Chapter 5. Using SPSS to Analyze Frequency Data

Introduction to the Chi-square Statistic

Sometimes when we collect data, we are simply counting up how many times something happens; that is, we are collecting frequency data. Unlike our more precise measures such as hours studied, reaction time or number correct, we need to treat these data a bit differently. We cannot calculate descriptive statistics such as means and standard deviations with frequency data, but that does not imply there is no way in which to better understand the patterns in these data.

We have available to us the Chi-square statistic, and there are two forms of this test; the first is the test of "Goodness of Fit," and the second is the "Test of Independence." With each of these tests we can see if the data we collected, our observed frequencies, are in line with what we might have expected to see, our expected frequencies.

Think about it. There is roughly an equal number of males and females in the young adult population. Now, look around the room. Is there roughly an equal number of male and female students in your class? If not, how different is the count from what you should have expected to see? The Chi-square statistic allows us to obtain answers to these types of questions.

Data Set 5.1 – Chi-square Goodness of Fit Test

The owner of a family amusement center is trying to better understand the interests of the families who come to enjoy time at the fun center. He is hoping that if he understands the likes and interests of his patrons, he can improve his facility, and hopefully, improve business. To gain a better understanding of the activities families enjoy, he offers 75 families a choice of one, $5-off coupon for one of three activities: bowling, mini golf or an arcade card. (Assume all are equally priced.)

He has no reason to expect that there should be much of a difference in the number of families choosing each activity as all of these activities are generally very busy. He counts up the number of families taking each type of coupon. The data follow.

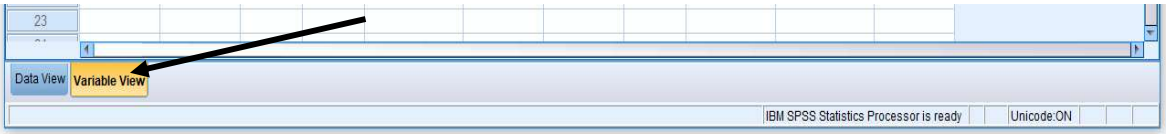
Activity	Number of Families Choosing the Activity
Mini Golf	35
Bowling	23
Arcade	17

Calculating Chi-square Goodness of Fit

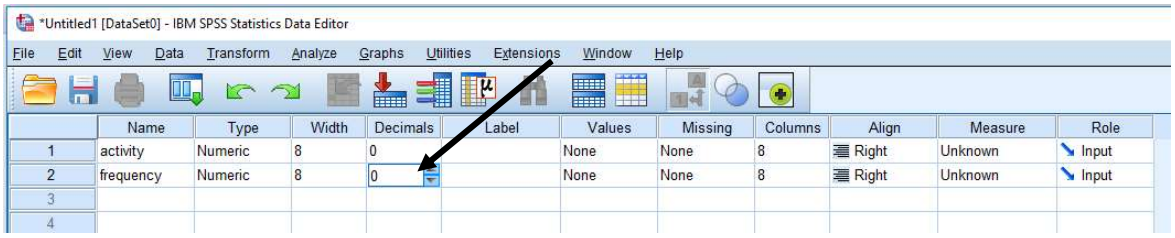
1. Open SPSS by double clicking the SPSS icon. 

   You can also find SPSS using the **"Start"** window button at the bottom, left of the screen, and choosing it from the menu.

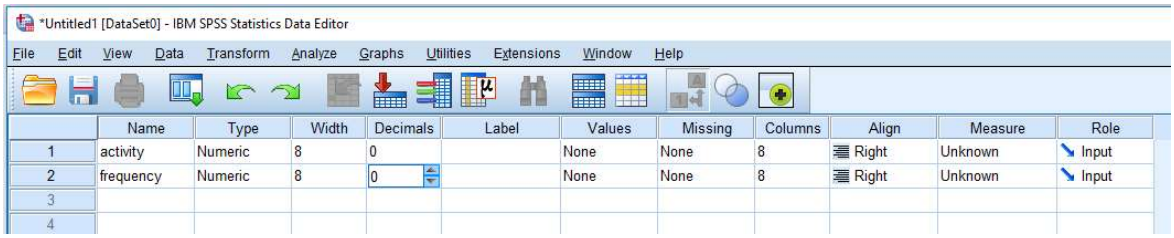
2. You will be brought to a new data screen. You will notice that there are two tabs at the bottom of the screen – data view and variable view. Select the **"Variable View"** tab (see below).



