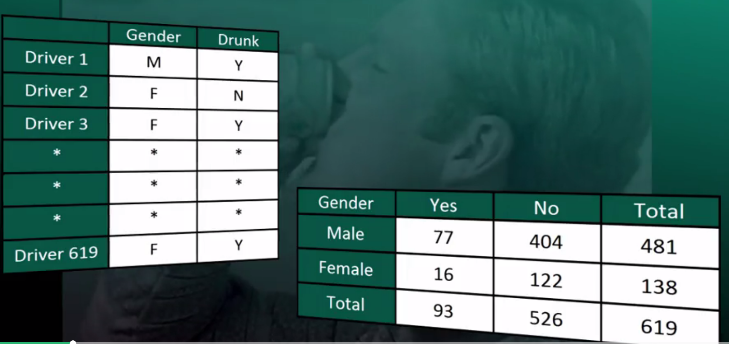
C2 Week 2 Notes

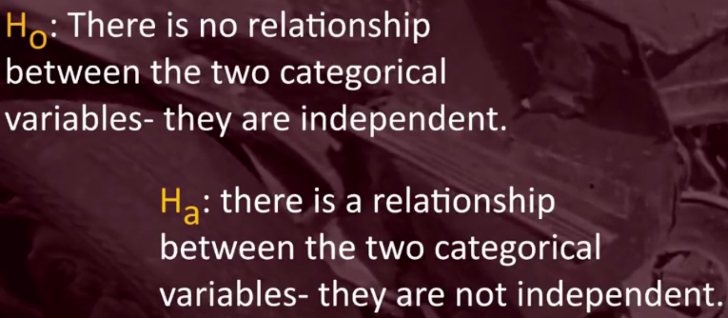
**Chi Square teste of independence**

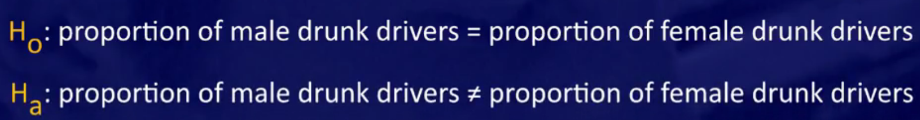
V1

\*Ideas behind Chi Square test of independence

Chi Square Test of Independence - inferences about the relationships between two categorical variables.

