alternative hypothesis:

H_1: There is an association between treatment and performance.

The commands for computing chi-square in case of two samples are different than that of one sample computed in Example 3.4.

In two-sample case, chi-square is computed using **Crosstabs** option in **Descriptive statistics** command of SPSS. The chi-square so obtained shall be used for testing the above-mentioned null hypothesis. Computation of chi-square for two samples using SPSS has been shown in the following steps:

Computation of Chi-Square for Two Variables Using SPSS

(a) ***Preparing Data File***

As discussed in Example 3.4, a data file needs to be prepared for using the SPSS commands for computing chi-square. Follow the below-mentioned steps to prepare data file:

(i) *Starting the SPSS:* Use the following command sequence to start SPSS:

Start → All Programs → IBM SPSS Statistics → IBM SPSS Statistics 20

By checking the option **Type in Data** on the screen you will be taken to the **Variable View** option for defining the variables in the study.

(ii) *Defining variables:* Here, two variables *Treatment* and *Performance* need to be defined. Since both these variables are classificatory in nature, they are treated as nominal variables in SPSS. The procedure of defining variables and their characteristics in SPSS is as follows:

1. Click **Variable View** to define variables and their properties.
2. Write short name of the variables as *Treatment* and *Performance* under the column heading **Name.**

3. Under the column heading **Label**, full name of the *Treatment* and *Performance* variables may be defined as *Treatment groups* and *Performance status,* respectively. There is flexibility in choosing full name of each variable.
4. In the *Treatment* row, double-click the cell under the column **Values** and add the following values to different labels:

Value	Label
1	Counseling group
2	Control group

5. Similarly in the *Performance* row, double-click the cell under the column **Values** and add the following values to different labels:

Value	Label
3	Improved
4	Unchanged
5	Deteriorated

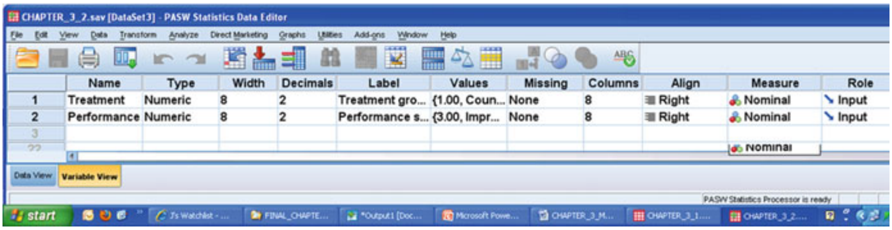
   There is no specific rule of defining the values of labels. Even "Improved," "Unchanged," and "Deteriorated" may be defined as 1, 2, and 3, respectively.
6. Under the column label **Measure,** select the option "Nominal" for both the variables *Treatment* and *Performance*.
7. Use default entries in rest of the columns. .
   After defining the variables in **Variable View,** the screen shall look like Fig. 3.7.

(iii) *Entering data*: Once the variables *Treatment* and *Performance* have been defined in the **Variable View**, click **Data View** on the left bottom of the screen to open the format for entering the data column wise.

In this example, there are two variables *Treatment* and *Performance*. *Treatment* has two value levels: "1" for counseling group and "2" for control group. Since there are 40 students in counseling group and 40 in control group in this example, under the *Treatment* column, write first 40 data as 1 and next 40 data as 2.

Since out of 40 students of counseling group, 22 showed "improved" (value = 3), 8 showed "unchanged" (value = 4), and 10 showed "deteriorated" (value = 5) performance, under the *Performance* column, type first 22 data as 3, next 8 data as 4, and subsequent 10 data as 5.

Similarly out of 40 students of control group, 4 showed "improved" (value = 3), 5 showed "unchanged" (value = 4), and 31 showed "deteriorated" (value = 5) performance; therefore, after typing the above data under the *Performance* column, type next 4 data as 3, 5 data as 4, and 31 data as 5.


CHAPTER_3_2.sav [DataSet3] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Data View
Variable View
PASW Statistics Processor is ready
start
J's Watchlist - ...
FINAL_CHAPTE...
*Output1 [Doc...
Microsoft Powe...
CHAPTER_3_M...
CHAPTER_3_1....
CHAPTER_3_2....