3. Now we need to code our variables. Enter the word **"activity"** in row 1. Then enter the word **"frequency"** in row 2. Once you do each of these, you will see that the row fills with additional information. Click on one of the boxes under the **"Decimals"** column, and change the number of decimal places to zero. (See arrow below.)



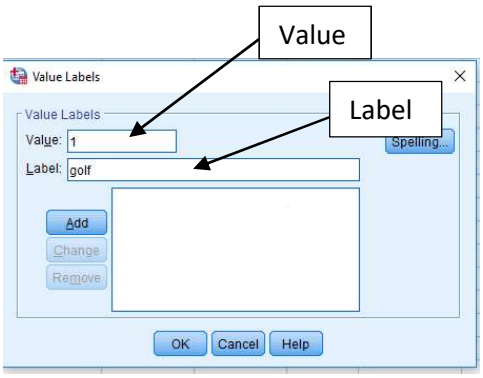
4. The next step is to "code" our variables. Click on the box in the **"Values"** column in the **"activity"** row. A gray button will appear with little dots inside. Click the button.

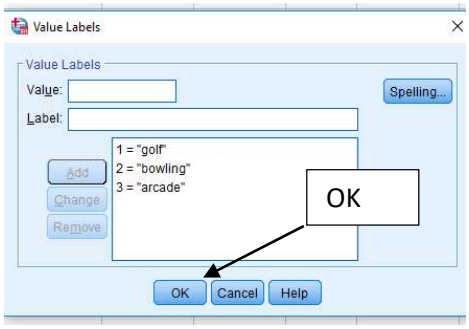


5. Once you click the button, a pop-up window will appear. This window is used for coding the levels of your variable, that is, the types of activities. Enter the number **"1"** for the value, and then type the word **"golf"** for the label. Next, click **"Add."** Repeat these steps to add the other two activities – **"bowling"** and **"arcade,"** giving them the values **2** and **3**.

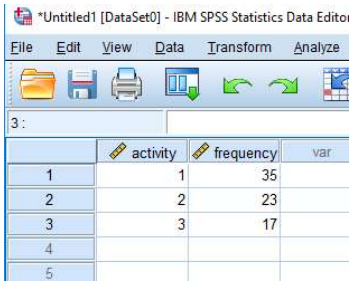
   Once you have entered all three values, click **"OK."**

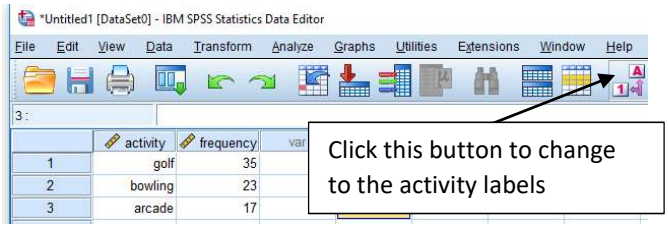
   The pop ups are shown on the next page





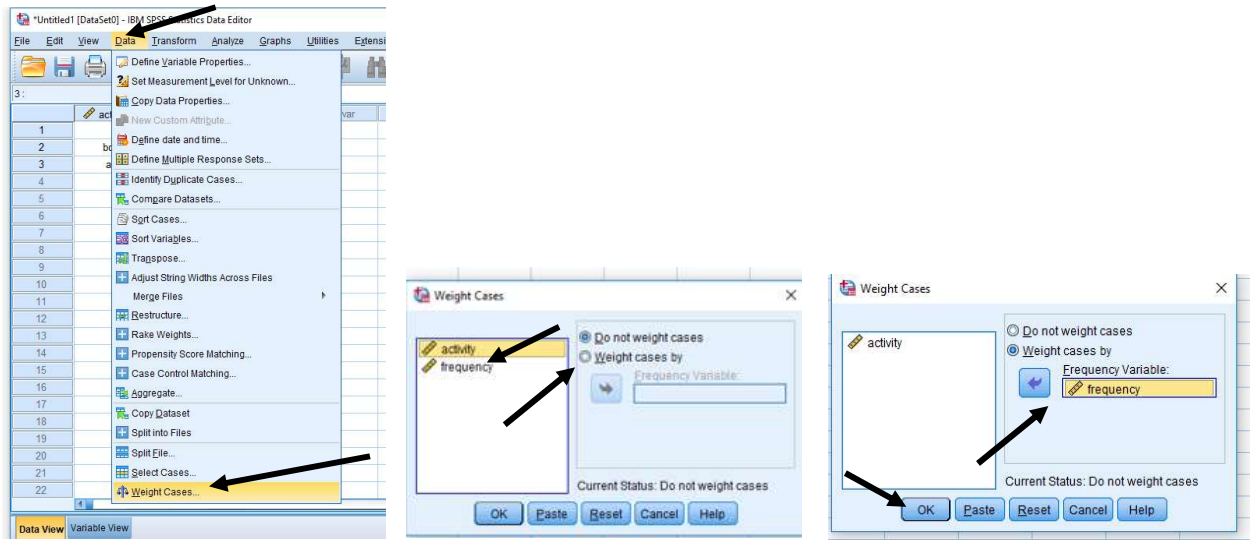
6. Now switch to the **"Data View"** tab. Here you will enter the frequencies. You will see that the **"Item"** column contains the values 1, 2 and 3. Enter the frequencies for each activity under the frequency column. We can also change our activity **"Values"** to the actual **"Labels"** by clicking the **"A→1"** button at the top of the screen. 



