**Transcript: Conducting Independent Samples *t* Tests in R Commander**

In this video we are going to go through how to conduct an independent samples *t* test in R Commander. We have loaded our data set which is called “Wells” and it is found in the “carData” package. We can look here at the variables within this data set, as well as the first few observations for each. This data set was drawn from a study in an area of Bangladesh where a lot of families have unsafe drinking water. All of the families in this data set were using wells that had an unsafe level of arsenic, and researchers recommended that they switch to a safer well. We have here some information about the wells that they were using prior to this study, so the well that was unsafe and had unsafe levels of arsenic. This “arsenic” variable tells you how much arsenic there was in their previous well. We also have “distance” which refers to distance to a safe well. “Education”, this is the years of education that the head of the household has, and “association”, which refers to whether or not any members of the household were part of a community association, and lastly, we have the “switch” variable which indicates whether or not the families made the switch to a safer well.

What we're going to look at today is whether or not there are significant differences in education level between families who did or did not switch wells. We might hypothesize that people with higher levels of education would be more likely to understand what the researchers were saying about unsafe levels of arsenic, and they also might have more trust in the researchers. And so, we might hypothesize that people who switched drinking wells, oh sorry, families who switched drinking wells, would have heads of households with higher levels of education compared to families that didn't switch drinking wells.

The first thing we want to do is just look at the descriptive statistics for each group, so we can closeout our data set here and look at the descriptive statistics. In the earlier labs we have looked at descriptive statistics for variables, but we've never broken it down by group before, so I'm going to show you how to do that in this video. We will navigate to “Statistics / Summaries” and “Numerical summaries”. This is the same window that we use to calculate descriptive statistics more generally. We can select our variable that we want descriptive statistics for—in this case, “education”. And the thing that we're going to do differently is click this “Summarize by”. This will bring up our grouping variables or categorical variables and we can select one of these to describe our “education” variable by. What this means is that we're going to see the mean and standard deviation and other descriptive statistics within each group, rather than for the sample as a whole. I'll select “switch” here and click “OK”. And once I click “OK” here, this will produce descriptive statistics for each group.

You can see here that under the “no” group—so this is people who did not switch wells—the mean level of education was 4.47 years. The standard deviation is 3.76 years. And we can also see the different percentiles as well as the number of people that fell within that group. Within the “yes” group, the mean level of education is 5.09 years. The standard deviation of 4.18 and again you can see the IQR and the different percentiles. Right off the bat we can see that the “yes” group does have a higher level of education, so 5.09 years compared to 4.47. But in order to see whether this is a significant difference, we need to conduct an independent samples *t* test.

To do that, we will navigate to “Statistics / Means” and “Independent samples *t* test”. Within this window we first need to pick our “Group”. This is the group that we're comparing our outcome variable between two levels of the group. OK, so we will select “switch” here. Our “Response Variable” is the continuous variable that we are comparing the level of between the two groups. We will select “education” here because we are comparing the level of education for the head of household between households that switched wells and households that did not.

Next, we're going to go into “Options” and we can look here at what our alternative hypothesis is. As we discussed in the one-sample *t* test lab, we can pick a two-sided hypothesis, which is testing whether the difference is significantly different from zero on either side, or we can select a one-sided test. In this case, our hypothesis is that families who switched wells would have heads of households that were more educated compared to families that did not switch wells. You can see here our difference is no compared to yes, so we hypothesize that the “no” group is going to have lower education compared to the “yes” group. Again, in plainer language, this is indicating that people, families who did not switch wells would have heads of households that were less educated compared to families that did switch wells. This corresponds to this option, “Difference < 0”. We also can see here this “Assume equal variances”. We want to make sure that “No” is selected.

Once we click “OK”, R Commander will run our independent samples *t* test. And we will be talking in a future video about how to interpret this output.