**Transcript: Resistant Indicators in SPSS**

This video will cover how to calculate the interquartile range and how to create a boxplot in SPSS. Here I've loaded in our data for this week, which is the SLID data set containing information about wages in Ontario. You can see our variables here. The one we're going to be looking at today is “wages”.

As you can see, “wages” has been read in as a string variable. When it really should be a numeric variable. This is because wages are reported in dollars per hour in this data set, which are numbers measured on a numerical scale, so it should be a numeric variable. If we go into “Data view”, you can see the reason why wages has been written as a string variable instead of numeric is because we have these “NA”s, which are letters and so they're being read in as a string. But really, this “NA” just means that there is a missing value, so we need to change this variable to a numeric variable by hand. To do that, we can go back to “Variable view”. Here where it says “String”, we can click on these three dots to the right and change this to “Numeric”. Once we click “OK” it will change to a numeric variable. We still have to go into this “Measure” and change this to “Scale”, so if we just click on that we can select “Scale”, and then we have our “wages” variable as a numeric variable. If we go back to our data view, we can see that these “NA” values have been replaced by dots, which is how SPSS represents missing values.

So now we're ready to look at our resistant indicators of interquartile range and boxplots. To do that, we will navigate to “Analyze”, “Descriptive Statistics”, and “Explore...”. This will pull up a list of variables and we can select which variable we want to look at. In this case, we're going to select “wages” and use this arrow to move it over to the field “Dependent List”. The “dependent” in this “Dependent List” field means that that's the variable that we're examining. This “Factor List” is another option that you have where you could look at, say, wages within sex. So within male and female participants you could look at wages separately but we're not going to do that today.

Next, we're going to click on “Statistics” and we're going to select “Descriptives”, which is already selected and press “Continue”. If we go into “Plots”, you can see that we can select which types of plots we want. We want to create a boxplot. We can leave this default option selected, and we don't need a stem and leaf display, so then we can click “Continue” and “OK”.

This will give us some basic descriptive statistics of the “wages” data set. We have the mean, the median and the variance, which you've seen previously in this data set and we also have here reported the interquartile range. This is the difference between the 75th percentile and the 25th percentile and it gives us a much better indication of the spread of the values than the range, which is 48.

If we look at our boxplot here, we can see why the interquartile range and the range are so different, which is that we have a lot of high outliers. We saw this in the R Commander video and we're seeing the same thing here, that the majority of people are making low to middle income wages with a small number making wages in the higher range.

SPSS has already created a boxplot for us, but we need to do a little bit of work to make it look prettier and have the axis labels and the titles that we want to see. To edit a plot, you right click on a PC or command, sorry control click on a Mac and select edit.

The first thing we want to do is add a y-axis label so we know what these units are and what they represent. To do that, we will select this “Y” button. Navigate to “Labels & Ticks” and check “Display axis title”. Once we select “Apply”, that will allow us to edit our y-axis. Once we click on that, we can fill in the variable, which is “wages”, and the units, which is dollars per hour.

We can delete this x-axis label as we discussed in the R Commander video. If you click on that, you can just backspace all the way through and you won't have an x-axis label anymore.

The last thing we need to do is add a title to our graph. This is represented by this button here, which is called “Insert a title”. We can write here our title, which should be descriptive and tell someone who's not familiar with the data, what they're seeing. In this case, “Distribution of Wages” is a good title. Once we're done with that, we can close out this window and our plot will update in our output field.

So that gives you a nice boxplot with the outliers displayed as well as the median, the 75th percentile, the 25th percentile and again, we can go up here if we want to see our interquartile range.