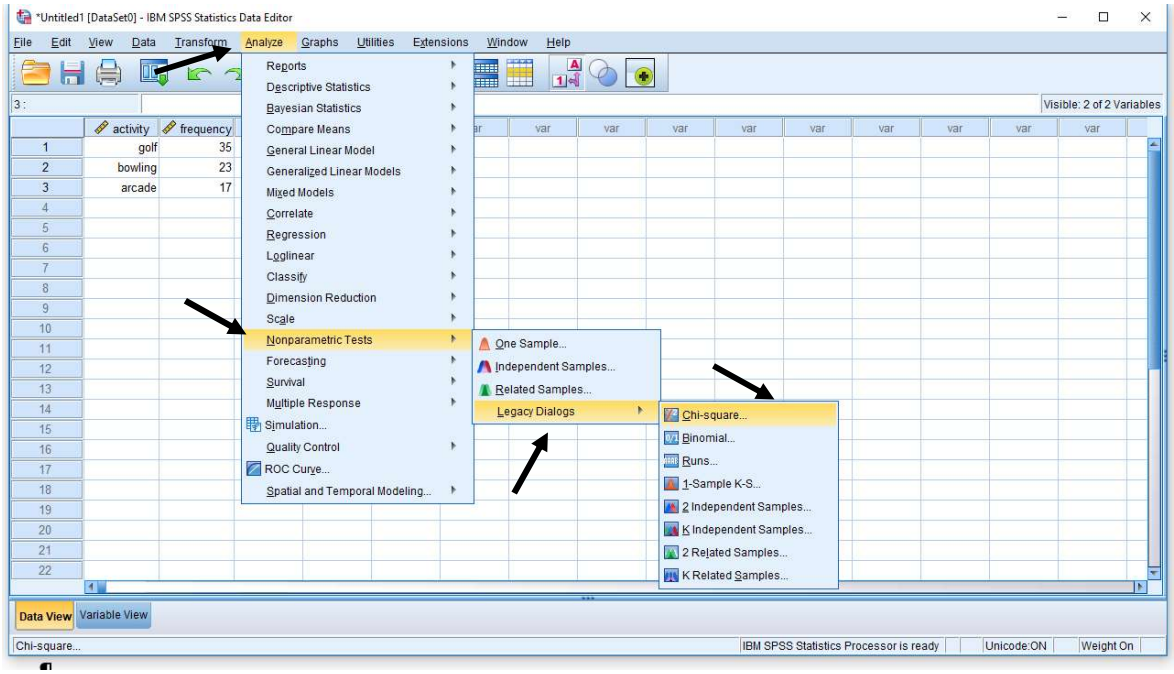


7. Now it is time to start analyzing the data. From the menu bar on the **"Data View"** page, choose **"Data."** A pop up will appear.

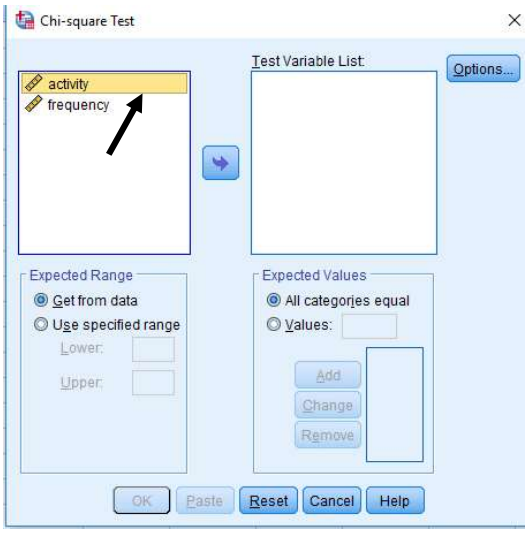
   Select **"Weight Cases,"** and then when the pop up appears, weight the cases by **"frequency"** by selecting **"frequency,"** and using the arrow to move it to the **"Frequency Variable"** box. Then click **"OK."**

