**Select random sample of data using 'Select Cases' tool**

In the last example we looked at selecting specific cases of data. This time we’re going to look at taking a random sample. So say for example you have a huge database with thousands and thousands or cases of participants and you just want to select a random sample of the data, you can do that using the select cases tool. So if you go back to data, select ‘cases’ and then instead of choosing the ‘If condition’ like we did last time, we’re going to choose either the first or the second so random sample of cases, click on this and then click ‘sample’. I can then choose if I want, say 50% of all the cases that would- so if I’ve got 70 in my data set now it would randomly select 35 of them, or I can specify exactly how many I want. So say for example, I want to take exactly 50 from the first 70 cases so I only have 70 in my data set, so if you’ve got thousands and thousands you may want to only select a hundred from the first thousand to that effect, so you can choose how you and only select your cases. I’m going to back here and choose approximately 50%, click ‘continue’ and then our output options- ‘filter out unselected cases’ or we could ‘copy selected cases to a new data set or delete them. I said before I don’t like the delete option, I like to keep my original data set and either use filter out or copy. So say if I had a huge data set, I might want to copy them to a new data set, so I’m going to give it a name and call them Random Sample, and go ahead and click ‘okay.’

Okay so if we open up our new data set, I can see that it’s called Random Sample, my data file is called Untitled Two, so I will need to name the file but my data set has been named, and if I scroll down here I’ve got 41 cases. Now you might wonder why I’ve got 41 when I asked for 50%. Well its approximately 50% so it’s around half of the data set, I should expect around 35 so in this case it has selected 41 and I can see here by the id column it has taken a random sample from my original data set. If we go back to the original one, select ‘cases’ again. This time, I’m going to choose based on time or case range, so if you click on ‘range’, I can choose to select a range of cases. So say I want observations 10 through 20, click ‘continue’, and this time I’m just going to filter out on selected cases, click ‘okay’ and if we have a look its only cases 10 to 20 who have actually been selected. Now in this one it’s not really random selection but it is a different select case option.

END.