PERFORMING THE χ^2 TEST USING SPSS

OBJECTIVE

In this lecture, you will learn the following items:

- Performing the χ^2 Goodness-of-Fit Test Using SPSS

- Performing the χ^2 Test for Independence Using SPSS

Performing the χ^2 Goodness-of-Fit Test Using SPSS

We will analyze the data from the example earlier using SPSS.

1. Define Your Variables

First, click the "Variable View" tab at the bottom of your screen (see Fig. 1). The χ^2 goodness-of-it test requires two variables:
one variable to identify the categories and a second variable to identify the observed frequencies. Type the names of these variables in the "Name" column. In our example, we define the variables as "Program" and "count."

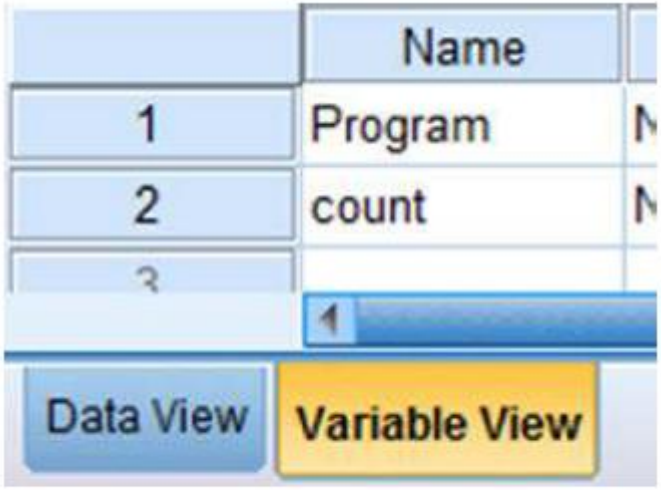

FIGURE 1

You must assign values to serve as a reference for each category in the observed frequency variable. It is often easiest to assign each category a whole number value.

As shown in Figure 2, our categories are “Program 1,” “Program 2,” and “Program 3.”

First, we selected the “count” variable and clicked the gray square in the “Values” field.

Then, we set a value of 1 to equal “Program 1,” a value of 2 to equal “Program 2,” and a value of 3 to equal “Program 3.” We use the “Add” button to move the variable labels to the box below.
Repeat this procedure for the “Program” variable so that the output tables will display these labels.

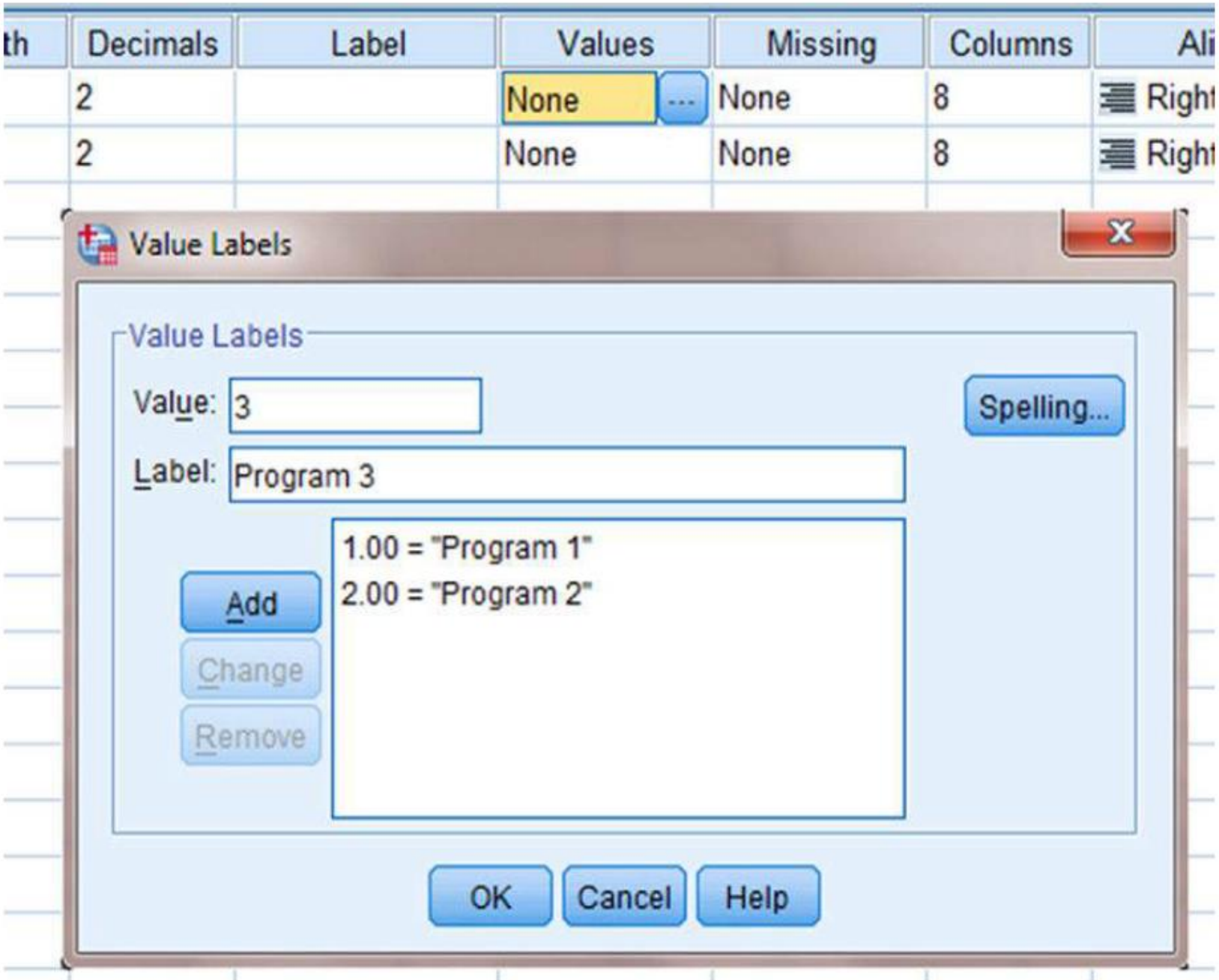
th
Decimals
Label
Values
Missing
Columns
Ali
2
None
None
8
Right
2
None
None
8
Right
Value Labels
Value Labels
Value: 3
Spelling...
Label: Program 3
1.00 = "Program 1"
2.00 = "Program 2"
Add
Change
Remove
OK
Cancel
Help

FIGURE 2

2. Type in Your Values

Click the “Data View” tab at the bottom of your screen. First, enter the data for each category using the whole numbers you assigned to represent the categories. As shown in Figure 8.3, we entered the values “1,” “2,” and “3” in the “Program” variable. Second, enter the observed frequencies next to the corresponding category values. In our example, we entered the observed frequencies “110,” “55,” and “85.”

As shown in Figure 3, we entered the values “1,” “2,” and “3” in the “Program” variable. Second, enter the observed frequencies next to the corresponding category values. In our example, we entered the observed frequencies “110,” “55,” and “85.”