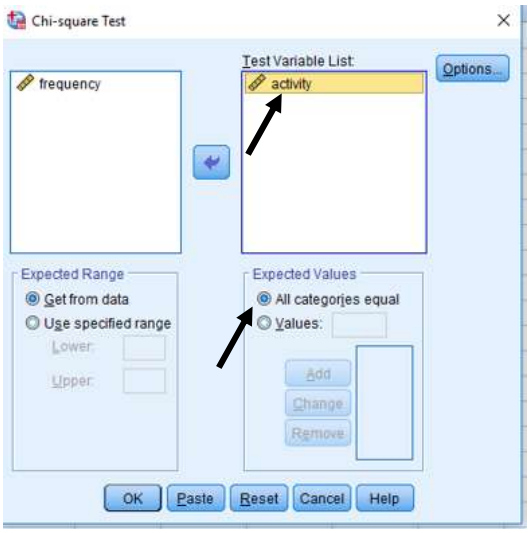


8. To analyze the data, stay on the **"Data View"** page, and select **"Nonparametric Tests,"** then **"Legacy Dialogs,"** followed by **"Chi-square."**



9. A pop up will appear. Select **"Activity,"** and move it to the **"Test Variable List"** by clicking the arrow button. Now look at the **"Expected Values"** box. We expect that all three activities should occur with equal frequency, so leave the **"Expected Values"** set at **"All categories equal."** Now click **"OK,"** and the results of the analysis will appear

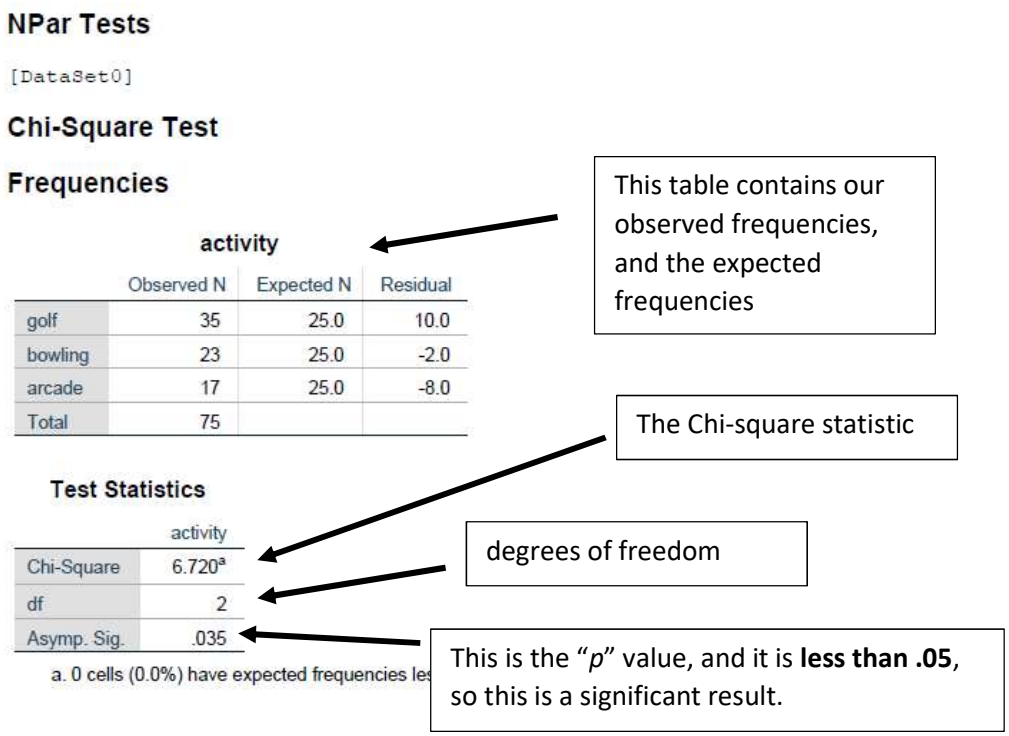




10. The results of the analysis appear in an SPSS output window.

The output appears on the next page

As you can see, the output provides a variety of information. The arrows indicate the key information needed to interpret the results of the Chi Square analysis.