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Treatment	Numeric	8	2	Treatment gro...	{1.00, Coun...	None	8	Right	Nominal	Input
2	Performance	Numeric	8	2	Performance s...	{3.00, Impr...	None	8	Right	Nominal	Input
3											
22										Nominal	

Fig. 3.7 Defining variables along with their characteristics

Data feeding procedure for the data of Table 3.11 in SPSS under **Data View**

S.N.		Treatment	Performance	
1		1	3	
2		1	3	
3		1	3	Type “3” twenty-two times as there are 22 students showed Improved performance in counseling group
4		1	3	
5		1	3	
6		1	3	
7		1	3	
8		1	3	
9		1	3	
10		1	3	
11		1	3	
12	Type “1” forty	1	3	
13		1	3	
14	times as there are 40	1	3	
15	students in the counseling group	1	3	
16		1	3	
17		1	3	
18		1	3	
19		1	3	
20		1	3	
21		1	3	
22		1	3	
23		1	4	
24		1	4	Type “4” eight times as there are 8 students showed Unchanged performance in counseling group
25		1	4	
26		1	4	
27		1	4	
28		1	4	
29		1	4	
30		1	4	
31		1	5	
32		1	5	Type “5” ten times as there are 10 students showed Deteriorated performance in counseling group
33		1	5	
34		1	5	
35		1	5	

(continued)

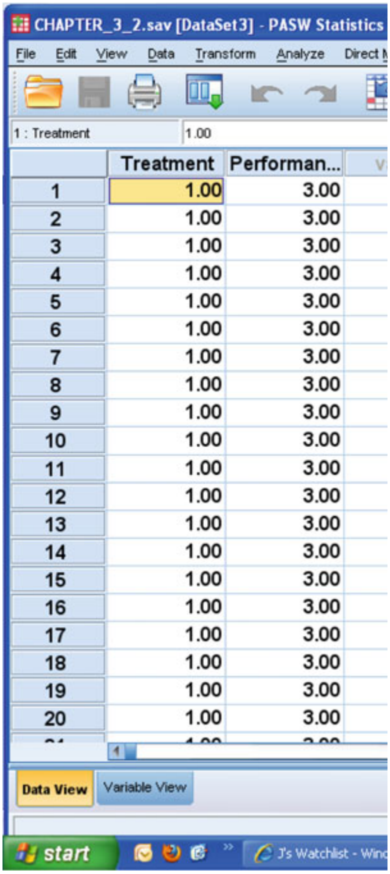
(continued)

S.N.		Treatment	Performance	
36		1	5	
37		1	5	
38		1	5	
39		1	5	
40		1	5	
41		2	3	Type "3" four times as there are 4 students showed Improved performance in control group
42		2	3	
43		2	3	
44		2	3	
45		2	4	
46		2	4	Type "4" five times as there are 5 students showed Unchanged performance in control group
47	Type "2" forty times as there are 40 students in the control group	2	4	
48		2	4	
49		2	4	
50		2	5	
51		2	5	
52		2	5	Type "5" thirty-one times as there are 31 students showed Deteriorated performance in control group
53		2	5	
54		2	5	
55		2	5	
56		2	5	
57		2	5	
58		2	5	
59		2	5	
60		2	5	
61		2	5	
62		2	5	
63		2	5	
64		2	5	
65		2	5	
66		2	5	
67		2	5	
68		2	5	
69		2	5	
70		2	5	
71		2	5	
72		2	5	
73		2	5	
74		2	5	
75		2	5	
76		2	5	
77		2	5	
78		2	5	
79		2	5	
80		2	5	

Treatment coding: 1 = Counseling group, 2 = Control group

Performance coding: 3 = Improved, 4 = Unchanged, 5 = Deteriorated

Fig. 3.8 Screen showing entered data for the *Treatment* and *Performance* variables in the data view


CHAPTER_3_2.sav [DataSet3] - PASW Statistics
File Edit View Data Transform Analyze Direct M
1 : Treatment
1.00
Data View
Variable View
start
J's Watchlist - Winc

	Treatment	Performan...	V
1	1.00	3.00	
2	1.00	3.00	
3	1.00	3.00	
4	1.00	3.00	
5	1.00	3.00	
6	1.00	3.00	
7	1.00	3.00	
8	1.00	3.00	
9	1.00	3.00	
10	1.00	3.00	
11	1.00	3.00	
12	1.00	3.00	
13	1.00	3.00	
14	1.00	3.00	
15	1.00	3.00	
16	1.00	3.00	
17	1.00	3.00	
18	1.00	3.00	
19	1.00	3.00	
20	1.00	3.00	

After entering the data, the screen will look like Fig. 3.8. The screen shows only the partial data as the data is entered column wise which takes two-page-long entries. Save the data file in the desired location before further processing.