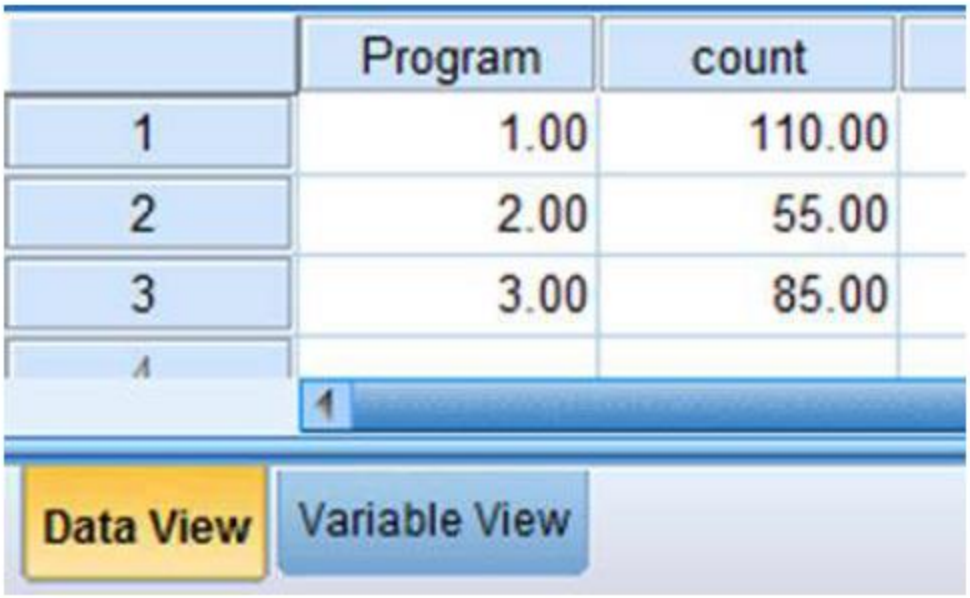

FIGURE 3

3. Analyze Your Data

First, use the “Weight Cases” command to allow the observed frequency variable to reference the category variable.

As shown in Figure 4, use the pull-down menus to choose “Data” and “Weight Cases . . .”.

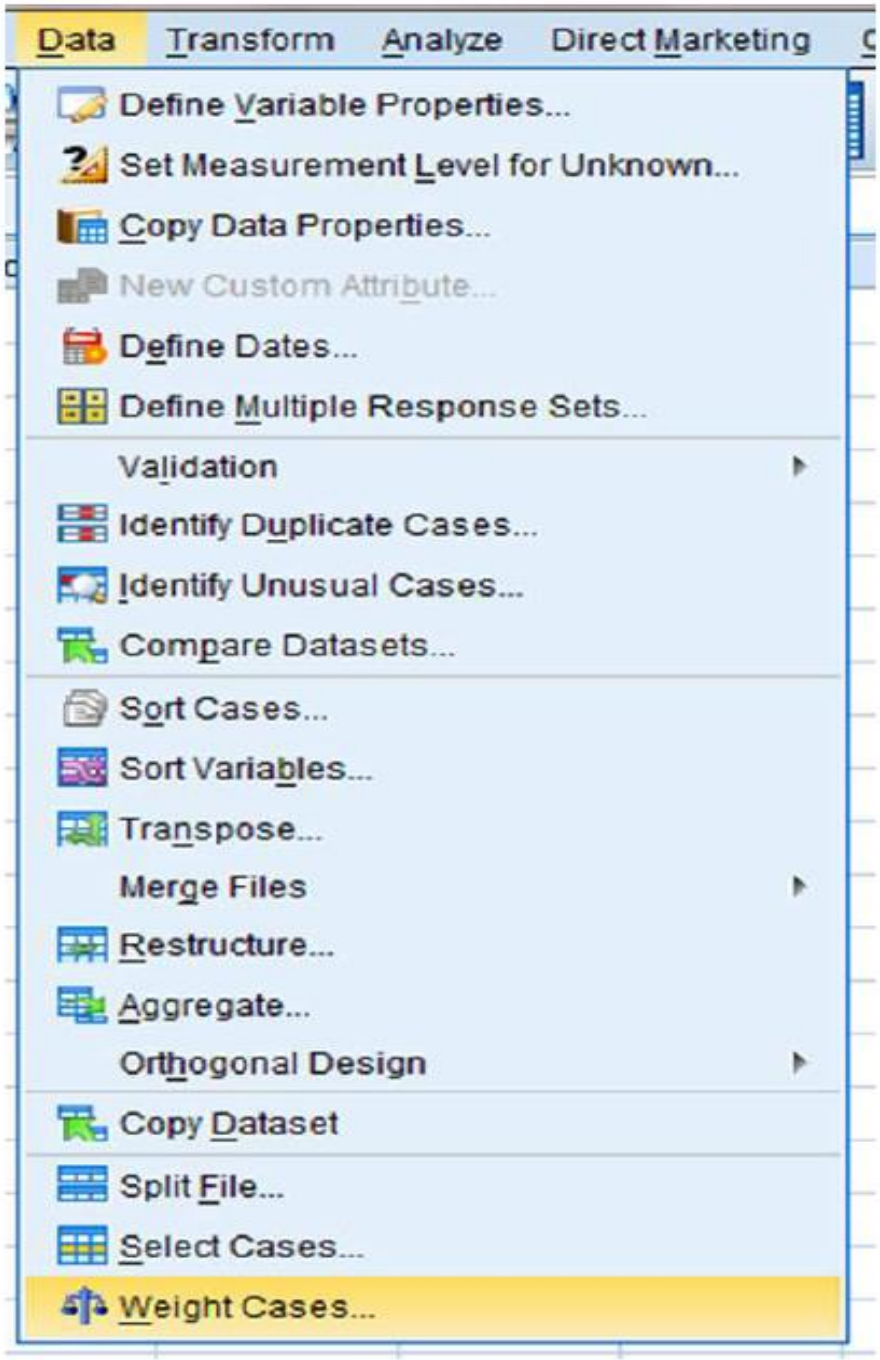
Data
Transform
Analyze
Direct Marketing
Define Variable Properties...
Set Measurement Level for Unknown...
Copy Data Properties...
New Custom Attribute...
Define Dates...
Define Multiple Response Sets...
Validation
Identify Duplicate Cases...
Identify Unusual Cases...
Compare Datasets...
Sort Cases...
Sort Variables...
Transpose...
Merge Files
Restructure...
Aggregate...
Orthogonal Design
Copy Dataset
Split File...
Select Cases...
Weight Cases...

FIGURE 4

The default setting is “Do not weight cases.” Click the circle next to “Weight cases by” as shown in Figure 5. Select the variable with the observed frequencies.

Move that variable to the “Frequency Variable:” box by clicking the small arrow button. In our example, we have moved the “count” variable. Finally, click “OK.”

