

Chapter 18: Categorical data

Self-test answers



SELF-TEST Run a multiple regression analysis using **Cat Regression.sav** with **LnObserved** as the outcome, and **Training**, **Dance** and **Interaction** as your three predictors.

The multiple regression dialog box will look like Figure 1. We can leave all of the default options as they are because we are interested only in the regression parameters. The regression parameters are shown in the book.

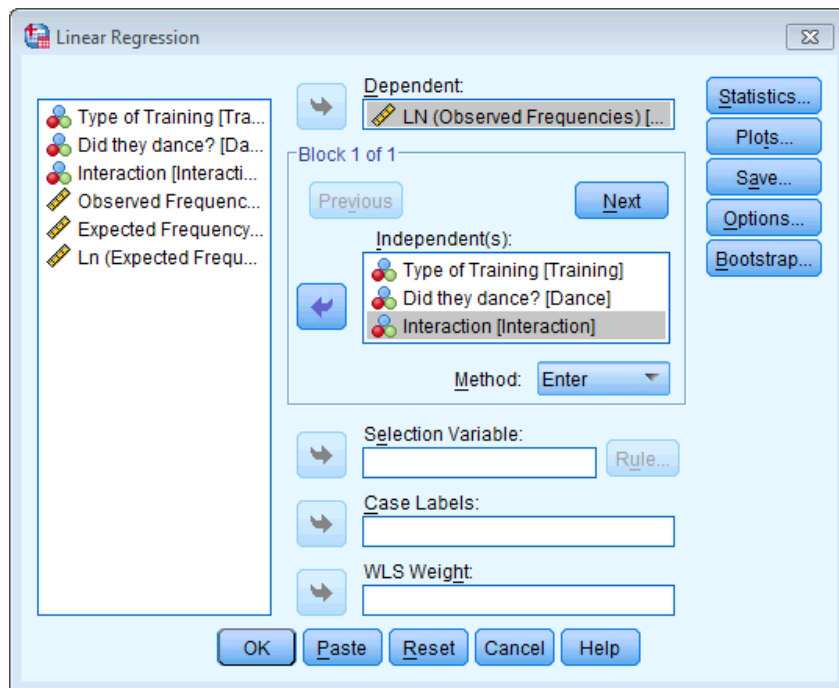


Figure 1



SELF-TEST To show that this all actually works, run another multiple regression analysis using **Cat Regression.sav**. This time the outcome is the log of expected frequencies (**LnExpected**) and **Training** and **Dance** are the predictors (the interaction is not included).

The multiple regression dialog box will look like Figure 2. We can leave all of the default options as they are because we are interested only in the regression parameters.

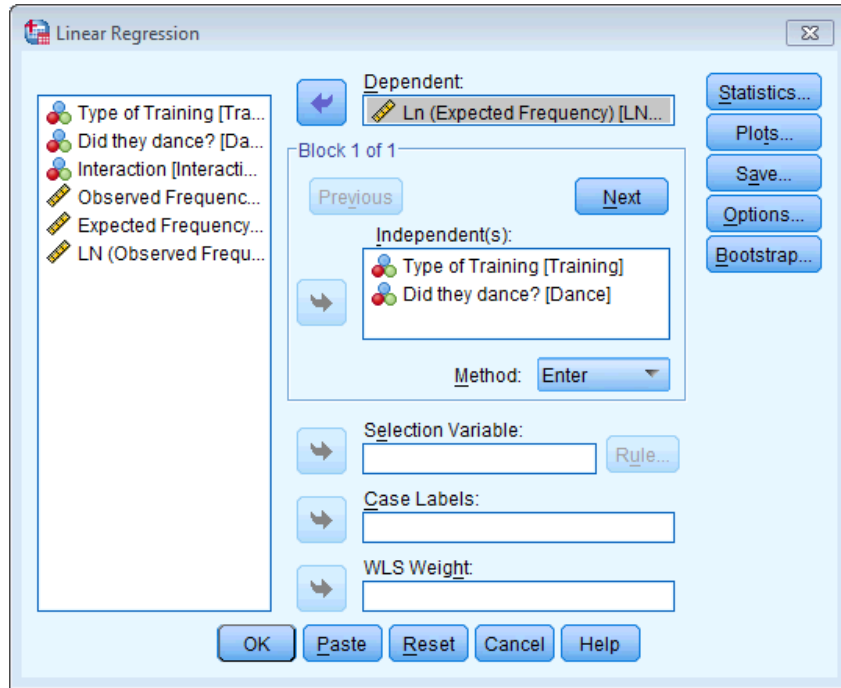


Figure 2

The resulting regression parameters are given in Output 1.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.160	.000		366695255	.000
	Type of Training	1.450	.000	.824	173423149	.000
	Did they dance?	-.490	.000	-.344	-72443568	.000

a. Dependent Variable: Ln (Expected Frequency)

Output 1

Note that $b_0 = 3.16$, the beta coefficient for the type of training is 1.45 and the beta coefficient for whether they danced is -0.49 . All of these values are consistent with those calculated in the book chapter.



