

Crosstabs

Is there a relationship between a respondent's education level and his or her awareness of neighbourhood policing?

Let's say you want to examine the relationship between two categorical variables, such as gender and income level, or location and homeownership. You can use crosstabulation (and eventually a chi-squared test) to determine whether there is a significant relationship between the two categorical variables you are interested in.

Crosstabulation allows you to summarize the data in categorical variables and examine it to determine if there are any relationships present. SPSS provides cross tabulation charts that show you how many individuals (or cases) are present in each group. For example, if you ran a crosstabulation on gender and income bracket, with gender having two categories (female and male) and income level having five categories (very high, high, average, low, and very low), you would be able to see, for example, how women have high incomes and how many men have average incomes.

For our crosstabulations, we should consider variables that reflect our research question, which is, "Is there a relationship between a respondent's education level and his or her awareness of neighbourhood policing?" We can continue to use **neighpol1**, as it concerns respondent awareness of neighbourhood policing. Because we're interested in illuminating the relationship between respondent neighbourhood policing awareness and education level, the second categorical variable we'll use is **educat3**, which catalogues respondent educational attainment.

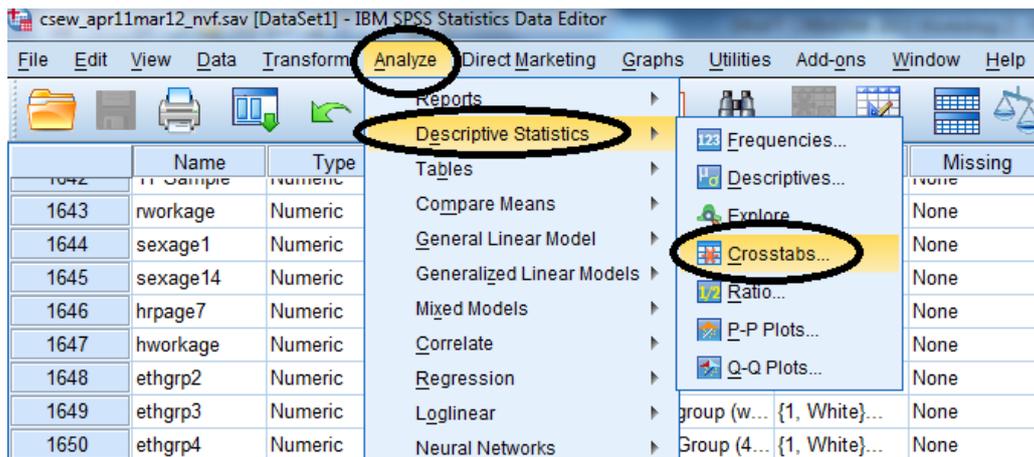
Before we begin this exploratory analysis, we should check the frequencies of data in **neighpol1** and **educat3**. We're running a quick frequency check in an effort to identify any potential issues with our data before we use it in further analyses.

Just select **Analyze, Descriptive Statistics**, and then **Frequencies**. Move **neighpol1** and **educat3** into the **Variable(s)** list on the right of the dialogues box and click **OK**. In the SPSS Output window, you should now have some output tables that will give you information about the two variables. In these tables, you should look to make sure that the only data available for analysis is data from within the categories of your variables. You'll want to exclude any data that doesn't look like it belongs, as this could be coded missing data or data entered into the dataset in error. Using a full dataset provides you with the ability to make any changes necessary to improve your analysis of the data. This is both a blessing and curse, as while you have access to more variables and improved freedom of analysis, you must also be wary of any errors that may be present in the dataset.

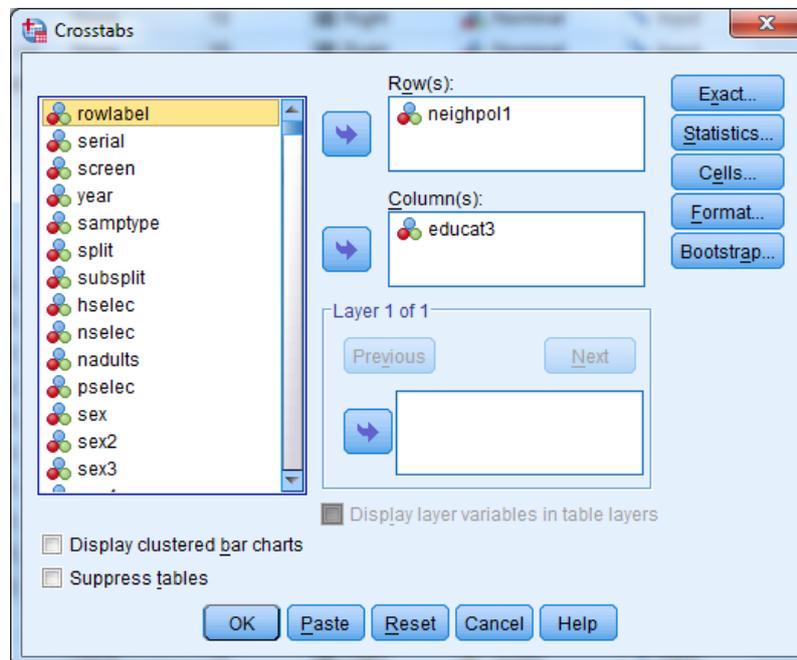
Luckily for us, the frequency tables produced for **neighpol1** and **educat3** look good! There are no problematic categories or data in these variables. In both variables, there are cases listed as System Missing, which means that data was not available for the chosen variables in these cases. System Missing data is not an issue in analysis, because SPSS recognizes these data points as being troublesome and excludes them from any models or tests you run.

