**Chi-square test + interpretation**

In this video we’re going to looking at running a chi-squared test and interpreting the results when assumptions are not violated. A chi-squared test is a non-parametric test and this means both of the variables we are going to use are not normally distributed, they’re not continuous, they’re not scale. Specifically, they’re nominal variables. So I have here gender and smoking status, and both of these are nominal categorical variables. Now a chi-squared test looks to see if there is an association between these two. You can also express it as a test of independence so for example, a smoking status independent from gender. Now to run a chi-squared test we need to go to Analyse, Descriptives and Crosstabs. Move one variable to the row and one to the column. If you go to the cells button, we’re going to tick on expected count. Now if you want to see more about expected count, have a look at the video on crosstabs. Click ‘continue’, and then click on the statistics button. We’re running a chi-squared test. Now the chi-squared test is only going to tell us if there is a significant association but it doesn’t give us an indication of strength in the relationship, so for that, we’re going to see an affect size, either Phi or Cramer’s V. Click ‘continue’, and then ‘okay’.

Here’s the output for the chi-squared test that we just ran. Our second table here is our crosstabulation table and you can see we have count which is how many we observed in our sample so 10 non-smoking males. We also have something called expected count. This is what we would expect to observe if there was no association between gender and smoking status, so I wold have expected there to be about 7 non-smoking males and about 7 smoking males. About 5 non-smoking females and about 6 smoking females, if there was no relationship between these two variables.

Now the chi-squared test is trying to establish if these two numbers are different enough, to justify saying if there is a significant relationship between gender and smoking status, so we’re going to look down here at our table. Before we read the chi-squared value and p-value, we need to look down at the bottom of the table to see our assumption had been broken, and you see it says 0%. If this value is more than 20%, then that means we violated one of the assumptions and we need to read a different value, but in our case, none of our cells have an expected count less than 5 so these are our expected counts, all of them are bigger than 5, so we’ve not violated any assumptions, so we’re going to go ahead and read of the Pearson chi-squared value which is 6.997 with 1 degree of freedom and our p-value which is under the ‘Asymp. Sig’ column of .008. Now just like with all of the statistical tests, we’re comparing our p-value to our alpha value. Now my alpha value is .05, but make sure to compare yours to your alpha value, which may be .01. Now if this p-value is less than .05 it means our result is statistically significant, so in other words we’re going to reject our H0 and accept our alternate hypothesis to conclude there is a significant relationship between gender and smoking status. Another way to phrase that is to say smoking status is not independent from gender, in other words, it’s dependent upon your gender. So whether you’re male or female will affect whether you’re a smoker or a non-smoker. Now if this value is bigger than your alpha value, then you need to accept your null hypothesis and the null for a chi-squared test says there is no relationship between the two variables. Now because my association is significant, I’m going to go down here to my Cramer’s v table and determine how big of an effect gender has on smoking status, and that’s under the value column. Now if your table is 2x2, which ours is, I’ve got two categories here and two categories at the top, I’m going to raid my 5 value. If your table is bigger than 2x2, so in other words, one of your variables has at least more than 2, so 3 or more, you’re going to read your Cramer’s v value.

Now I recommend that you have a look at your lecture notes, to see what cut off points you should use, or any statistics text books that have been recommended in your course or module. I’m going to use some values that are quoted in Julie Pallant’s SPSS Survival Guide, and a value of .529 would be a large effect, so gender is having a large significant effect on smoking status. If you want to see how to interpret a result where assumptions are broken have a look at my video ‘Chi-squared tests + assumptions violated.’

END.