SELF-TEST Using the **Cats Weight.sav** data, change the frequency of cats who had food as reward and didn't dance from 10 to 28. Re-do the chi-square test and select and interpret z-tests (Compare column proportions). Is there anything about the results that seems strange?

You need to change the score so your data look like **Error! Reference source not found**.Figure 3.

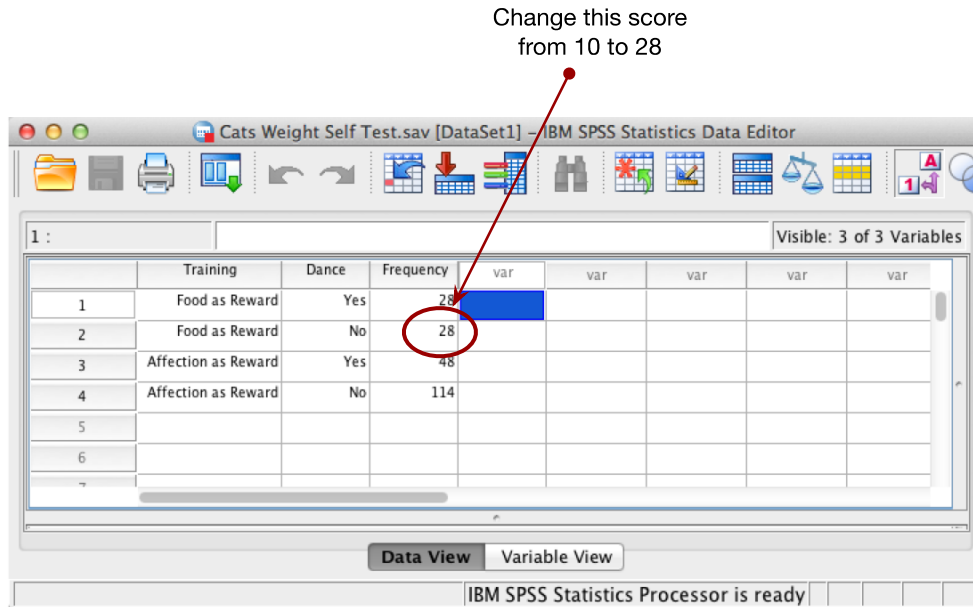


Figure 3

The data are the same as in the chapter so you can follow the instructions in the book to run the analysis. The contingency table you get looks like **Error! Reference source not found**. Output 2. In the row labelled *Food as Reward* the count of 28 in the column labelled *No* has a subscript letter *a* and the count of 28 in the column labelled *Yes* has a subscript letter *b*. These subscripts tell us the results of the *z*-test that we asked for: columns with different subscripts have significantly different column proportions. This is what should strike you as strange: how can it be that two identical counts of 28 can be deemed significantly different? The answer is that despite the subscripts being attached to the counts, that isn't what they compare: they compare the *proportion* of the total frequency of that column that falls into that row against the *proportion* of the total frequency of the second column that falls into that row. In this case, it's testing whether 19.7% is different from 36.8%, and it is ($p < .05$), which is why the column counts have been denoted with different letters. So, of all the cats that danced, 36.8% had food, and of all the cats that didn't dance, 19.7% had food. These proportions are significantly different.

Type of Training * Did they dance? Crosstabulation

			Did they dance?		Total
			No	Yes	
Type of Training	Food as Reward	Count	28 ^a	28 ^b	56
		Expected Count	36.5	19.5	56.0
		% within Type of Training	50.0%	50.0%	100.0%
		% within Did they dance?	19.7%	36.8%	25.7%
		Std. Residual	-1.4	1.9	
	Affection as Reward	Count	114 ^a	48 ^b	162
		Expected Count	105.5	56.5	162.0
		% within Type of Training	70.4%	29.6%	100.0%
		% within Did they dance?	80.3%	63.2%	74.3%
		Std. Residual	.8	-1.1	
Total		Count	142	76	218
		Expected Count	142.0	76.0	218.0
		% within Type of Training	65.1%	34.9%	100.0%
		% within Did they dance?	100.0%	100.0%	100.0%

Each subscript letter denotes a subset of Did they dance? categories whose column proportions do not differ significantly from each other at the .05 level.