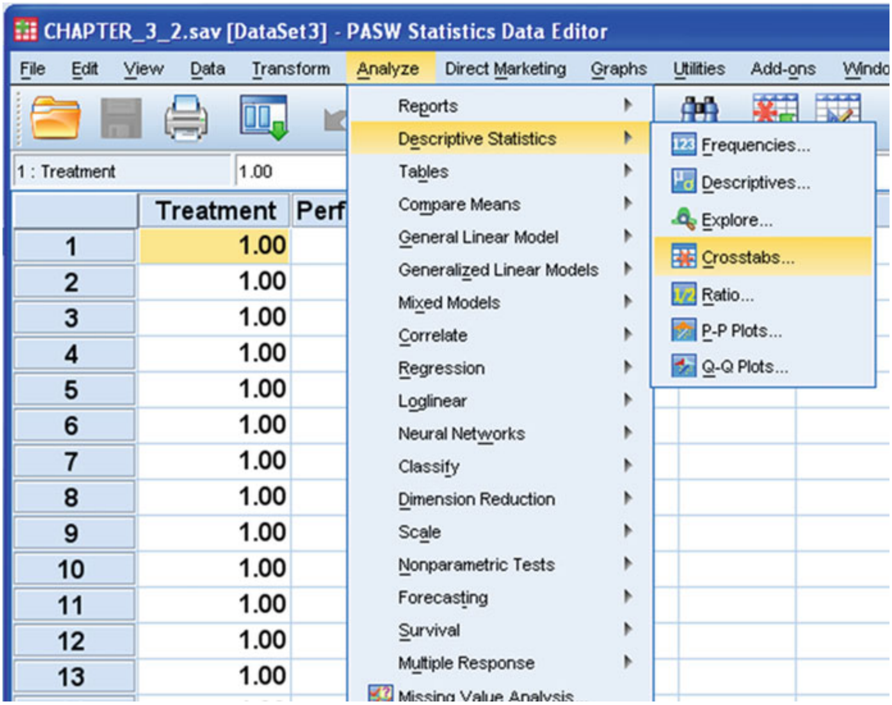
(b) **SPSS *Commands for Computing Chi-square with Two Variables***
After entering all the data by clicking the data view, take the following steps for computing chi-square:

(i) *Initiating the SPSS commands for computing chi-square:* In Data View, click the following commands in sequence:

Analyze → Descriptive Statistics → Crosstabs

The screen shall look like Fig. 3.9.

(ii) *Selecting variables for computing chi-square*: After clicking the "Crosstabs" option, you will be taken to the next screen for selecting variables for the crosstabs analysis and computing chi-square. Out of the two variables, one has to be selected in the Row(s) panel and the other in the Column(s) panel.



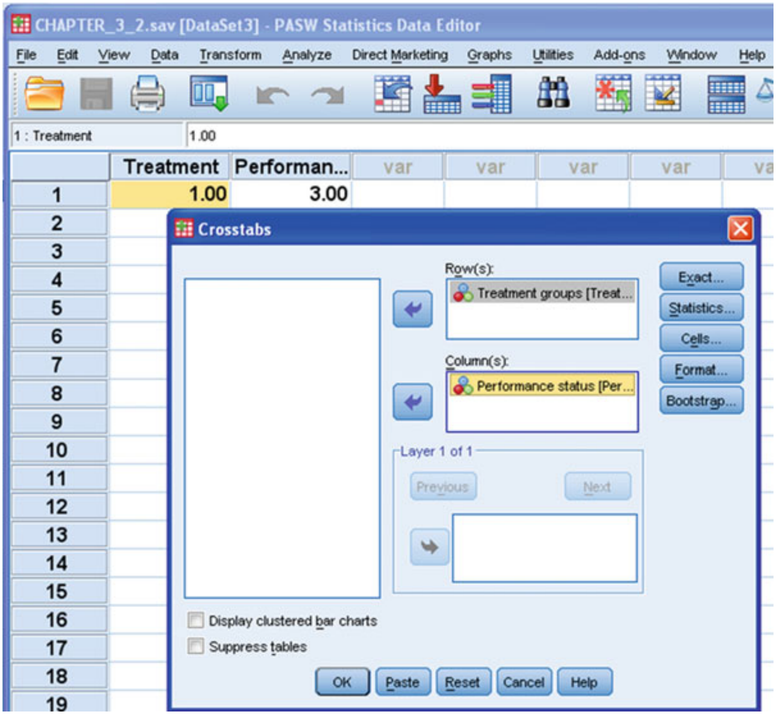
	Treatment	Perf
1	1.00	
2	1.00	
3	1.00	
4	1.00	
5	1.00	
6	1.00	
7	1.00	
8	1.00	
9	1.00	
10	1.00	
11	1.00	
12	1.00	
13	1.00	