Now that we're sure that our data is clean and ready for analysis, we can run some cross tabulations, asking SPSS to calculate the counts and percentage totals for respondents who fall into each group.

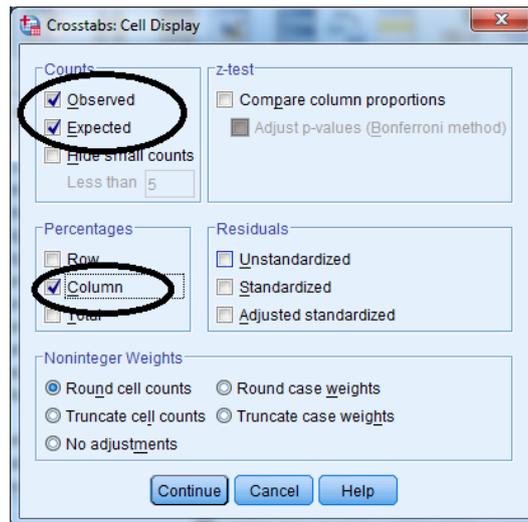
Select **Analyze, Descriptive Statistics**, and then **Crosstabs**.



Find **neighpol1** in the variable list on the left, and move it to the **Row(s)** box. Find **educat3** in the variable list on the left, and move it to the **Column(s)** box.



Click on the **Cells** button, and then select **Column** under the **Percentages** header. While you're here, select both **Observed** and **Expected** under the **Counts** header. (You'll see why we've done this in a moment.)



Click **Continue**, and then click **OK** in the original dialogue box. Your output should look like this:

Aware of Neighbourhood Policing Team in your local area - recoded * Respondent education (5 categories) Crosstabulation

			Respondent education (5 categories)					Total
			None	O level/GCSE	Apprenticeship or NVQ level	Degree or diploma	Other	
Aware of Neighbourhood Policing Team in your local area - recoded	Yes	Count	971	991	920	1920	193	4995
		Expected Count	1196.7	993.1	894.2	1702.0	208.9	4995.0
		% within Respondent education (5 categories)	36.0%	44.3%	45.6%	50.0%	41.0%	44.4%
	No	Count	1727	1248	1096	1917	278	6266
		Expected Count	1501.3	1245.9	1121.8	2135.0	262.1	6266.0
		% within Respondent education (5 categories)	64.0%	55.7%	54.4%	50.0%	59.0%	55.6%
Total	Count	2698	2239	2016	3837	471	11261	
	Expected Count	2698.0	2239.0	2016.0	3837.0	471.0	11261.0	
	% within Respondent education (5 categories)	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	

You can see in the output table above that SPSS has displayed for us the total counts and percentages for each cell. Looking at the “Yes” row at the top of table, which includes all those survey participants who were aware of neighbourhood policing, we can see that 971 respondents who had a level of education coded as “None” answered “Yes” to **neighpol1**.

What percentage of respondents with a degree or diploma were aware of neighbourhood policing?

How many respondents with an O level/GCSE were not aware of neighbourhood policing?

Now, why have we calculated both the actual observed counts and the expected counts for each cell? The difference between the observed count and the expected count for each cell lets us know that there is some relationship between the two variables. Expected values are what we would expect if there was no association between education and awareness of neighbourhood policing. If there is a difference between observed counts and expected counts, then there may be an association between the two variables in question. However, we don’t yet know if this difference is statistically significant. We can determine if there is a statistically significant relationship between these two

categorical variables by running a chi-squared test. A chi square test will determine whether the difference between the observed counts and the expected counts is big enough to say that there is an association in the population.

Summary

You've just run a crosstabulation comparing neighpol1, our categorical dependent variable, with educat3, a categorical variable we've chosen as an independent variable in our analyses. This crosstabulation showed you that there were differences between the observed counts and the expected counts. Can you say anything about the relationship between these variables yet? What do you need to do next?