Output 2



SELF-TEST Create a contingency table of these data with dance as the columns, the type of training as rows and the type of animal as a layer.




To use the *crosstabs* command select **Analyze** > **Descriptive Statistics** > **Crosstabs...** We have three variables in our crosstabulation table: whether the animal danced or not (**Dance**), the type of reward given (**Training**), and whether the animal was a cat or dog (**Animal**). Select **Training** and drag it into the box labelled *Row(s)* (or click on ). Next, select **Dance** and drag it to the box labelled *Column(s)* (or click on ). We have a third variable too, and we need to define this variable as a layer. Select **Animal** and drag it to the box labelled *Layer 1 of 1* (or click on ). Then click on **Cells...** and select the options shown on the right in Figure 4.



Figure 1



SELF-TEST Can you use the chart builder to replicate the graph in Figure 18.8?



Actually this self-test is not as easy as it looks. Figures 5 and 6 guide you through the process.

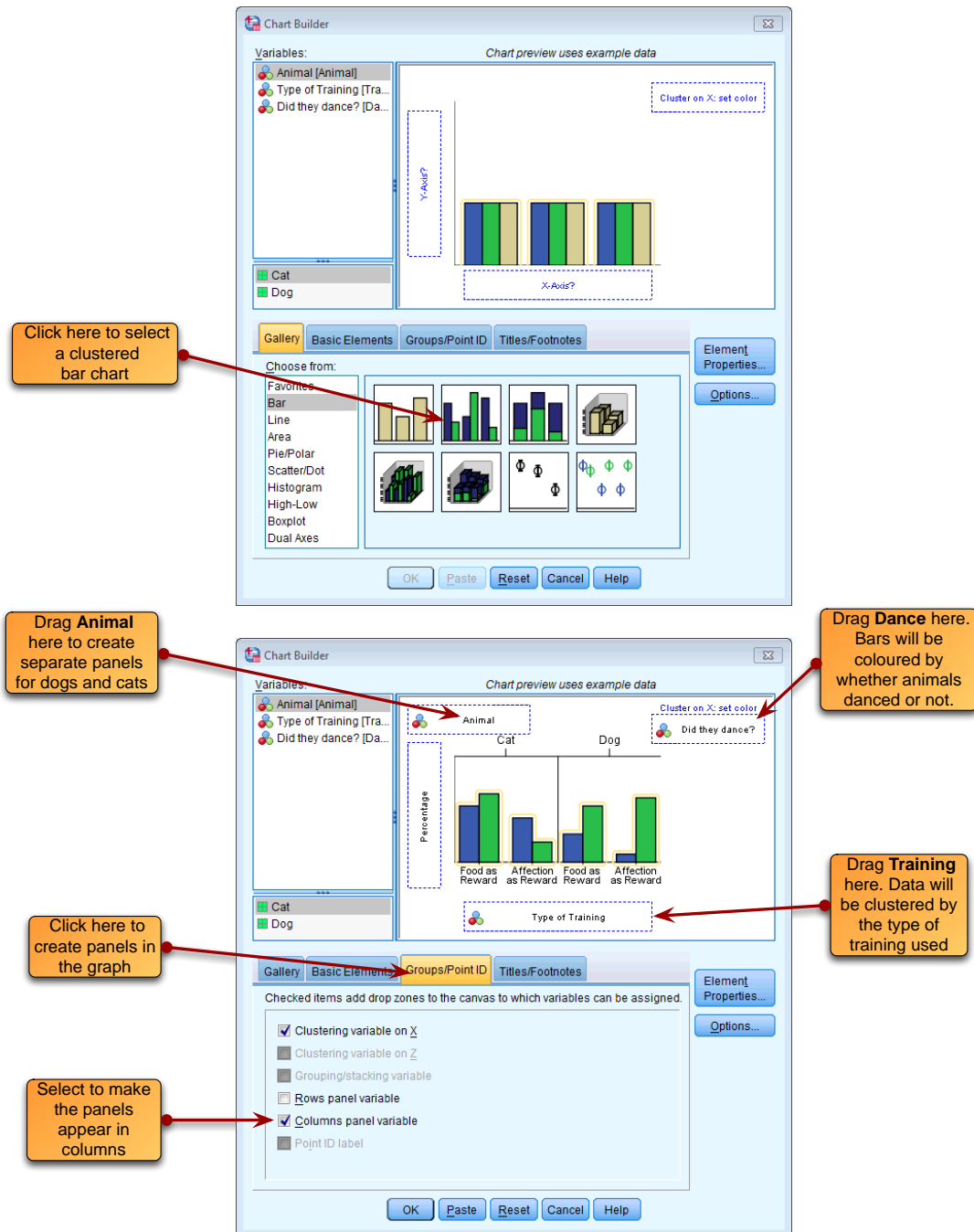


Figure 2

The figure shows two dialog boxes from SPSS. The main dialog is 'Element Properties' for 'Bar1'. It has a 'Statistics' section where 'Percentage()' is selected in the 'Statistic' dropdown. Below it is a 'Set Parameters...' button. There is also a 'Display error bars' section with radio buttons for 'Confidence intervals', 'Standard error', and 'Standard deviation'. At the bottom are 'Apply', 'Close', and 'Help' buttons. A callout box on the left explains that percentages are chosen over counts for direct comparison. A second callout box on the right explains that 'Total for Each Panel' is selected in the 'Denominator for Computing Percentage' dropdown of the 'Set Parameters...' dialog to calculate percentages within each animal group.

We want to display percentages rather than counts because there were more cats than dogs and this will allow us to compare animals directly. To do this, click here and select **Percentage()** from the list

