Dummy Variables

In Chapter 5, we discussed dummy variables (often referred to as "dummies"). Dummies indicate if something is present or not, indicated by the values 1 (present) or 0 (not present). If, for example, advertising is used, a dummy variable takes on the value 1. If no advertising is used, the dummy variable takes on the value 0. Such dummies can be used for many purposes.

Dummies can, for example, be used for descriptive purposes where we can plot the use of advertising against time, indicating patterns of when advertising is used. Dummy variables are often used in regression analysis (discussed in Chapter 7 of the book), but also in other statistical techniques, such as t-tests and as part of cluster analysis. There are several types of dummy variables:

- Dummy variables that are naturally coded as 1 or 0. For example, variables such as gender are often coded as 1 (e.g., for females) and 0 (for males). There is no need to explicitly create dummies because these variables already only take on values of 0 and 1.
- Dummy variables created by splitting a nominal variable into multiple levels. For example, if we have a nominal variable, such as *type of shop*, we can create multiple dummies for different levels. For example, we can define a nominal variable *shop type* as having three different levels; department stores, supermarkets, and discount stores. The number of levels are the different values a variable takes on. We can indicate these three different levels by creating two (not three!) dummies. A first dummy can be created to indicate department stores (1 for department stores, 0 for supermarkets and department stores). A second dummy can indicate supermarkets (1 for supermarkets, 0 for department stores and discount stores). A third dummy should not be created because if we already know a store is not a department store or supermarket, it is a discount store. Therefore, always create one dummy less than the number of categories or levels! Thus for 4 levels, you only need to create 3 dummies. An issue to consider is the level for which we do not create a dummy. The level in which you are the least interested research-wise would be best for this.

– Dummy variables can also be used to split up ordinal, interval, or ratio scaled variables. For example, if we have a ratio scaled variable called *age* measuring the age of customers, we can define customers as underage (younger than 18 years) or mature (18 or older). In doing so, we create two categories. We could, of course, also split age into multiple categories. Again, create one dummy less than the number of categories. If you have five categories, only four dummies are needed.

How can we create dummies in SPSS? We use a dataset called *retailer.sav* (Web Appendix → Chapter 5) for the subsequent examples and illustrations. This dataset contains information on how participants feel about the Internet and how they experienced their last Internet session. Moreover, it contains the participant's age, gender, income, and the size of city in which he/she lives.

This dataset contains one naturally coded dummy, *Gender*. This variable takes on the value of zero for males and 1 for females. We can create dummies for the other variables from the original variables provided. For example, when taking the first variable, *Enjoy1*, we have seven different levels, 0 for completely disagree, 1 for disagree, 2 for somewhat disagree, 3 for neutral, 4 for somewhat agree, 5 for agree, and 6 for completely agree. To compare the seven responses, we should create only six dummies. Imagine we are least interested in those that completely agree, making that our base category and level for which we do not create a dummy. Thus we should create dummies as follows:

Values variable can take on	Which dummies to create
Completely disagree	Create dummy with value=1 if respondent completely disagrees and the value of 0 else
Disagree	Create dummy with value=1 if respondent disagrees and the value of 0 else
Somewhat disagree	Create dummy with value=1 if respondent somewhat disagrees and the value of 0 else
Neutral;	Create dummy with value=1 if respondent is neutral and the value of 0 else
Somewhat agree	Create dummy with value=1 if respondent somewhat agrees and the value of 0 else
Agree	Create dummy with value=1 if respondent agrees and the value of 0 else
Completely agree	Do not create a dummy for this category

Using SPSS, we should therefore create five dummies. To do so, we need SPSS's *Recode into Different Variables menu* included under ▸ Transform ▸ Recode into Different Variables.

