**Chi-square test + interpretation (assumptions violated)**

Okay, let’s have a look at a chi-squared test when the assumptions have been violated. I’m going to do two examples, one is for a 2x2 table and one is for a table that’s bigger than 2x2. Now the advice I am going to give in this video is from a couple of stats books. One is Andy Field’s Discovering Statistics using SPSS, his recommendation, and the other is by Julie Pallant. Now if your lecturer tells you something different, please do use their advice and what they want you to use, but this is just standard in most texts. Okay, let’s look at our 2x2 example first. If we go to ‘Analyse’, ‘Descriptive Statistics’, then ‘Crosstabs’, just as we did before. We’re going to look at gender and smoking status. Under Statistics, choose Chi-squared and the effects size phi will be the one we want to look at for a 2x2. ‘Continue.’ Under cells we want to tick ‘expected’ because this is how we’re going to check whether or not the assumptions have been violated. Click ‘continue’, and then ‘okay’.

Alright, in our output, here’s our crosstabs table. Now the assumption for a 2x2 table, meaning two categories by two categories, is that all the expected counted are at least ten. Now you can see that two of them here, one is 7.8, the other 8.2, both less than ten which means we violated that assumption. When that happens with a 2x2 table, we go down here and we use Fisher’s exact test. The likelihood ratio is also an option, but with just a 2x2 table, Fisher’s exact test is common to use. I’m just going to the two sided figure here and that’s when your alternate hypothesis is two-tailed. In other words, you don’t specify a direction of the difference. For example, you don’t say there’s more male smokers or more male non-smokers, you just say there is an association between gender and smoking status. Once you determine which significance value is best to use according to your alternate hypothesis, you want to compare it to your alpha just as usual. Now alpha, which is our level of significance, is typically .05. We can see that this is much bigger than .05 which means that a result is not significant and we do accept null hypothesis, also known as H0. Our null hypothesis in a chi-squared test is always, there is not a significant difference or not a significant association between the two variables. In other words, smoking status is completely independent of someone’s gender. So that is the result we would accept in this case because this not significant and when you get a non-significant result you do not look down here at the effect size, because it doesn’t matter, because the result is not significant.

Now let’s have a look at an example with a table bigger than 2x2. Here I have some insurance data, and the two variables we’re going to use are the type of claim someone has filed and whether or not that claimed turned out to be fraudulent. So let’s conduct our chi-squared test as normal, ‘Analyse’, ‘Descriptive statistics’, ‘Crosstabs’. ‘Type of Claim’- because this has more categories I’m going to make it my row variable and fraudulent, only two so its yes or no, so I’m going to make it my column, just so my table is easier to read. Click the ‘Statistics’ button. Tick ‘Chi-squared’, and this time we’re going to use Cramer’s V, ‘continue’, and in cells, I’m going to tick ‘Expected’, but it’s not absolutely necessary this time. Click ‘continue’ and ‘okay’. Here’s my Crosstabs table. It’s a bit bigger this time, it’s bigger than a 2x2, and our assumption for a table bigger than 2x2 is that the expected count is not less than 5, or 20% of the cells have expected count greater than 5. So if we look down here at the bottom its six cells, or 60% have an expected count, less than 5, this violates the assumption because 60% is much bigger than 20%, so I’ll say it again. We want this value to be 20% or less, otherwise the assumption has been violated. If you need to rewind and listen to that again, go ahead. I know it’s a lot to take in in one go. When that’s the case, we’re going to read off the likelihood ratio. Here’s my statistic, my degrees of freedom and my significance value. Again, I’m going to compare it to my level of significant, .05. This is much bigger, which means I accept my null, and I conclude there is no association between the type of claim and whether or not the type of claim is fraudulent.

END.