By default SPSS will display the percentage of the total sample. However, we want the percentage to be calculated within each animal (i.e. the percentage of cats that danced for food). To display these percentages, select **Total for Panel** from the drop-down list. This will calculate the percentage within each panel (not all panels combined). This means that we will get the percentage of cats and dogs, not the percentage of all animals.

Don't forget to click here to apply the changes to the graph

Figure 3





SELF-TEST Use the *split file* command to run a chi-square test on **Dance** and **Training** for dogs and cats.

First, to split the file we need to select **Data** **Split File...** and then select the *Organize output by groups* option (Figure 7). Once this option is selected, the *Groups Based on* box will activate. Select the variable containing the group codes by which you wish to repeat the analysis (in this example select **Animal**), and drag it to the box or click on

The 'Split File' dialog box shows three radio button options: 'Analyze all cases, do not create groups', 'Compare groups', and 'Organize output by groups'. The 'Organize output by groups' option is selected. Below it, the 'Groups Based on:' box contains 'Animal [Animal]'. At the bottom, there are 'OK', 'Paste', 'Reset', 'Cancel', and 'Help' buttons. The status at the bottom reads 'Current Status: Analysis by groups is off.'

Figure 7

To run the chi-square tests, select **Analyze** **Descriptive Statistics** **Crosstabs...** (Figure 8). First, select one of the variables of interest in the variable list and drag it into the box labelled *Row(s)* (or click on ). For this example, I selected **Training** to be the rows of the table. Next, select the other variable of interest (**Dance**) and drag it to the box labelled *Column(s)* (or click on ). Select the same options as in the book (for the cat example).

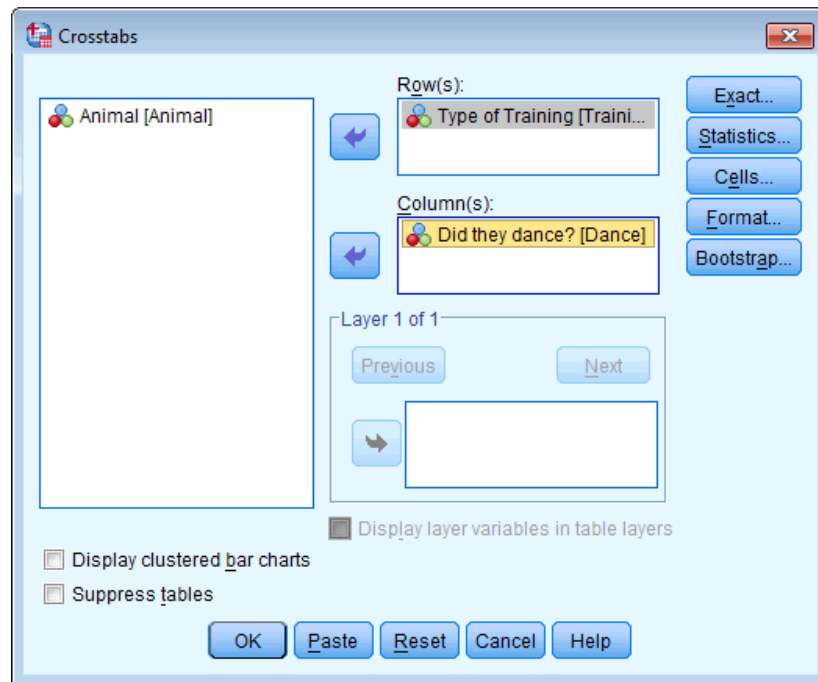


Figure 8