NPar Tests

[DataSet0]

Chi-Square Test

Frequencies

activity

	Observed N	Expected N	Residual
golf	35	25.0	10.0
bowling	23	25.0	-2.0
arcade	17	25.0	-8.0
Total	75		

This table contains our observed frequencies, and the expected frequencies

Test Statistics

	activity
Chi-Square	6.720[a]
df	2
Asymp. Sig.	.035

a. 0 cells (0.0%) have expected frequencies les

The Chi-square statistic

degrees of freedom

This is the "*p*" value, and it is **less than .05**, so this is a significant result.

11. Once we have the value of the Chi-square statistic, and the *p* value, we can interpret the results.

    Our obtained Chi-square is significant. This means that the obtained frequencies are not in line with expectations. Some of the activities have higher frequencies than we expected, and others occur less often than expected.

12. Think about the scenario. The owner wondered if some activities are preferred over others. Our result provides evidence that the answer to this question is, "YES!"

13. We can now report our finding according to APA guidelines, the format we use in psychological research. We use **"χ^2"** to represent the Chi-square statistic. Our degrees of freedom (*df*) are placed in parentheses next to the Chi-square symbol. We report the statistic to two decimal places. A sample interpretation follows.

    The families showed significant preferences among the three activities at the family amusement center, $\chi^2(2) = 6.72$, $p = .035$.

Chi-square Test of Independence

The Chi-square Test of Independence allows us to examine how two variables impact our observed frequencies. If they have separate influence on the observed frequencies, then we say they work independently. In contrast, if they work together as a "combined force" to influence the observed frequencies, then their impact is not independent; that is, there is a relationship between the two variables.

Data Set 5.2 – Chi-square Test of Independence

An owner of a small, but rapidly growing accounting firm is thinking of moving to a new location. Her business is currently located in the heart of the city, but she knows that several of her employees live in the suburbs. She is wondering whether it is better to keep her business in the city or move it closer to the suburbs. She thinks it is best to poll her employees to see if they have a strong preference for the new location. She also thinks that married versus single employees might have different preferences.

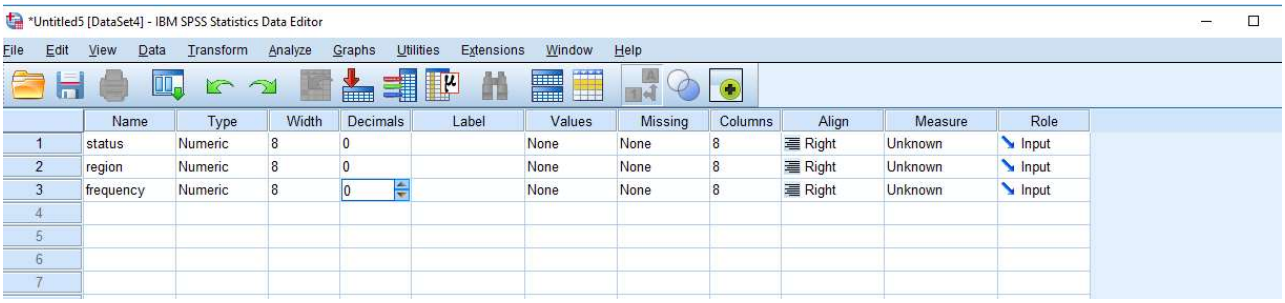
To better understand this situation, and to gain a better idea as to where she should relocate her business, she asks 50 single employees whether they would prefer to remain working in the city or work in the suburbs. She asks the same of 50 married employees. Their responses are presented in the table below.

		Region	
		City	**Suburbs**
Status	**Single**	**28**	**22**
	Married	**17**	**33**

Calculating Chi-square Test of Independence

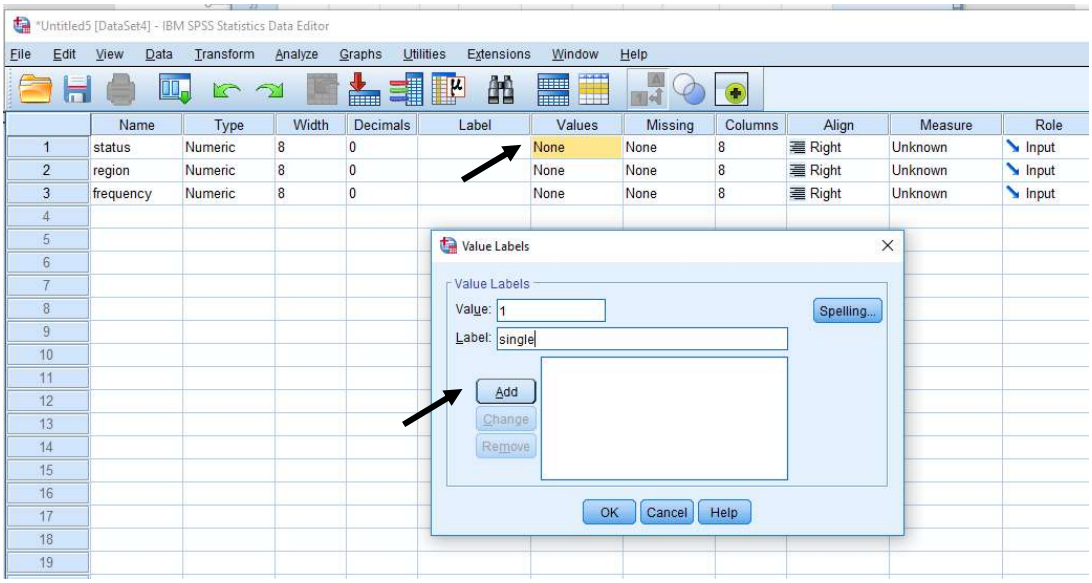
1. Open a new data screen by selecting **"File," "New"** and **"Data"** from the menu bar.