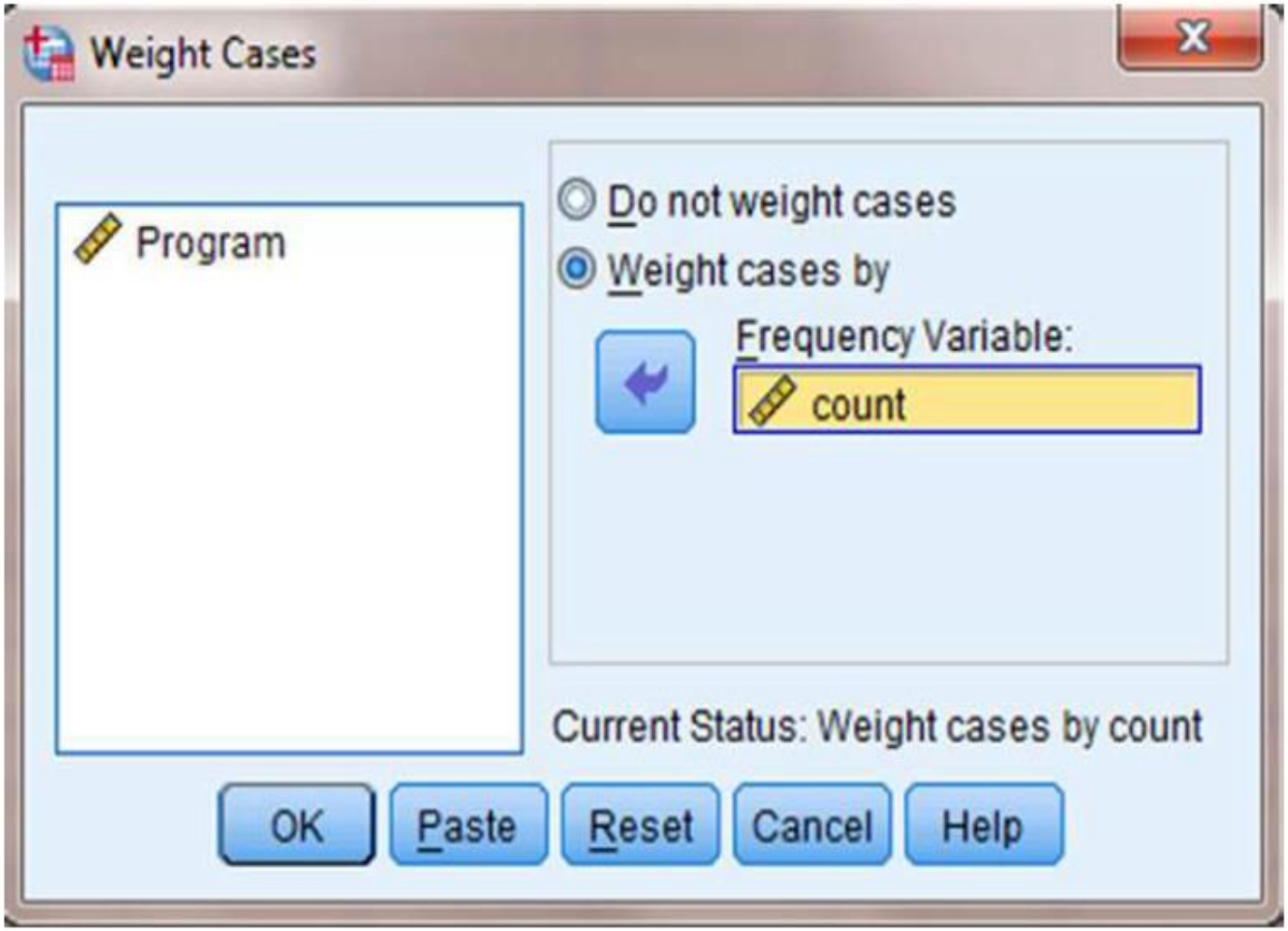
Weight Cases
Program
Do not weight cases
Weight cases by
Frequency Variable:
count
Current Status: Weight cases by count
OK
Paste
Reset
Cancel
Help

FIGURE 5

As shown in Figure 6, use the pull-down menus to choose “Analyze,” “Nonparametric Tests,” “Legacy Dialogs,” and “Chi-square. . .”.

First, move the category variable to the “Test Variable List:” box by selecting that variable and clicking the small arrow button near the center of the window.

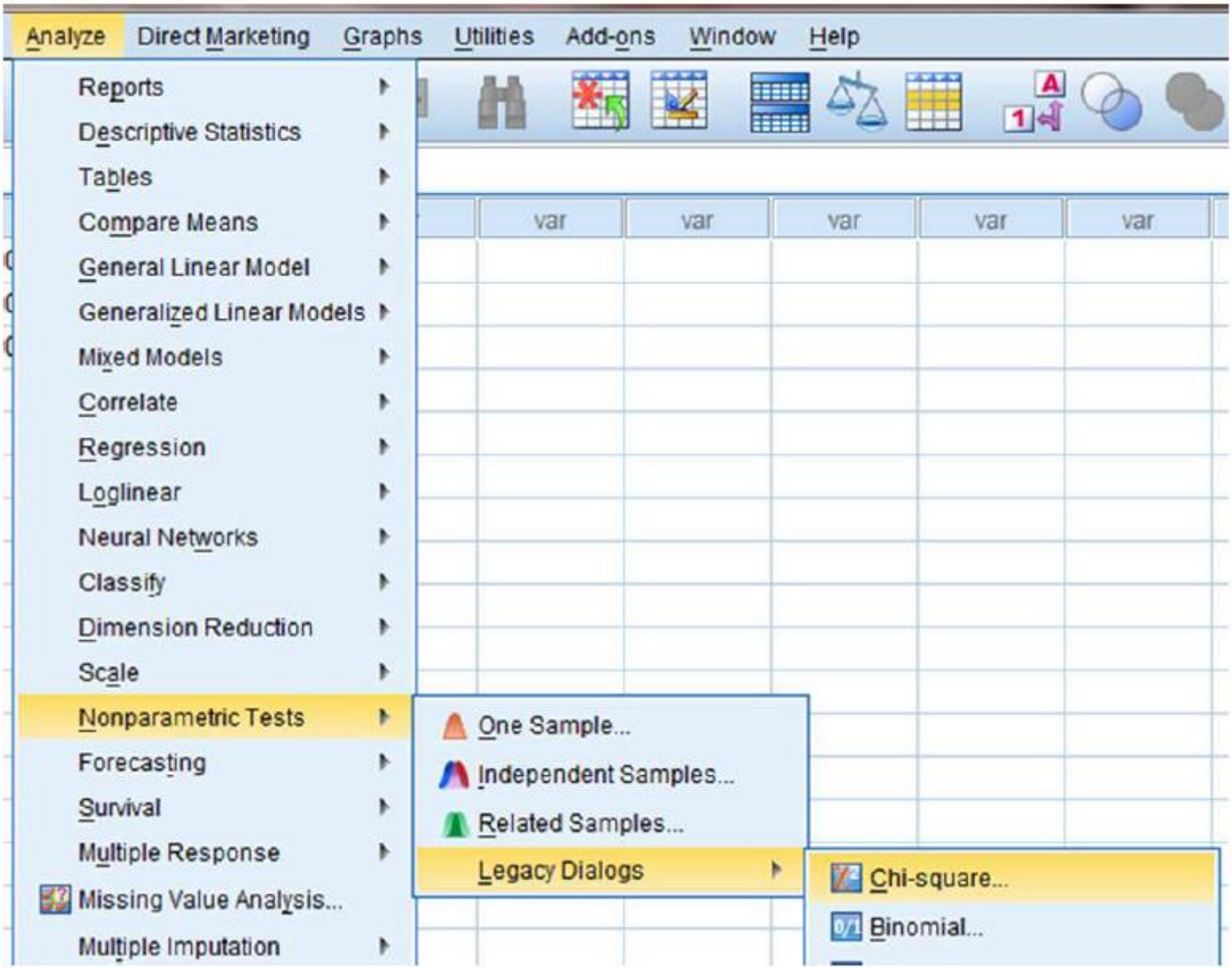
Analyze
Direct Marketing
Graphs
Utilities
Add-ons
Window
Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
One Sample...
Independent Samples...
Related Samples...
Legacy Dialogs
Chi-square...
Binomial...
var
var
var
var
var

FIGURE 6

As shown in Figure 7, we have chosen the “Program” variable.

Then, enter your “Expected Values.”

Notice that the option “All categories equal” is the default setting.

Since this example does not have equal categories, we must select the “Values:” option to set the expected values. Enter the expected frequencies for each category in the order that they are listed in the SPSS Data View.

After you type in an expected frequency in the “Values:” field, click “Add.”
For our example, we have entered 80, 55, and 115, respectively. Finally, click “OK” to perform the analysis.