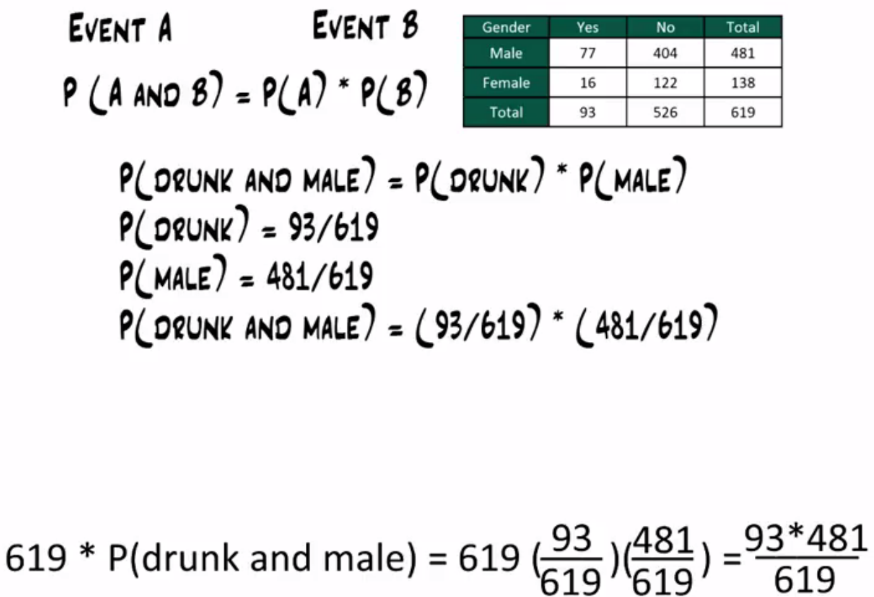


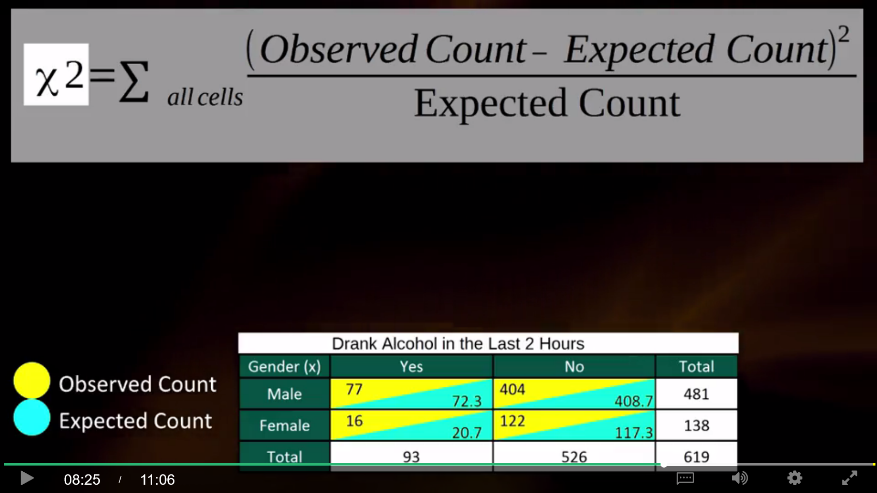


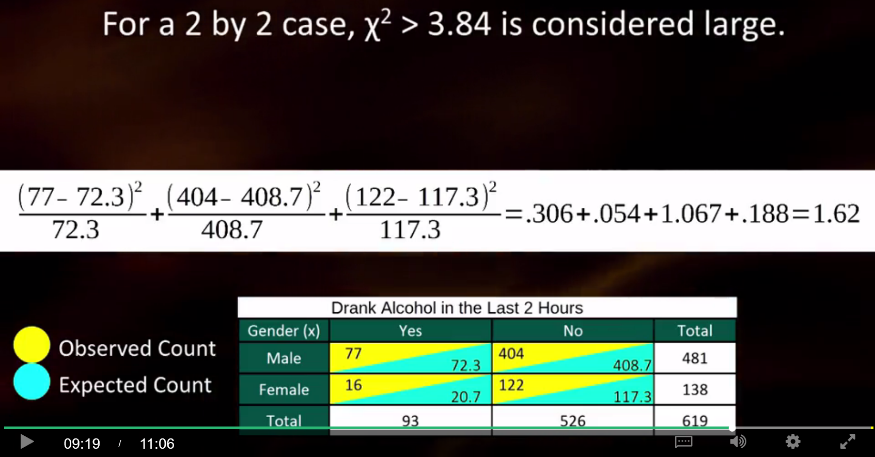


The idea behind the chi-square test of independence, much like the analysis of variance is to measure how far the data are from what is claimed in the null hypothesis. The further the data are from the null hypothesis, the more evidence the data presents against it.

If the null hypothesis were true, how many female drunk drivers would we expect to see instead of 16? How many non-drunk driving males would we expect to see instead of 404? How many non-drunk driving females would we expect to see instead of 122? >> In other words, we will have two sets of counts. The Observed Counts, that is the data. And the Expected Counts, if the null hypothesis were true. We will measure how far the observed counts are from the expected ones. We will base our decision on the size of the discrepancy between what we observe and what we would expect to observe, if the null hypothesis were true.







It turns out that for two by two case like ours, we're inclined to call the chi-square statistic large if it's larger than 3.84. Therefore, our test statistic is not large indicating that the data are not different enough from the null hypothesis for us to reject it.

For cases other than a two by two, there are different cutoffs for what's considered large, which are determined by the null distribution in that case.

Even though we cannot really use the chi-square statistic, it was important to learn about it, since it encompasses the idea behind the test. The p value for the chi squared test of independence is the probability of getting counts like those observed, assuming that the two variables are not related. Which is what is claimed by the null hypothesis. The smaller the p value, the more surprising it would be to get counts like we did, if the null hypothesis were true. Technically, the p value is the probability of observing a chi-square at least as large as the one observed.

**Using our statistical software, we'll find that the p value for this test is 0.201. The p value of 0.201 is not small at all. There's no compelling statistical evidence to reject the null hypothesis.** And so, we'll continue to assume it may be true. Gender and drunk driving may be independent.