**Create dummy variables from an existing categorical variable in SPSS**

In this video I’m going to show you how to create dummy variables. Now dummy variables are useful when we have a nominal variable that we want to use in something like regression. I’m going to use a variable called ‘ethnic group’. This variable is nominal because I can’t rank the ethnicities in any particular order. If we go to the variable view, we can see what our current coding is. I’ve got five categories: White European, Asian, West Indian, African, and other. Now dummy variables can be a bit confusing sometimes. I highly recommend, and if you haven’t already looked into what a dummy variable is, that you have a look either online or get a good stats textbook and try to get to grips with what it is. Basically, we’re creating dichotomist variables, meaning two groups, zeros and ones, for each of our categories. Now, it turns out that if we have five categories, we’re always going to have four dummy variables, so the number of dummy variables should always be one less than the number of categories that you currently have. I don’t have a whole lot of time you go into why, so please do look at that and see why that is. So I’ve got five categories, and I’m going to make four dummy variables. So if we go to the transform menu, I want to use the recode into different variables option. I don’t want to overwrite my current data; I want to create new variables, each one with a zero-one coding. Hopefully it will make a bit more sense once we do an example.

So let’s take ethnic group and move it to our variables box. Now the first one I’m going to code is White European. We can also add a label if we want to. And click ‘change’ when you’re done. Now we’re going to define what the old and new values are. So White European, the current coding for that category is a one. So I’m going to type that into my old value box and the new value. Dummy values, like I said, are always zero and one, so one is given the category name which is White European and zero is given to everything else so I’m going to put a one here. Click ‘add’, and everything else, so that means all other categories I want to code as zero because they’re not White European, they’re something else. So click ‘continue’ and then ‘okay.’ If we go to our data file, we can see we’ve got a new variable called White European, and it’s got a label which you put in. We want to change the decimal places to zero, because we don’t need them, and we want to put in some value labels. So one represents White European, and zero is everything else, so we’re going to call that ‘other.’ So basically, they’re not White European. Click ‘okay’ and make sure that your measurement is at nominal because we can’t order these zeros and ones, there’s no meaningful rank to them. They’re just nominal. If we have a look at our data view and go all the way to the end we can see we’ve got a new variable here with our value labels showing and our codes are zero and one.

So let’s do one more example. We’re going to use Asian. If we go to transform, recode into different variables, go ahead and click ‘reset’, and move our ethnic group variable in. This one’s called Asian. I’m going to put the same label in. Click ‘change’, and again go to old and new values. Now Asian is coded as a two, so I’m going to enter two as my old value. And the new value, just like before is one. So we’re always coding them to be zero and one, no matter what the original coding is. Click ‘add’ and then again, everything else, so all other values, I want to be zero. So every other ethnic group is going to be ‘other.’ Click ‘continue’ and then ‘okay’. If we scroll here to the end we’ve got Asian, and then we just need to define it, give it some value labels. So from the variable view, I’m going to change decimal places to zero, and then I’m going to add some value labels. So one represents Asian, and zero represents ‘other’. Click ‘okay’, and then again just make sure your level of measurement is correct. So if we go to the data view, we can see White European. There should be ‘other’ here because we’ve only given them option to one ethnic group, and we’ve got Asian and ‘other’ here, and White European and ‘other’ here. So we would continue doing this for two more ethnic groups, West Indian and African because those are our third and fourth groups. And we would not create a dummy variable for the fifth one. So we always have one less dummy variable than we do categories. And that is because, by default, if a participant is not White European, is not Asian, is not West Indian, is not African, it means that they are ‘other’, so we don’t need to have that fifth category, or dummy variable. Again, if you’re unsure about what dummy variables are , what they’re used for, please do have a look and see what they’re for in either a textbook or have a search online.

END.