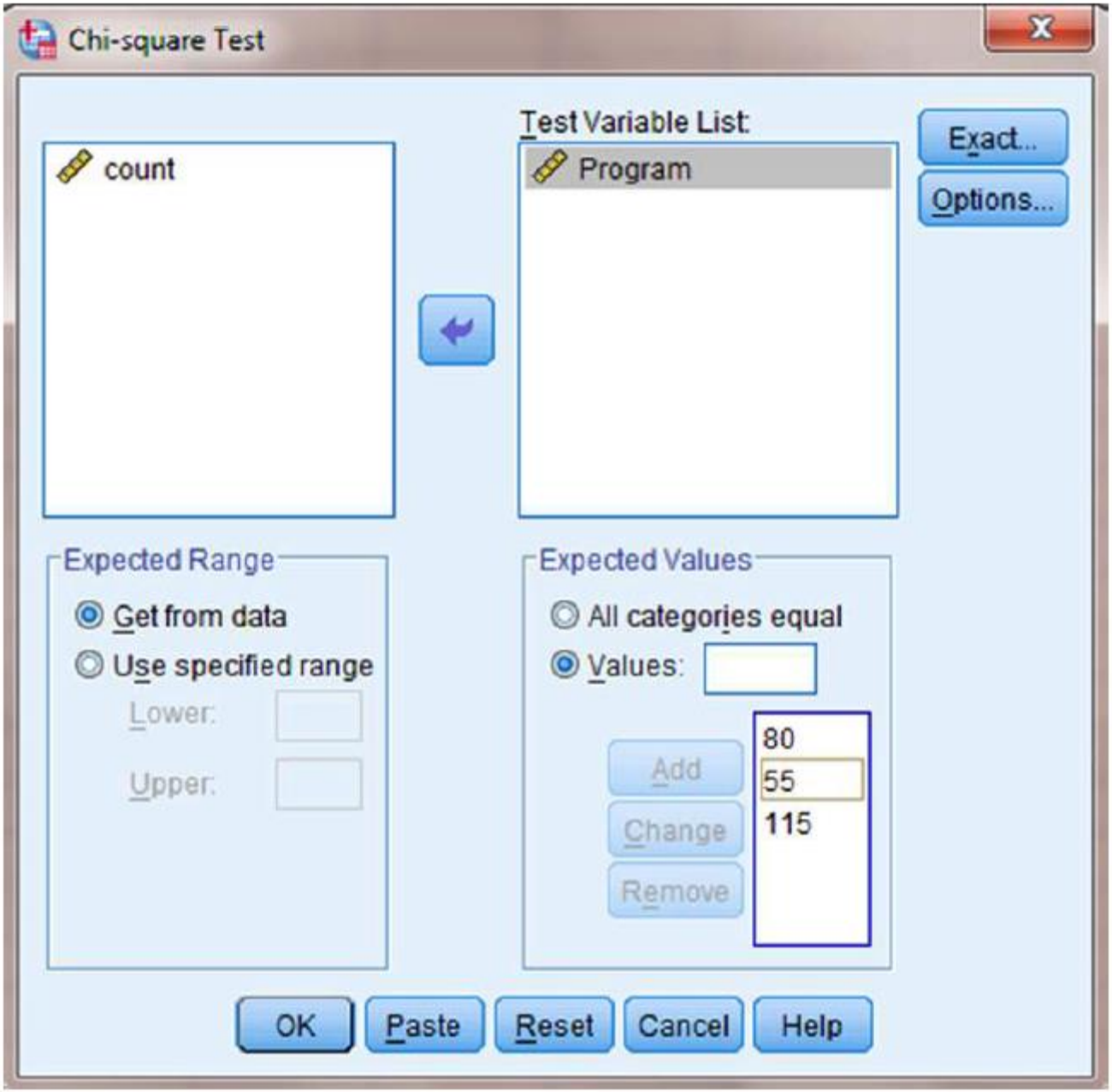
Chi-square Test
count
Test Variable List:
Program
Exact...
Options...
Expected Range
Get from data
Use specified range
Lower:
Upper:
Expected Values
All categories equal
Values:
80
55
115
Add
Change
Remove
OK
Paste
Reset
Cancel
Help

FIGURE 7

4. Interpret the Results from the SPSS Output Window

The first output table (see SPSS Output 1A) provides the observed and expected frequencies for each category and the total count.

Program

	Observed N	Expected N	Residual
Program 1	110	80.0	30.0
Program 2	55	55.0	.0
Program 3	85	115.0	-30.0
Total	250		

SPSS OUTPUT 1A

The second output table (see SPSS Output 1B) provides the χ^2 statistic (χ^2 = 19.076), the degrees of freedom (df = 2), and the significance (p 0.000).

Test Statistics

	Program
Chi-Square	19.076[a]
df	2
Asymp. Sig.	.000

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 55.0.

SPSS OUTPUT 1B

Based on the results from SPSS, three programs were compared with unequal expected frequencies. The χ^2 goodness-of-it test was significant ($\chi^2_{(2)}$ =19 08, $p < 0.01$).

Based on these results, program χ^2 was least effective in both cases, with no difference between the two. Program 1 became more effective and program 3 became less effective.

Performing the χ^2 Test for Independence Using SPSS

We will analyze the data from the example earlier using SPSS.

1. Define Your Variables

First, click the “Variable View” tab at the bottom of your screen, as shown in Figure 8.

The 2-test for independence requires variables to identify the conditions in the rows: one variable to identify the conditions of the rows and a second variable to identify the conditions of the columns.

According to the previous example, the “Behavior” variable will represent the columns. “School_Type” will represent the rows. Finally, we need a variable to represent the observed frequencies. “Frequency” represents the observed frequencies.

You must assign values to serve as a reference for the column and row variables.

It is often easiest to assign each category a whole number value. First, click the gray square in the “Values” field to set the desired values.

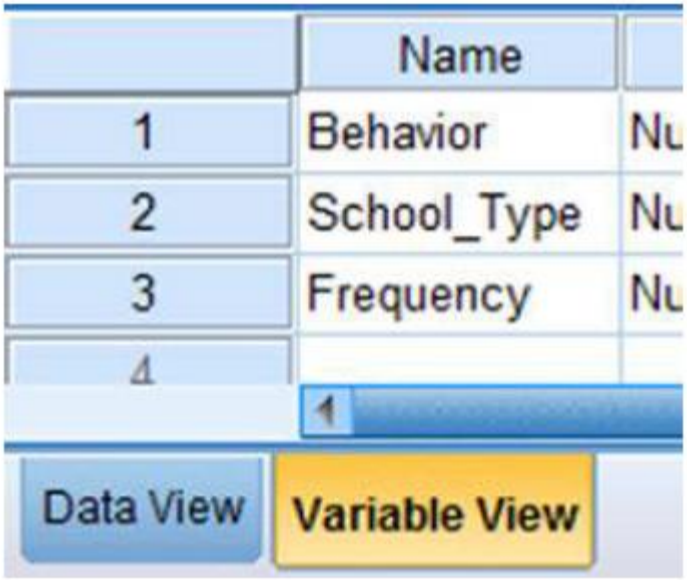

FIGURE 8

As shown in Figure 9, we have already assigned the value labels for the “Behavior” variable.

For the “School_Type” variable, we set a value of 1 to equal “Public Preschool,” a value of 2 to equal “Private Preschool,” and a value of 3 to equal “No Preschool.” Clicking the “Add” button moves each of the value labels to the list below.

Finally, click “OK” to return to the SPSS Variable View screen.

