

# SPSS Lesson 2

## Variable Transformations

Although in most cases the variables we code are the variables we use in our analyses, sometimes we need to do something to them in order to make them more usable for us. There are three variable transformations we are interested in: (1) COMPUTEs; (2) RECODEs; and (3) COUNTs.

### COMPUTE

We use COMPUTE when we want SPSS to take our raw scores and apply a formula to them in order to compute a new variable. For example, if our survey included Likert-type items, we might want to compute a “total score” on our attitude measure. For another example, in the QuickSurvey database used in tutorial, one of the questions asked the person to specify the year in which they were born. This is a good way to ask the question, particularly when dealing with older samples who often will balk at telling you their age but are willing to tell you what year they were born, but normally we ask this sort of question because we are interested in their age, *not* the year of their birth. In order to compute a new variable called AGE from the YEARBORN variable we coded, you would do the following:

- Start SPSS, open up the QuickSurvey file, and make sure it is on “data view.”
- Click TRANSFORM on the menu at the top of the page, and a drop down menu will appear.
- Click COMPUTE on the drop down menu and a dialogue box will appear.
- In the space that says TARGET VARIABLE, supply a name for the new variable you want to create via the COMPUTE statement.
- Note that as soon as you start typing a name for the target variable, the “TYPE & LABEL” button lit up. You can use that to create a variable label. Just click on the button and write in the label. [Note you can also do this later through the “Variable view” page.
- To the right of the TARGET VARIABLE box is an equals (=) sign, and to the right of that is another box. It is in that box to the right where we write the formula that will change YEARBORN (in this case) to AGE. So ... the first thing we need to decide is what the formula is. Because the question asked respondents to specify only the last two digits of their year of birth, and that is what we coded, and given that this year is 2002, note that the formula we need is that  $AGE = 102 - YEARBORN$ . For example, if a person is born in 1970, so that the number we coded for them was 70, then in 2002 they should be 32 years old. When we subject them to the formula we find  $102 - YEARBORN = 102 - 70 = 32$ . Good.
- So ... first place 102 in the “NUMERIC EXPRESSION” box to the right; then click on the minus (-) sign; then bring YEARBORN into the equation by first clicking

on it in the list of variables at left, and then clicking on the little arrow (▶) close to the right box.

- After you are satisfied the formula is OK, click “OK” and when you go to the extreme right hand side of your data matrix, you will see that a new variable called AGE has been created.
- When engaging in more complex compute statements, note that SPSS follows the algebraic rules of precedence you probably learned in Grade 10, i.e., expression in parentheses are computed before those outside of them; multiplications and divisions are done before additions and subtractions, and so on.
- When engaging in more complex calculations, it is often a good idea to create some “dummy” cases where you know what the person’s score should be IF you get the formula right. For example, when I use SPSS to compute your final grades on the basis of all the course components and using the weighting scheme specified in the course syllabus, before running the COMPUTE I will add in some fictitious students – one who gets 100% on everything, for example, and another who gets 50% on everything, and then check to see whether they end up with 100% and 50%, respectively, as final scores. If they do, then my formula is correct; if they don’t, then I made a mistake somewhere and need to try again.

## RECODE

There are many situations where we want to change the way we coded our original variables, although most frequently this occurs when we want to collapse down a larger number of categories into a smaller set. For example, sometimes we may have used a 7-category or 5-category Agree-Disagree scale for our Likert items, but decide in the end that we want to simplify things and scrunch those 5 (or 7) categories down into 3 (e.g., agree, neutral, disagree). Or we might want to take the 20 different AGEs we have and separate them into three groups – older, average, and younger respondents – or the 15 different INCOME categories we have and simplify it down to poorer, average, and wealthier respondents. In the example below, I will take one of the Likert questions from the QuickSurvey database – which used 5 response categories – and recode it into three groups (disagree, neutral, agree).

- Start SPSS, open up the QuickSurvey file, and make sure it is on “data view.”
- Click TRANSFORM on the menu at the top of the page, and a drop down menu will appear.
- Click RECODE on the drop down menu and you will be asked whether you want to recode “into same variables” or “into different variables.” If you pick “into same variables” your original variable and coding will disappear. Even if this is our intention, it is generally better to choose “into different variables” so that you can do the transformation and make sure you get it right, because you can always destroy the “old” variable if you find you do not need it any more.
- The RECODE INTO DIFFERENT VARIABLES dialogue box opens, and on the left you can see the list of variables. Scroll down to the one with the label “Immigration levels are too high” [IMMIGRNT]. Click on it, and then click on the arrow (▶) to move it into the box that says INPUT VARIABLE → OUTPUT VARIABLE. Note that as soon as you do so, “INPUT VARIABLE” turns into

“NUMERIC VARIABLE.” Nothing of any great significance to that; just don’t be surprised when you see it.