2. Switch to the **"Variables View"** tab, and label your variables. Under **"Name,"** type in **"status," "region"** and **"frequency."** (See below.)

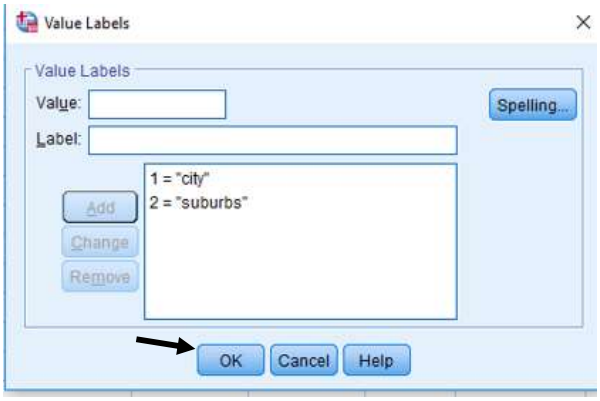


3. Once completed, the variables need to be coded. This process is very similar to the steps in the Goodness of Fit Test

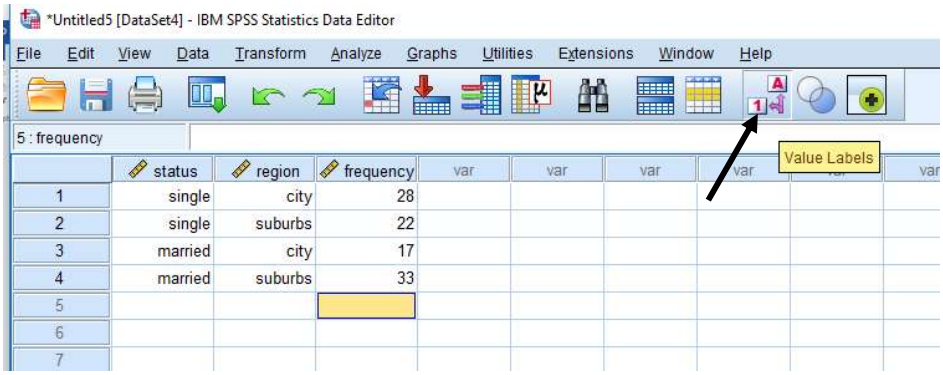
   Under the **"Values"** column, click on the cell in the status row. Click the gray button with the little dots inside. A pop up will appear. Type **"1"** in **"Value,"** and **"single"** in **"Label."** Then click add. Repeat this process using **"2"** for the **"Value,"** and **"married"** for the **"Label."** Click **"Add,"** and then **"OK"** to complete the coding for status.



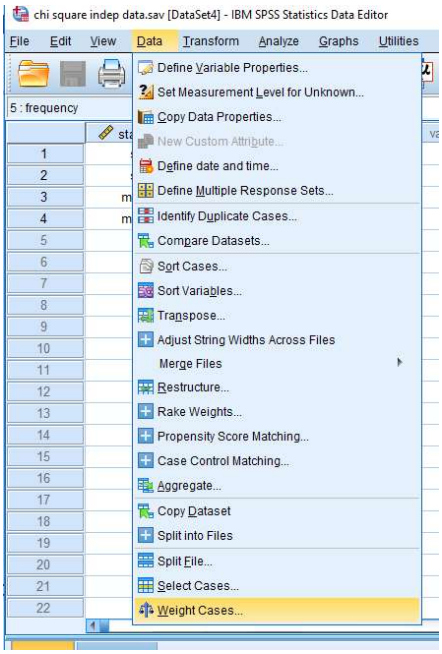
4. Now that status is coded, it is time to code **"region."** Repeat the steps you used above. Click on "region," and then the gray button with the dots. Once the pop up appears, use **"1"** for **"single,"** and **"2"** for **"married."** Once you have added both, click **"OK."**