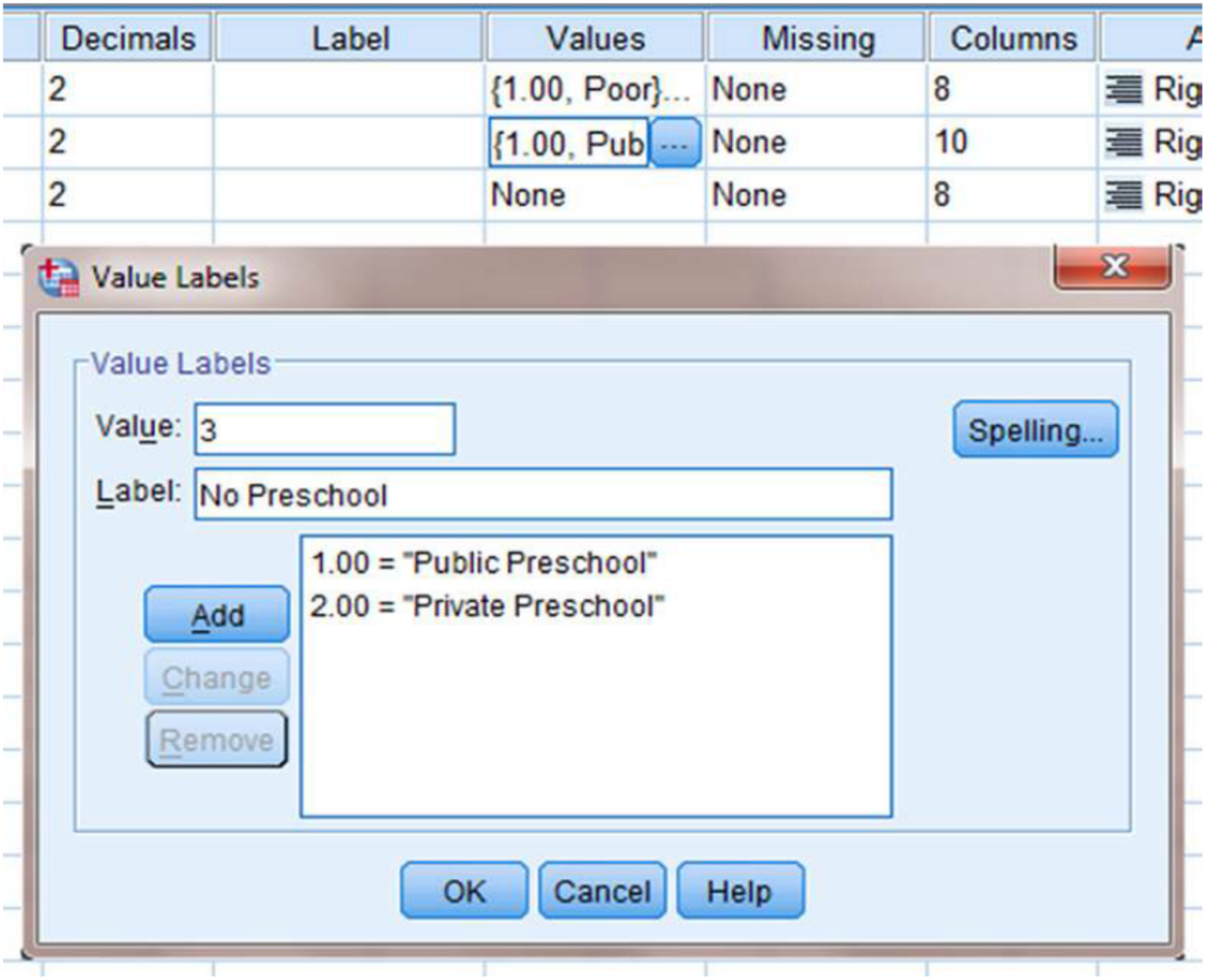
Decimals
Label
Values
Missing
Columns
2
{1.00, Poor}...
None
8
Rig
2
{1.00, Pub
None
10
Rig
2
None
None
8
Rig
Value Labels
Value Labels
Value: 3
Spelling...
Label: No Preschool
1.00 = "Public Preschool"
2.00 = "Private Preschool"
Add
Change
Remove
OK
Cancel
Help

FIGURE 9

2. Type in Your Values

Click the “Data View” tab at the bottom of your screen, as shown in Figure 10.

Use the whole number references you set earlier for the row and column variables. Each possible combination of conditions should exist. Then, enter the corresponding observed frequencies.

In our example, row 1 represents a “Behavior” of 1 which is “Poor” and a “School_Type” of 1 which is “Public School.” The observed frequency for this condition is 12.

	Behavior	School_Type	Frequency
1	1.00	1.00	12.00
2	2.00	1.00	25.00
3	3.00	1.00	10.00
4	1.00	2.00	6.00
5	2.00	2.00	12.00
6	3.00	2.00	.00
7	1.00	3.00	2.00
8	2.00	3.00	23.00
9	3.00	3.00	10.00
10			

Data View | Variable View

FIGURE 10

3. Analyze Your Data

First, use the “Weight Cases” command to allow the observed frequency variable to reference the category variable. As shown in Figure 11, use the pull-down menus to choose “Data” and “Weight Cases . . .”.

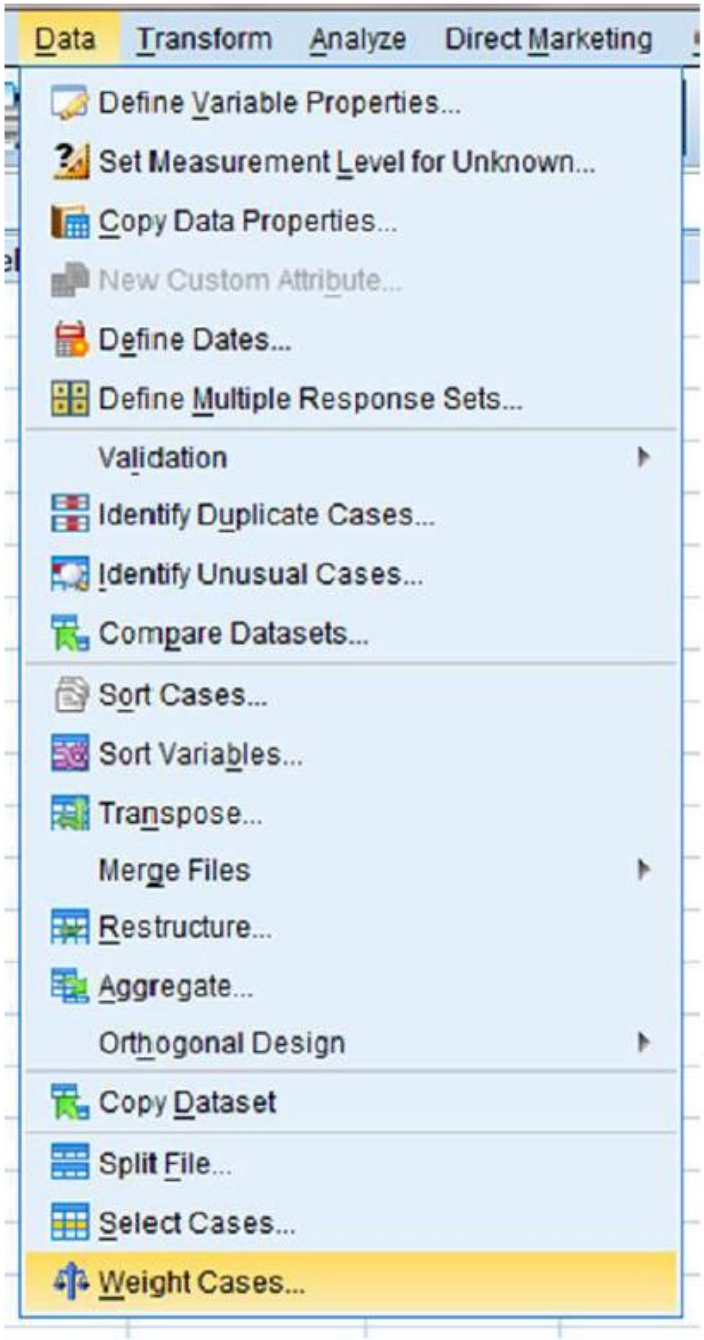
Data
Transform
Analyze
Direct Marketing
Define Variable Properties...
Set Measurement Level for Unknown...
Copy Data Properties...
New Custom Attribute...
Define Dates...
Define Multiple Response Sets...
Validation
Identify Duplicate Cases...
Identify Unusual Cases...
Compare Datasets...
Sort Cases...
Sort Variables...
Transpose...
Merge Files
Restructure...
Aggregate...
Orthogonal Design
Copy Dataset
Split File...
Select Cases...
Weight Cases...

FIGURE 11

The default setting is “Do not weight cases.” Click the circle next to “Weight cases by” as shown in Figure 12.

Select the variable with the observed frequencies. Move that variable to the “Frequency Variable:” box by clicking the small arrow button.

In our example, we have moved the “Frequency” variable. Finally, click “OK.”

