14.6: Using SPSS

Let's return to our example about Statistonia College from the beginning of our chapter. For that example, we supposed that among the population of college students, 25% were business majors, 25% were psychology majors, 20% were nursing majors, and 30% were statistics majors. Suppose that you wanted to test whether the proportion of students in each of these majors at a small college in Statistonia were similar to that of college students, overall, in SPSS. In order to test a hypothesis using SPSS, we need the raw data rather than the summarized counts like we saw in Data Set 14.1. Thus, we will use Data Set 14.3 to test the current hypothesis using SPSS. Data Set 14.3 shows the qualitative data for majors for a sample of 30 in the left column. To help us summarize the hypothesized counts, the expected proportions for each category are shown in the right column but are *not* data.

Data Set 14.3. Majors at Statistonia College (n = 30)

Majors Data	Expected Proportions
Business	25%
Business	
Business	
Business	
Business	
Business	
Psychology	
Psychology	
Psychology	
Psychology	
Psychology	
Psychology	
Psychology	
Psychology	
Psychology	
Psychology	
Nursing	20%
Nursing	
Nursing	
Nursing	
Nursing	
Nursing	
Nursing	
Statistics	30%
Statistics	
Statistics	
Statistics	
Statistics	

Majors Data	Expected Proportions
Statistics	

Data need to be organized and entered into SPSS in ways that serve the analysis to be conducted. Thus, this section focuses on how to enter and analyze data for a chi-squared goodness of fit using SPSS. SPSS version 29 was used for this book; if you are using a different version, you may see some variation from what is shown here.

Entering Data

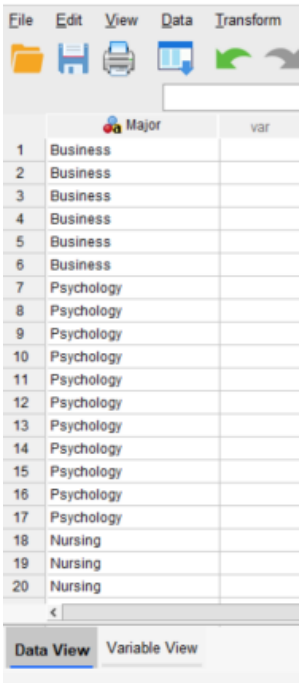
Open the SPSS software, click “New Dataset,” then click “Open” (or “OK” depending on which is shown in the version of the software you are using). This will create a new blank spreadsheet into which you can enter data. Click on the Variable View tab on the bottom of the spreadsheet. This tab of the spreadsheet has several columns to organize information about the variables. The first column is titled “Name.” Start here and follow these steps:

1. Click the first cell of that column and enter the name of your test variable using no spaces, special characters, or symbols. For Data Set 14.3 the variable name is Major. Hit enter and SPSS will automatically fill in the other cells of that row with some default assumptions about the data.
2. Click the first cell of the column titled “Type” and then click the three dots that appear in the right side of the cell. Ideally, we would select “string” which allows us to use words to name each group. However, SPSS works best if we recode each group name as number (i.e. 1 = Business, 2 = Psychology, 3 = Nursing, and 4 = Statistics). We will use this coding to enter our data. For type, specify that the data for that variable will appear as numbers by selecting “Numeric.” We will need to use numbers to represent each major.
3. Click the first cell of the column titled “Label.” This is where you can specify what you want the variable to be called in output, including in tables and graphs. You can use spaces or phrases here, as desired.
4. Click on the first cell of the column titled “Measure.” A pulldown menu with three options will allow you to specify the scale of measurement for the variable. Select the “Nominal” option because the variables for a chi-squared are qualitative.