Fig. 3.9 Screen showing the SPSS commands for computing chi-square in crosstabs

Select the variables *Treatment group* and *Performance status* from the left panel and bring them to the "Row(s)" and "Column(s)" sections of the right panel, respectively, by arrow button. The screen shall look like Fig. 3.10.

(iii) *Selecting option for computation:* After selecting variables, option needs to be defined for the crosstabs analysis and computation of chi-square. Take the following steps:

- Click **Statistics** option to get the screen shown in Fig. 3.11.
  - Check the options "Chi-square" and "Contingency coefficient."
  - Click ***Continue***.
- Click **Cells option to** get the screen shown in Fig. 3.12. Then,
  - Check the options "Observed" and "Expected" under the Counts section. Observed is checked by default.
  - Click ***Continue***. You will be taken back to the screen shown in Fig. 3.10.
  - Use default entries in other options. Readers are advised to try other options and see what changes they are getting.
  - Click **OK.**


CHAPTER_3_2.sav [DataSet3] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Crosstabs
Exact... Statistics... Cells... Format... Bootstrap...
Layer 1 of 1
Previous Next
OK Paste Reset Cancel Help

1 : Treatment 1.00

	Treatment	Performan...	var	var	var	var
1	1.00	3.00				

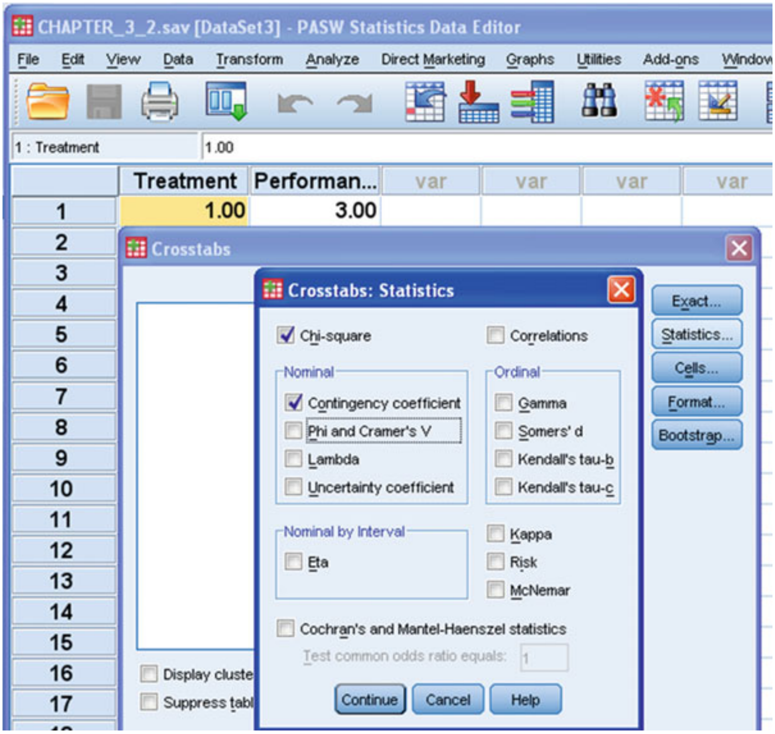
Row(s): Treatment groups [Treat...

Column(s): Performance status [Per...