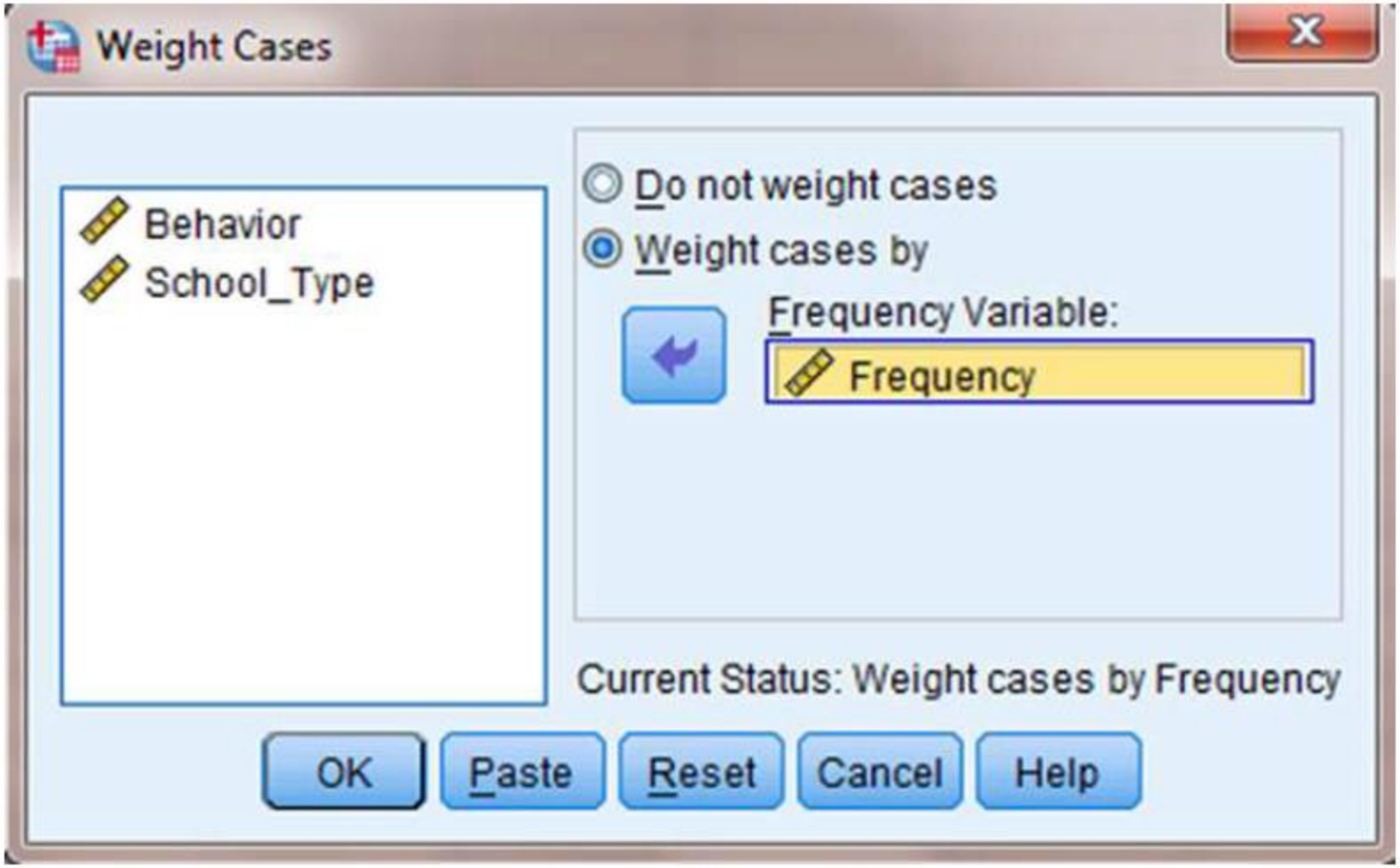
Weight Cases
Behavior
School_Type
Do not weight cases
Weight cases by
Frequency Variable:
Frequency
Current Status: Weight cases by Frequency
OK
Paste
Reset
Cancel
Help

FIGURE 12

As shown in Figure 13, use the pull-down menus to choose “Analyze,” “Descriptive Statistics,” and “Crosstabs . . . ”.

When the Crosstabs window is open, move the variable that represents the rows to the “Row(s):” box by selecting that variable and clicking the small arrow button next to that box.

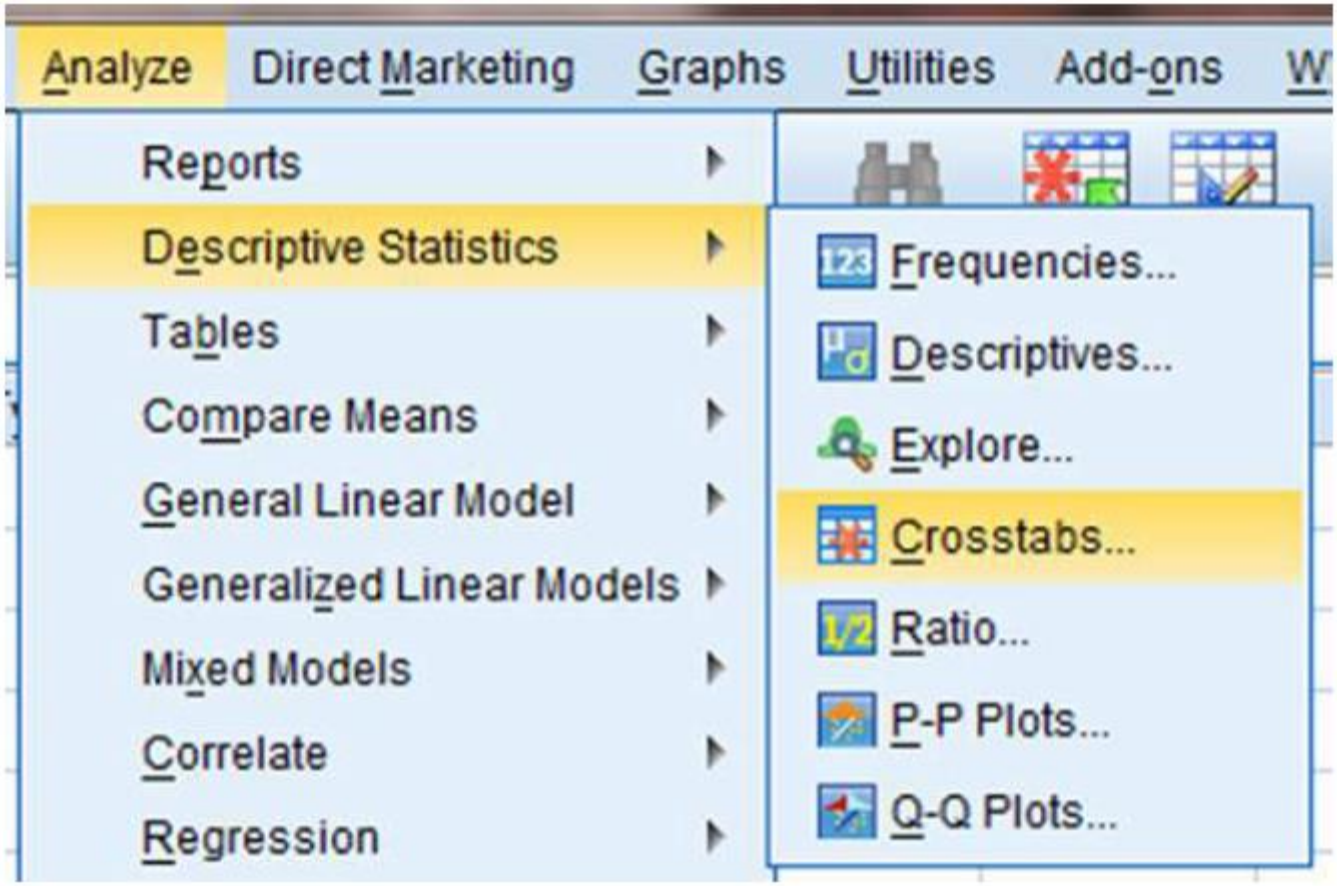
Analyze
Direct Marketing
Graphs
Utilities
Add-ons
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Frequencies...
Descriptives...
Explore...
Crosstabs...
Ratio...
P-P Plots...
Q-Q Plots...