In this box (Figure A7.1) you can indicate for which variable you want to create dummies. The original variable should be moved from the left-hand box to the center box title **Numeric → Variable Output Variable**. The name of the dummy

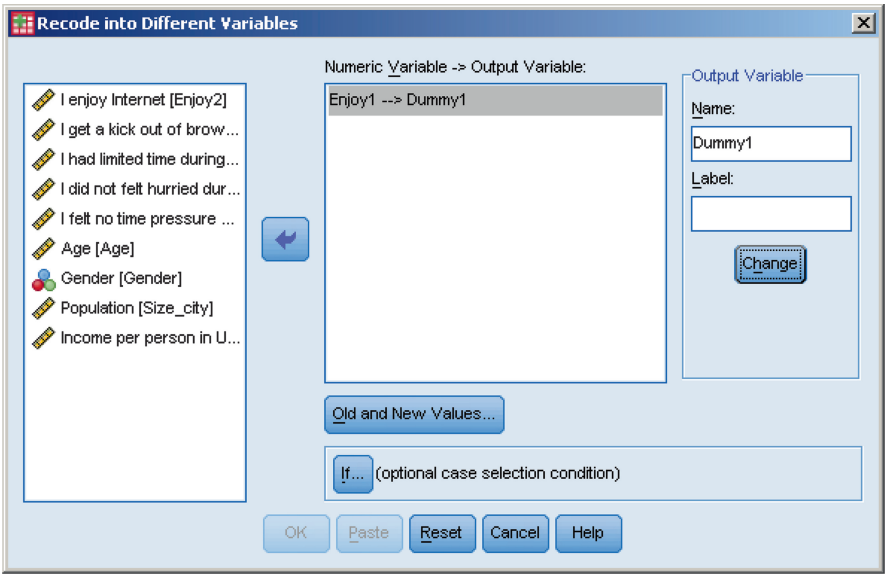

Figure A7.1 Recode into Different Variables.

variable (for example *Dummy1*) should be typed under **Output Variable**. Then click on **Change** first and then click on **Old and New Values**. This will open up a new dialog box shown in Figure A7.2.

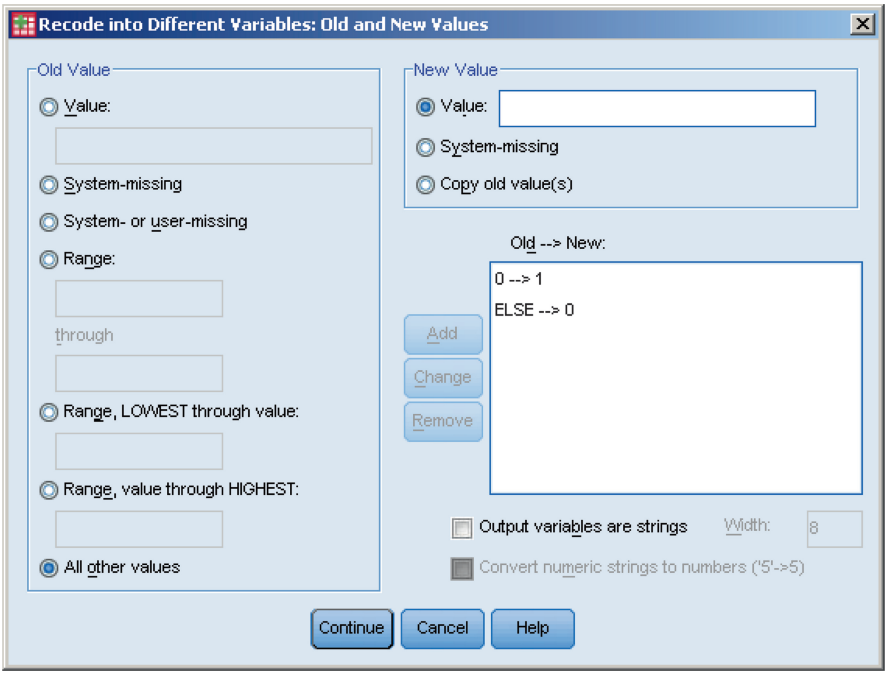

Figure A7.2 Recode into Different Variables: Old and New Values.

In this dialog box, you should indicate that if a respondent completely disagrees, this is given the value of 1 and in all other cases the value of 0. Do this by indicating that if the value of Enjoy1 is 0, this should be given the value of 1 and 0 else (under **Old** - -> **New**). Then click on **Continue** and **OK**. Repeat this for values one through five of Enjoy1.

Similarly, you can create dummies from interval or ratio-scaled variables such as Age. For example, you could create a dummy that defines customers as underage (younger than 18 years) or mature (18 or older). This can be done as described above, except that you should indicate the value of 18 in the box under **Range, LOWEST through value**, and indicate that this should be given the value of 0 under **New Value**. Then click on **All other values** and indicate under **New Value** again that this should get the value of 1.