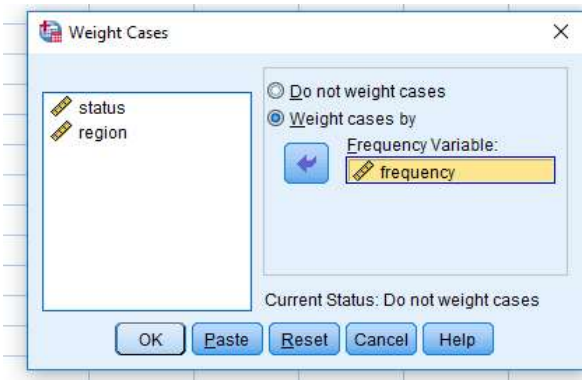


5. Now that we have our variables coded, we need to add the observed frequencies. Switch to the **"Data View"** tab, and add the observed frequencies. Once that is finished, click the **"A→1"** button  to change the numbers to the labels



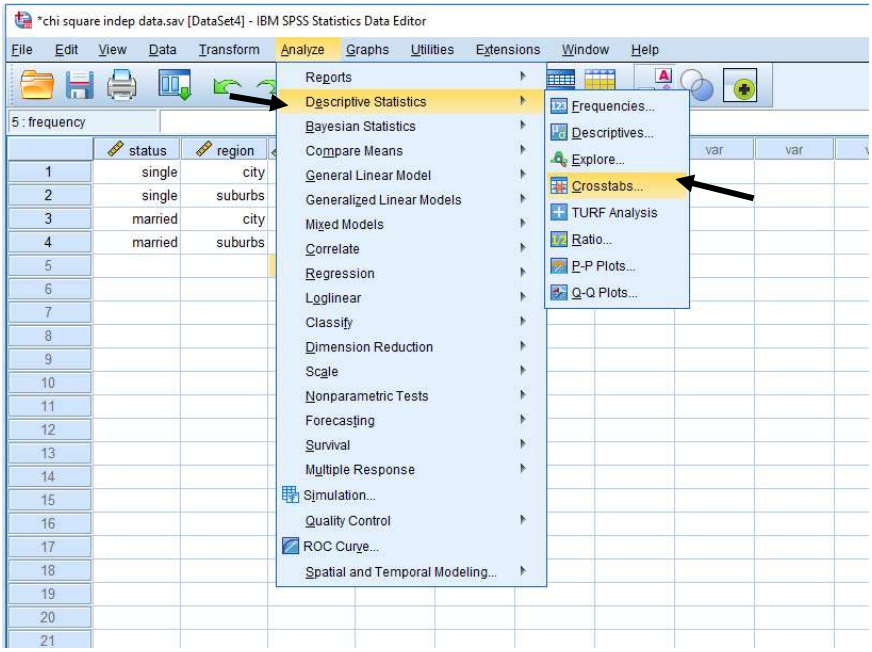
6. The next step, is to weight the cases. Select the **"Data"** tab at the top of the menu bar, and then select **"Weight Cases."** Once this is completed, a pop up will appear. Choose **"Weight cases by,"** and then select **"frequency"** and move it over using the arrow button. Next, click **"OK"**



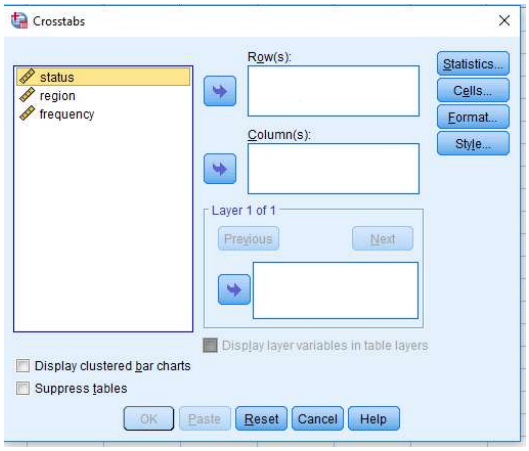


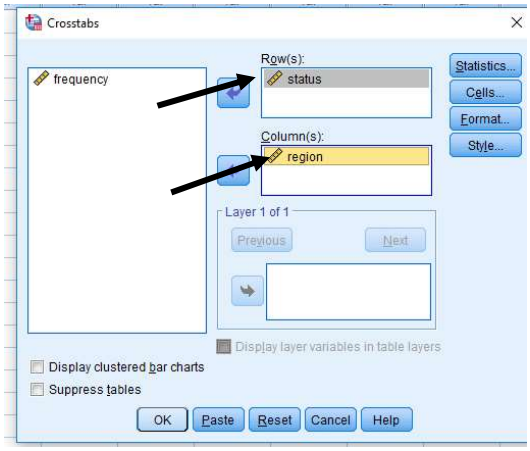
7. We can now begin our analysis of the data. The steps are different from the Goodness of Fit test.