FIGURE 13

As shown in Figure 14, we have chosen the “School_Type” variable. Then, move the variable that represents the column to the “Column(s):” box.

In our example, we have chosen the “Behavior” variable. Next, click the “Statistics . . .” button.

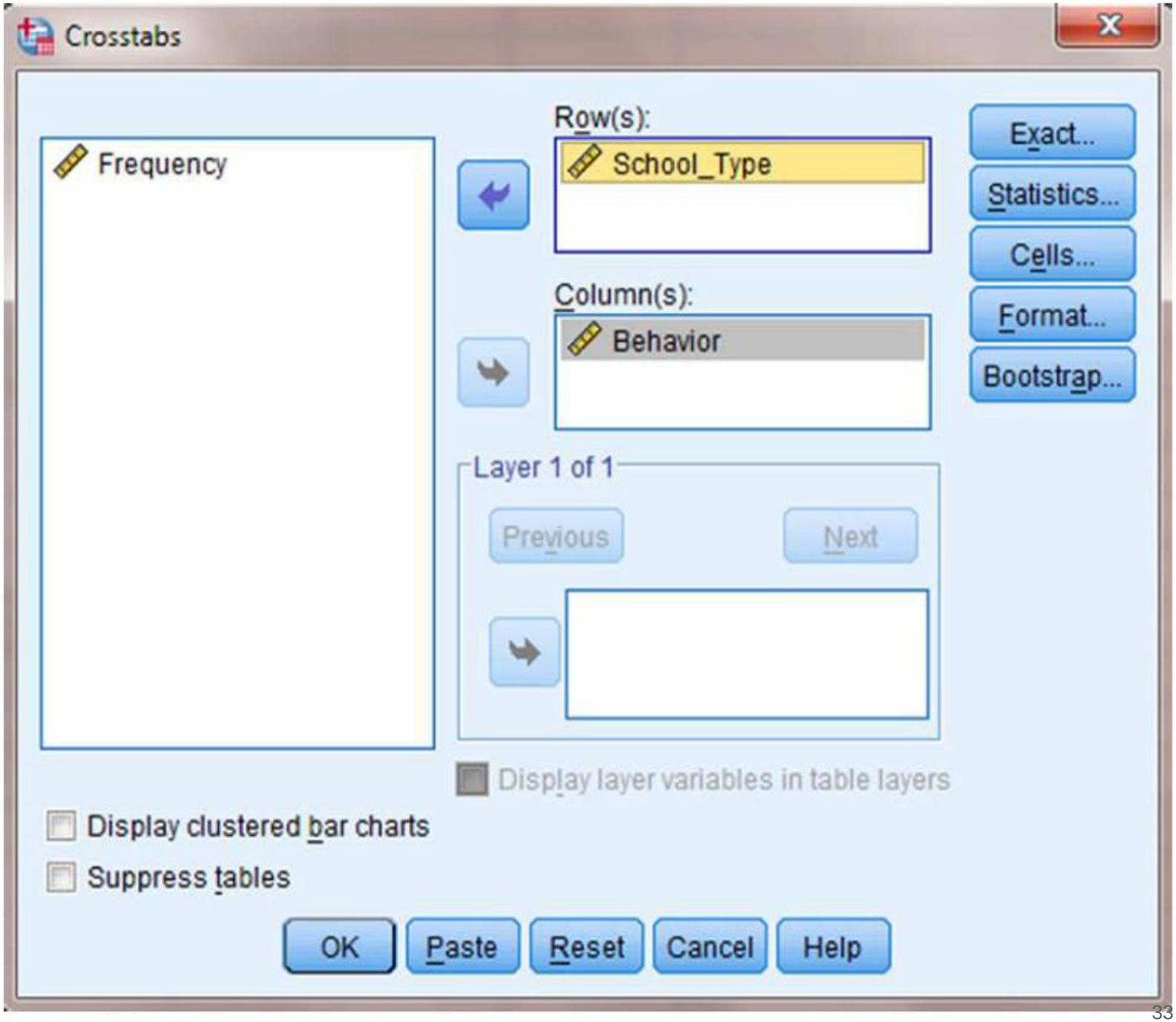
Crosstabs
Frequency
Row(s):
School_Type
Column(s):
Behavior
Layer 1 of 1
Previous
Next
Display layer variables in table layers
Display clustered bar charts
Suppress tables
Exact...
Statistics...
Cells...
Format...
Bootstrap...
OK
Paste
Reset
Cancel
Help

FIGURE 14

As shown in Figure 15, check the box next to “Chi-square” and the box next to “Phi and Cramer’s V.”

Once those boxes are checked, click “Continue” to return to the Crosstabs window. Now, click the “Cells . . .” button.

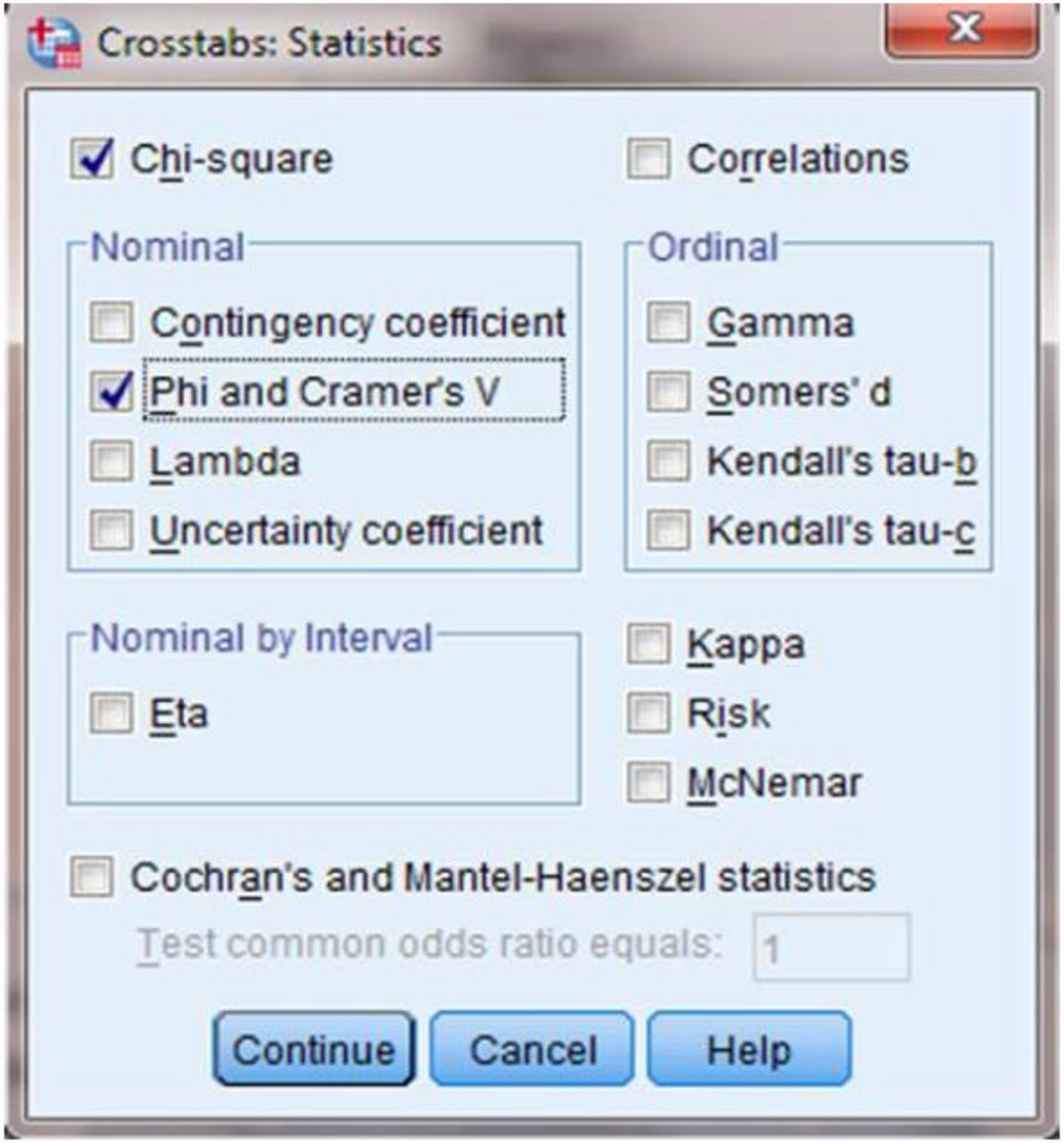
Crosstabs: Statistics
Chi-square
Correlations
Nominal
Contingency coefficient
Phi and Cramer's V
Lambda
Uncertainty coefficient
Ordinal
Gamma
Somers' d
Kendall's tau-b
Kendall's tau-c
Nominal by Interval
Eta
Kappa
Risk
McNemar
Cochran's and Mantel-Haenszel statistics
Test common odds ratio equals: 1
Continue
Cancel
Help