□ Display clustered bar charts

□ Suppress tables

Figure 3.10 Screen showing selection of variables for chi-square in crosstab


CHAPTER_3_2.sav [DataSet3] - PASW Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window
Crosstabs
Crosstabs: Statistics
Exact... Statistics... Cells... Format... Bootstrap...
Continue Cancel Help

1 : Treatment 1.00

	Treatment	Performan...	var	var	var	var
1	1.00	3.00				

☑ Chi-square □ Correlations

Nominal
☑ Contingency coefficient
□ Phi and Cramer's V
□ Lambda
□ Uncertainty coefficient

Ordinal
□ Gamma
□ Somers' d
□ Kendall's tau-b
□ Kendall's tau-c

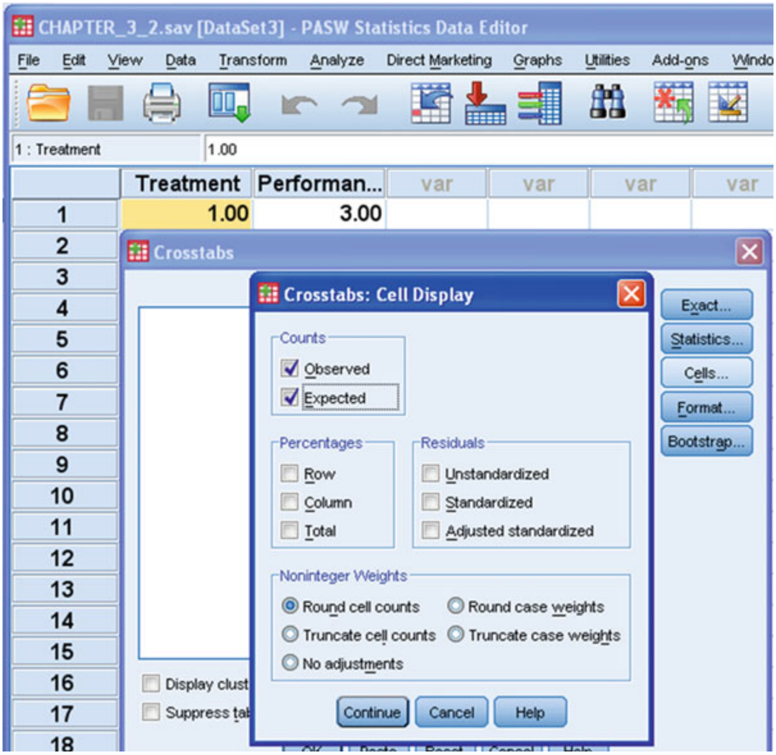
Nominal by Interval
□ Eta

□ Kappa
□ Risk
□ McNemar

□ Cochran's and Mantel-Haenszel statistics
Test common odds ratio equals: 1

□ Display cluste
□ Suppress tabl

Fig. 3.11 Screen showing option for computing chi-square and contingency coefficient



CHAPTER_3_2.sav [DataSet3] - PASW Statistics Data Editor

1 : Treatment 1.00

	Treatment	Performan...	var	var	var	var
1	1.00	3.00				

Crosstabs: Cell Display

Counts
☑ Observed
☑ Expected

Percentages
□ Row
□ Column
□ Total

Residuals
□ Unstandardized
□ Standardized
□ Adjusted standardized

Noninteger Weights
● Round cell counts ○ Round case weights
○ Truncate cell counts ○ Truncate case weights
○ No adjustments

□ Display clust
□ Suppress tal

Fig. 3.12 Screen showing option for computing observed and expected frequencies

Table 3.12 Treatment groups × Performance status cross tabulation

			Performance status			
			Improve	Unchanged	Deteriorated	Total
Treatment groups	Counseling Gp	Count	22	8	10	40
		Expected count	13.0	6.5	20.5	40.0
	Control Gp	Count	4	5	31	40
		Expected count	13.0	6.5	20.5	40.0
Total		Count	26	13	41	80
		Expected count	26.0	13.0	41.0	80.0

(c) ***Getting the Output***

Clicking option **OK** will lead you to the output window. The output panel will have lots of results. It is up to the researcher to decide the relevant outputs to be shown in their report. The relevant output can be selected by using the right click of the mouse and copying in the word file. In this example, the output so generated by the SPSS will look like as shown in Tables 3.12, 3.13, and 3.14.