Now you are ready to enter your data. Click on the Data View tab toward the bottom of the spreadsheet. This tab of the spreadsheet has columns into which you can enter the data. Click the cell of the first column and follow these steps:

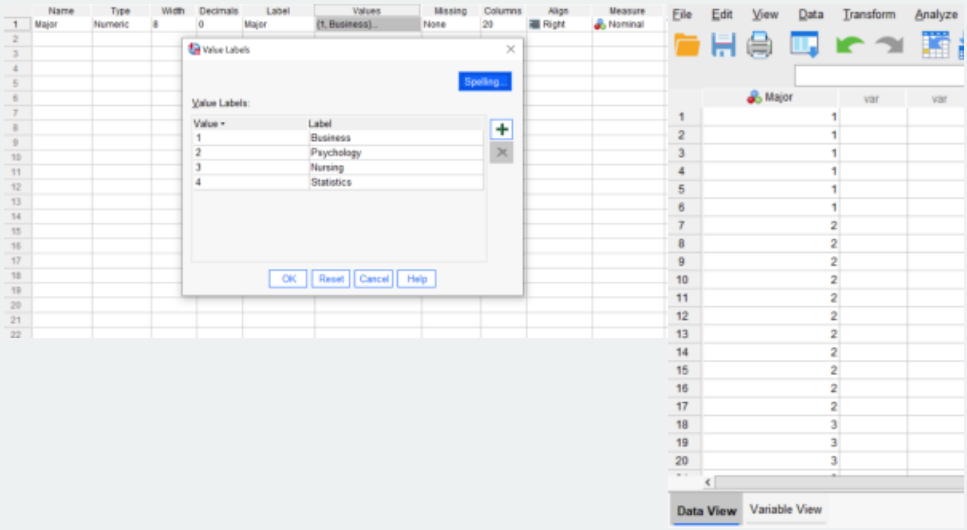
1. Enter the data for the test variable moving down the rows under the first column. If your data are already on your computer in a spreadsheet format such as excel, you can copy-paste the data in for the variable.
2. Then hit save to ensure your data set will be available for you in the future.

Here is how the first 20 data points for Data Set 14.3 looks after data are entered into SPSS using the group names:



Note

SPSS Version 29 is fussy and only seems to work if you code the names with numeric values. Thus, if you get an error warning when trying to run your analysis, recode your majors as 1, 2, 3, and 4 in the data view tab and clarify what they stand for using the values cell of the variables tab as shown below:

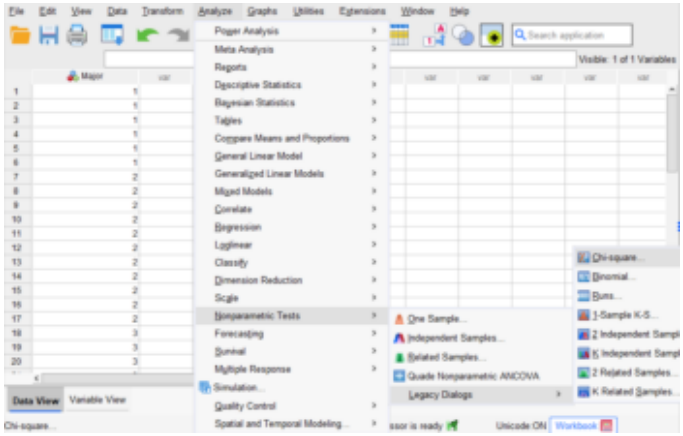


Once all the variables have been specified and the data have been entered, you can begin analyzing the data using SPSS.

Conducting a Chi-Squared Goodness of Fit Test in SPSS

The steps to running this form of chi-squared in SPSS are:

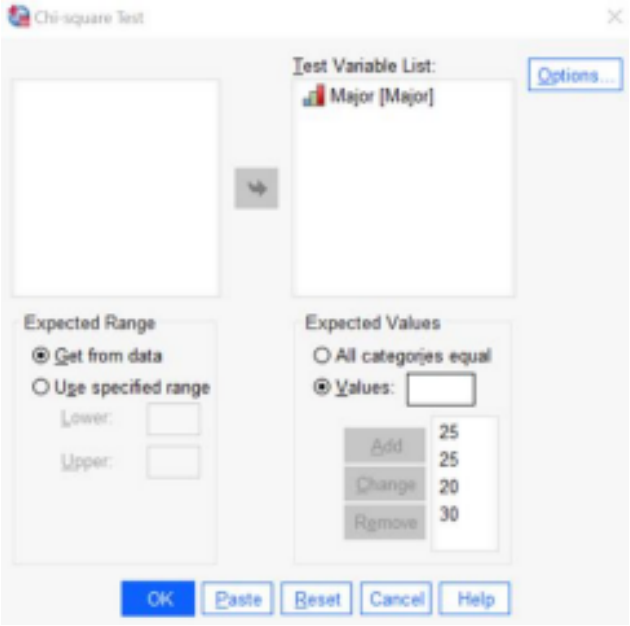
1. Click Analyze -> Nonparametric Tests -> Legacy Dialogs -> Chi-squared from the pull down menus.



2. Drag the name of the test variable from the list on the left into the box on the right of the command window. For Data Set 14.3, this means we are moving Major to the test box as shown below.

> **Note**
>
> When using SPSS Version 29, it may call your nominal variable ordinal at this step. This will not impact the analysis.

In this command window, we must set our expected values. We are testing whether the count of students who identified their majors as Business, Psychology, Nursing, and Statistics are similar to dissimilar to the proportions of 25%, 25%, 20%, and 30%, respectively. Thus, we need to enter those values as shown in the command window below:



3. Click “OK” to run the analyses.
4. The output (which means the page of calculated results) will appear in a new window of SPSS known as an output viewer. The results will appear in two tables as shown below:

	Major		
	Observed N	Expected N	Residual
Business	6	7.5	-1.5
Psychology	11	7.5	3.5

	Major		
	Observed N	Expected N	Residual
Nursing	7	6.0	1.0
Statistics	6	9.0	-3.0
Total	30		

Test Statistics

	Major
Chi-Square	3.100[a]
df	3
Asymp. Sig.	.376

a. 0 cells (0.0%) have expected frequencies less than 5.
The minimum expected cell frequency is 6.0.

The first table summarizes the observed and expected counts for each category (i.e. for each major in this example). Notice that the expected percents have been transformed into their proportional expected counts. The second table provides the results of the inferential test using the observed and expected counts from the first table. The chi-squared formula result appears on the top row; the result is χ^2 = 3.10 when rounded to the hundredths place. The degrees of freedom appear in the second row. There were 4 categories whose counts were being compared and, thus, the df = 3. Finally, the p-value (which represents the risk of a Type I Error) appears in the third row as the "Asymp. Sig.". This is the same as what SPSS has called "sig." in our earlier chapters. When the p-value is less than .05, the result is significant and when it is greater than .05 the result is not significant. We see that the p-value for the present analysis was .376 and, thus, that the result is not significant.

Note

An assumption of a chi-squared test is that no category has an expected frequency lower than 5. SPSS automatically checks this assumption and alerts the user when it has been met or violated. We can see that SPSS has confirmed that this assumption was met for the present data in a note it has included under the second table of results.

The results from this test are as follows: A chi-squared goodness of fit test was used to test whether the counts of students in each major at Statistonia College were similar to those of students in the population, overall. Specifically, students overall majored in Business (25.0%), Psychology (25.0%), Nursing (20.0%), and Statistics (30.0%). The counts of students at Statistonia College who were majoring in Business (n = 6), Psychology (n = 11), Nursing (n = 7), and Statistics (n = 6) were not significantly different than the proportions in the population, $\chi^2(3, N = 30) = 3.10$, p >.05. Thus, the counts of each major at the college were proportionally similar to those in the population of students.

Reading Review 14.5

1. Where do you specify the expected proportions when setting up a chi-squared goodness of fit analysis is SPSS?
2. What information is provided in each of the two tables of SPSS output for chi-square?
3. Which assumption of chi-squared is automatically checked and reported by SPSS?
4. What does a non-significant result indicate when using a chi-squared goodness of fit test?