- So we’ve told SPSS that the variable we want to recode is the one called IMMIGRNT, but now it wants to know the name of the new variable we want to create. We do this by writing the name of the new variable in the box labelled “OUTPUT VARIABLE” ; note that we can also insert its variable label there as well. Let’s say we call the new variable IMM.x.3 – note that I’m using the IMM stem from IMMIGRNT but putting .x.3 to convey to myself that it’s the 3-category version of it – and I’ll add the Variable Label “Immigration Levels Too High.”
- Now that we’ve told SPSS the old and new variables, the next step is to tell it the translation formula, i.e., what the old and new values will be, i.e., how we want it to recode the variable. We start this by clicking on the OLD AND NEW VALUES button in the RECODE INTO DIFFERENT VARIABLES dialogue box.
- When we click OLD AND NEW VALUES, a new dialogue box opens. Note that the left hand side is where we specify the old values, while the right hand side is where we specify the new ones. We want it to change as follows:

Old Codes	New Codes
1 = Disagree Strongly	1 = Disagree
2 = Disagree Somewhat	
3 = Neutral / Don’t know	2 = Neutral / Don’t know
4 = Agree Somewhat	3 = Agree
5 = Agree Strongly	

- We want to collapse categories 1 and 2 into a new Category 1 – Disagree. The Neutral/Don’t Know Category will stay on its own but will now become Category 2. And the former categories 4 and 5 will now be collapsed down into a single category called “Agree,” which we will code as 3. Let’s do this one at a time:
- On the left hand side, click your cursor on the circle at the start of the line that says RANGE \_\_\_\_ THROUGH \_\_\_\_\_. Place “1” and “2” in those boxes so that it reads RANGE 1 THROUGH 2. On the right hand side, click the circle beside VALUE up at the top, and place the value “1” in the box there. Then click ADD to register that change in the OLD  $\rightarrow$  NEW box below that. You’ll see “1 thru 2  $\rightarrow$  1” appear there, which shows you that the former 1 and 2 will be recoded into code 1.
- Go back to the left hand (OLD VALUE) side again and click the circle by the word VALUE at the top. Place “3” in that box (the code formerly associated with Neutral/Don’t Know). Then go over to the right hand (NEW VALUE) side, and beside where it says VALUE, insert “2.” Then click ADD. A new line gets added in the OLD  $\rightarrow$  NEW box, i.e., 3  $\rightarrow$  2, i.e., everyone in the former code 3 will now be given code 2.
- You should now know how to do the last part of the change. The Old values are 4 and 5, and the new value we want those groups coded to is 3.
- After you’ve completed that, click CONTINUE, and you will be returned to the RECODE INTO DIFFERENT VARIABLES dialogue box. We’re all ready now to create the newly recoded variable, so first click CHANGE, then OK.

- If you look to the right end of your data matrix, you will see that a new variable has been created – IMM.x.3 – that has levels 1, 2 and 3, just as we wanted.
- You can go a step further in defining this new variable by clicking on VARIABLE VIEW at the bottom left hand corner of the data matrix, and going to the new IMM.x.3 variable we created (which you'll find at the bottom of the list), and then go to the VALUES column where we can enter in labels for our new codes, i.e., 1 = Disagree; 2 = Neutral / Don't Know; 3 = Agree.

## COUNT

A final transformation you need to know is the COUNT transformation. Whenever you activate COUNT, you need to specify what you want SPSS to look for, and then it will count how many of these there are, and report back to you.

- For example, recall that in the QuickSurvey, one of the questions asked respondents to check all the different activities they had participated in in the last month. Seven different activities were included – team sport, individual sport, gone to a movie, attended a sporting event as a spectator, gone to a pub or bar with friends, relaxed in front of the TV, and/or listened to music.
- Because it was a “check all that apply” question, we had to code each alternative individually, which we did by creating 7 different variables, and assigning a “0” if they did *not* check it, and a “1” if they did.
- If we were now interested in seeing what sort of diversity of activities the students were engaged in, we could create a new variable called ACTIVITY by getting SPSS to go through all the codes and counting how many 1s the person had across the 7 different activities.
- To do so, start off by going to the menu at the top of the matrix and clicking on TRANSFORMATION, and then sliding down the dropdown menu and clicking COUNT.
- A new dialogue box opens that is entitled COUNT OCCURRENCES OF VALUES WITHIN CASES [SPSS doesn't exactly have the snappiest titles, but at least they're descriptive.] We are asked to specify the new TARGET VARIABLE (wherein we will place the new variable, ACTIVITY), and a new Variable Label (where we could say something like “Number of Different Activities Engaged In.”
- Note the list of variables at the left of that dialogue box. The next thing we have to do is find all the variables (activities) we want SPSS to go through and count. One by one, we click these into the VARIABLES box. [You can also do all 7 at once by clicking the first variable – “Played team-base sport [XTEAM]” – and then hitting the down arrow key on your keyboard while holding the shift key down until all 7 variables we want are highlighted.
- Then click the DEFINE VALUES button. A new dialogue box will open wherein SPSS wants us to specify what values we want it to count. At the top left hand corner of the box where it says VALUE, place a “1” (because 1 is the value we want it to count). Then click ADD over on the right hand side, at which point a “1” will appear in the large box at right.

- Finally, click CONTINUE, which will take you back to the earlier dialogue box, and then OK, and you're done. If you look at the right end of your data matrix, you will see that a new variable called ACTIVITY has been created, which has values that fall from 0 through to 7, which we can then enter in any other sort of analysis we are interested in (e.g., to determine whether men or women are involved in a broader range of activities).