Table 3.13 Chi-square for the data on Treatment × Performance

	Value	df	Asymp. sig. (2-sided)
Pearson chi-square	23.910[a]	2	.000
Likelihood ratio	25.702	2	.000
Linear-by-linear association	23.400	1	.000
N of valid cases	80		

[a] 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.50

Table 3.14 Contingency coefficient for the data on Treatment × Performance

		Value	Approx. sig. (*p* value)
Nominal by nominal	Contingency coefficient	0.480	0.000
N of valid cases		80	

Interpretation of the Outputs

The observed and expected frequencies of the *Treatment group* × *Performance status* can be seen in Table 3.12. Since no cell frequency is less than 5, therefore, no correction is required while computing the chi-square. If any of the cell frequency had value 5 or less, then SPSS would have computed the chi-square after applying the correction.

Table 3.13 shows the value of chi-square (χ^2) as 23.910, which is significant at 1% level as the p value is .000. Thus, we may reject the null hypothesis that "There is no association between *Treatment* and *Performance*." Hence, it may be concluded that there is a significant association between treatment and performance. In other words, it can be said that the pattern of academic performance is different in counseling and control group. Since the number of improved performance cases (22) is higher in counseling group than that of 4 in control group, it may be interpreted that the academic counseling is effective in improving the performance.

In Table 3.14, the value of contingency coefficient is 0.480. This is a measure of association between *Treatment* and *Performance*. This contingency coefficient can be considered to be significant as its p value is .000 which is less than .05. Thus, it may finally be concluded that counseling is significantly effective in improving the academic performance.

Summary of the SPSS Commands

(a) **For computing chi-square statistic (for testing equal occurrence hypothesis):**

1. Start SPSS by using the following sequence of commands:

Start → All Programs → IBM SPSS Statistics → IBM SPSS Statistics 20

2. Create data file by choosing the option **Type in Data**.
3. Click the tag **Variable View** and define the variable *Job_Sat* as "Nominal" variable.
4. For the variable *Job_Sat*, under the column heading **Values,** define "1" for low, "2" for medium, and "3" for high.
5. By clicking the **Data View,** enter first forty data of the variable *Job_Sat* as 1, next thirty as 2, and further 20 as 3 in the same column.
6. Click the following command sequence for computing chi-square:

Analyze → Nonparametric Tests → Legacy Dialogs → Chi – Square

7. Select the variable *Job_Sat* from left panel to the right panel.
8. Click the tag **Options** and check the box of "Descriptive." Press ***Continue***.
9. Click **OK** to get the output.

(b) **For computing chi-square statistic (for testing the significance of association between two attributes):**

1. Start SPSS by using the following command sequence:

Start → All Programs → IBM SPSS Statistics → IBM SPSS Statistics 20

2. Click **Variable View** tag and define the variable *Treatment* and *Performance* as "Nominal" variables.
3. In the *Treatment* row, double-click the cell under the column **Values** and add the values "1" for Counseling group and "2" for Control group. Similarly, in the *Performance* row*,* define the value "3" for Improved, 4 for Unchanged, and 5 for Deteriorated.
4. Use default entries in rest of the columns.
5. Click **Data View** tag and feed first forty entries as 1 and next forty entries as 2 for the *Treatment* variable.
6. Similarly for the *Performance* variable, enter first twenty-two entries as 3, next eight entries as 4, and further ten entries as 5. These three sets of entries are for counseling group. Similarly for showing the entries of control group, enter first four entries as 3, next five entries as 4, and after that thirty-one entries as 5 in the same column.
7. Click the following command sequence for computing chi-square:

Analyze → Descriptive Statistics → Crosstabs

8. Select variables *Treatment group* and *Performance status* from the left panel to the "Row(s)" and "Column(s)" sections of the right panel, respectively.
9. Click the option **Statistics** and check the options "Chi-square" and "Contingency coefficient." Press ***Continue***.
10. Click **OK** to get the output.

Exercise

Short-Answer Questions

Note: Write answer to each of the questions in not more than 200 words:

Q.1. Responses were obtained from male and female on different questions related to their knowledge about smoking. There were three possible responses Agree, Undecided, and Disagree for each of the questions. How will you compare the knowledge of male and female about smoking?

Q.2. Write in brief two important applications of chi-square.

Q.3. How will you frame a null hypothesis in testing the significance of an association between gender and IQ where IQ is classified into high and low category? Write the decision criteria in testing the hypothesis.

Q.4. Can the chi-square be used for comparing the attitude of male and female on the issue of "Foreign retail chain may be allowed in India" if the frequencies are given in 3 × 5 table below? If so or otherwise, interpret your findings. Under what situation chi-square is the most robust test?