   Stay in the **"Data View"** tab. Select **"Analyze"** then **"Descriptive Statistics"** followed by **"Crosstabs."**

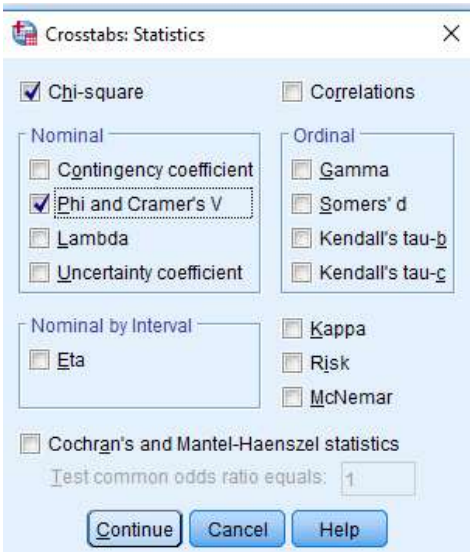


8. A pop up will appear. Move **"status"** to the **"Row"** box using the arrow button. Then move **"region"** to the **"Column"** box, and then click **"OK."**





9. Another pop up will appear. Here we can select some additional analyses. **"Phi"** is a measure of effect size. Effect size is an estimate of the strength of our variables in influencing the observed frequencies. Select **"Phi and Cramer's V,"** and then click **"Continue."**



10. Once you click **"Continue,"** the results of the analyses will appear in an output screen.

Please see the next page for the results

11. The table below provides a summary of the data for each of the variables; that is the obtained frequencies. This is separated by region (column totals), and status totals (row totals). The numbers inside the box (see arrow) are the frequencies that occur when the two variables are combined. These are the data you entered into SPSS.

status * region Crosstabulation

Count

		region		
		city	suburbs	Total
status	single	28	22	50
	married	17	33	50
Total		45	55	100

12. The tables that follow provide the information needed to interpret the results. As shown, the Chi-square value is 4.889, and the *p* value is .027. As our *p* value is **less than .05**, therefore, this result is **significant**. This means the effects of our variables on the observed frequencies are **NOT** independent. Our Phi value is also significant (*p* = .027), suggesting a strong effect.

Chi-square value

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.889[a]	1	.027		
Continuity Correction[b]	4.040	1	.044		
Likelihood Ratio	4.931	1	.026		
Fisher's Exact Test				.044	.022
Linear-by-Linear Association	4.840	1	.028		
N of Valid Cases	100				

p value

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 22.50.

b. Computed only for a 2x2 table

Phi value

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.221	.027
	Cramer's V	.221	.027
N of Valid Cases		100	

13. We now need to work on interpreting out results. Because we have a significant Chi-square, it means that there is a relationship between status and region. Look back at the data table. Single employees prefer to remain in the city. In contrast, married employees would prefer to work in the suburbs. Unfortunately, the owner of the accounting firm has no clear information to decide whether the new location should be in the city or suburbs because preference varies depending upon marital status.

    Interpretation

    The analysis showed a significant relationship between marital status and location preference, $\chi^2(2) = 4.89$, $p = .027$. Married employees would prefer to work in the suburbs, whereas single employees prefer to remain in the city.

Practice Problem

Consider the following scenario.

The director of college housing noticed that there is unused space in one of the largest dorms. He is trying to decide how to make best use of the space, and is undecided as to whether to convert the space to a fitness room or and entertainment room. He wants to make sure that he makes the correct decision based on the preferences of his student residents, he does, however, worry that there may be differences based on sex. He asks 60 male and 60 female residents whether they would prefer that the space be converted to a fitness room or an entertainment center. The results of the survey are presented below.

		Sex	
		Male	**Female**
Space Preference	**Fitness**	**20**	**28**
	Entertainment	**40**	**32**

Perform the appropriate Chi-square statistic on these data to determine if there is a preference in how the space is used, and determine whether preference does this vary by sex.

Once you have completed the analysis, try to write an interpretation of the results.