FIGURE 15

As shown in Figure 16, check the boxes next to “Observed” and “Expected.”

Then, click “Continue” to return to the Crosstabs window.

Finally, click “OK” to perform the analysis.

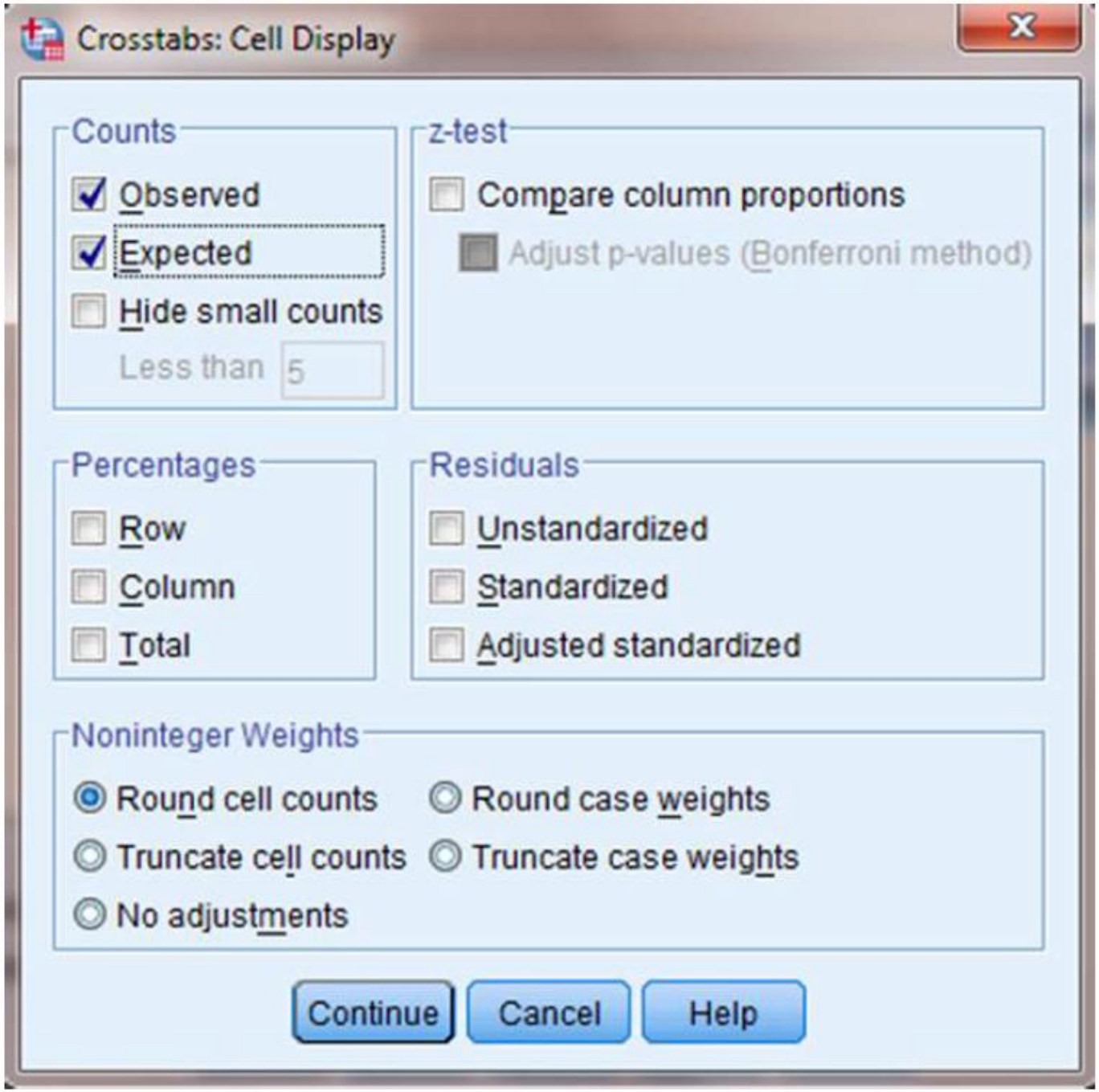
Crosstabs: Cell Display
Counts
Observed
Expected
Hide small counts
Less than 5
z-test
Compare column proportions
Adjust p-values (Bonferroni method)
Percentages
Row
Column
Total
Residuals
Unstandardized
Standardized
Adjusted standardized
Noninteger Weights
Round cell counts
Round case weights
Truncate cell counts
Truncate case weights
No adjustments
Continue
Cancel
Help

FIGURE 16

4. Interpret the Results from the SPSS Output Window

The second to fourth output tables from SPSS are of interest in this procedure.

The second output table (see SPSS Output 2A) provides the observed and expected frequencies for each category and the total counts.

School_Type * Behavior Crosstabulation

			Behavior			Total
			Poor	Average	Good	
School_Type	Public Preschool	Count	12	25	10	47
		Expected Count	9.4	28.2	9.4	47.0
	Private Preschool	Count	6	12	0	18
		Expected Count	3.6	10.8	3.6	18.0
	No Preschool	Count	2	23	10	35
		Expected Count	7.0	21.0	7.0	35.0
Total		Count	20	60	20	100
		Expected Count	20.0	60.0	20.0	100.0

SPSS OUTPUT 2A

The third output table (see SPSS Output 2B) provides the χ^2 statistic (χ^2 = 11.502), the degrees of freedom (df = 4), and the significance (p = 0.021).

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11.502[a]	4	.021
Likelihood Ratio	16.042	4	.003
Linear-by-Linear Association	3.072	1	.080
N of Valid Cases	100		

a. 2 cells (22.2%) have expected count less than 5. The minimum expected count is 3.60.

SPSS OUTPUT 2B

The fourth output table (see SPSS Output 2C) provides the Cramer's V statistic (V = 0.240) to determine the level of association or effect size.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.339	.021
	Cramer's V	.240	.021
N of Valid Cases		100	

SPSS OUTPUT 2C

Based on the results from SPSS, three programs were compared with unequal expected frequencies.

The χ^2 goodness-of-it test was significant ($\chi^2_{(4)}$ =11 502, $p < 0.05$).

Based on these results, there is a real association between type of preschool experience children obtained and their behavior in the kindergarten classroom during their first few weeks in school.

In addition, the measured effect size presented a medium level of association (V = 0.240).

SUMMARY

Nominal, or categorical, data sometimes need analyses. In such cases, you may be seeking to determine if the data statistically match some known or expected set of frequencies. Or, you may wish to determine if two or more categories are statistically independent. In either case, nominal data can be analyzed with a nonparametric procedure.

In this lecturer, we explained how to perform the procedures for examining nominal data:

chi-square χ^2 goodness of it and χ^2 -test for independence, using SPSS.