Response on "Foreign retail chain may be allowed in India"

		Strongly agree	Agree	Undecided	Disagree	Strongly disagree
Gender	Male	50	20	15	5	10
	Female	20	15	10	25	30

Q.5 If chi-square is significant, it indicates that the association between the two attributes exists. How would you find the magnitude of an association?

Q.6 What is phi coefficient? In what situation it is used? Explain by means of an example.

Multiple-Choice Questions

Note: For each of the question, there are four alternative answers. Tick mark the one that you consider the closest to the correct answer.

1. For testing the significance of association between Gender and IQ level, the command sequence for computing chi-square in SPSS is

   (a) Analyze -> Nonparametric Tests -> Chi-square
   (b) Analyze -> Descriptive Statistics -> Crosstabs
   (c) Analyze -> Chi-square -> Nonparametric Tests
   (d) Analyze -> Crosstabs -> Chi-square

2. Choose the most appropriate statement about the null hypothesis in chi-square.

   (a) There is an association between gender and response.
   (b) There is no association between gender and response.
   (c) There are 50−50% chances of significant and insignificant association.
   (d) None of the above is correct.

3. Response of the students on their preferences toward optional papers is as follows:

	Response of the students		
Subjects	Finance	Human resource	Marketing
No. of students	15	25	20

The value of chi-square shall be

(a) 2
(b) 2.5
(c) 50
(d) 25

4. The value of chi-square for the given data shall be

		Gender	
		Male	Female
Region	North	30	20
	South	10	40

(a) 16.67
(b) 166.7
(c) 1.667
(d) 1667

5. Chi-square is used for

(a) Finding magnitude of an association between two attributes
(b) Finding significance of an association between two attributes
(c) Comparing the variation between two attributes
(d) Comparing median of two attributes

6. Chi-square is the most robust test if the frequency table is

(a) 2×2
(b) 2×3
(c) 3×3
(d) $m \times n$

7. While using chi-square for testing an association between the attributes, SPSS provides Crosstabs option. Choose the most appropriate statement.

(a) Crosstabs treats all data as nominal.
(b) Crosstabs treats all data as ordinal.
(c) Crosstabs treats some data as nominal and some data as ordinal.
(d) Crosstabs treats data as per the problem.

8. If responses are obtained in the form of the frequency on a 5-point scale and it is required to compare the responses of male and female on the issue “Marketing stream is good for the female students,” which statistical test you would prefer?

   (a) Two-sample t-test
   (b) Paired t-test
   (c) One-way ANOVA
   (d) Chi-square test

9. If p value for a chi-square is .02, what conclusion you can draw?

   (a) Chi-square is significant at 95% confidence.
   (b) Chi-square is not significant at 95% confidence.
   (c) Chi-square is significant at .01 levels.
   (d) Chi-square is not significant at .05 levels.

10. The degree of freedom of chi-square in a $r \times c$ table is

    (a) $r + c$
    (b) $r + c-1$
    (c) rc
    (d) (r−1)(c−1)

11. Phi coefficient is used if

    (a) Both the variables are ordinal.
    (b) Both the variables are binary.
    (c) Both the variables are interval.
    (d) One of the variables is nominal and the other is ordinal.

12. Gamma coefficient is used if

    (a) Both the variables are interval.
    (b) Both the variables are binary.
    (c) Both the variables are ordinal.
    (d) Both the variables may be on any scale.

Assignments

1. Following are the frequencies of students in an institute belonging to Low, Medium, and High IQ groups. Can it be concluded that there is a specific trend of IQ’s among the students. Test your hypothesis at 5% level.

Frequencies of the students in different IQ groups

IQ categories	Low IQ	Medium IQ	High IQ
Frequency	20	65	35

2. In an organization following are the frequencies of male and female workers in the skilled and unskilled categories. Test whether nature of work is independent of the gender by computing chi-square. Also compute contingency coefficient

along with the expected frequency and percentage frequencies in the Crosstabs and interpret your findings. Test your hypothesis at 5% level.

Frequency of workers in different categories

		Workers	
		Skilled	Unskilled
Gender	Male	50	15
	Female	15	40

Answers to Multiple-Choice Questions

Q.1	b	Q.2	b
Q.3	b	Q.4	a
Q.5	b	Q.6	a
Q.7	a	Q.8	d
Q.9	a	Q.10	d
Q.11	b	Q.12	c