**MODIFYING DATA**

## 1.0 SPSS commands used in this unit

|  |  |
| --- | --- |
| **save outfile** | saves the data file |
| **display** | displays attributes of the data set |
| **variable labels** | labels a variable |
| **value labels** | adds labels to values of a variable |
| **autorecode** | recodes variables and automatically adds value labels |
| **rename variables** | renames variables |
| **recode** | recodes variables |
| **document** | adds a document to the data set |
| **compute** | creates new numeric variables |
| **summarize** | calculates descriptive statistics |
| **aggregate** | creates new variables with aggregated data |

## 2.0 Demonstration and explanation

Let’s begin by opening the data file.

|  |  |
| --- | --- |
| * **File** * **Open** * select the **C:** drive, the **temp** folder, and **hs0.sav** | \* open the data file.  **get file "c:\temp\hs0.sav".** |

It is often useful to see information regarding the data file, such as the number of cases and variables, any type of labels, etc.

|  |  |
| --- | --- |
| * **File** * **Display Data File Information** * **Working file** * select **c:\temp\hs0.sav** | \* using display dictionary to view the properties  \* of the data set.  \* Because we have not listed any variables  \* after the command, SPSS will show us the  \* codebook for all of the variables.  **display dictionary.** |

## 2.1  Reordering variables

Reordering variables in the data file is helpful both for organizational reasons as well as to minimize the amount of scrolling you need to do in order to see the variables that you are working with. Let’s make id the first variable.

|  |  |
| --- | --- |
| * Click on "Variable View" tab * Hover mouse pointer over blank column to the left of Name * Click and hold the second row (id) * Drag up to position id in the first row   + gender should now appear in the second row | \* ordering the variables in a way that  \* makes sense.  **save outfile = "c:\temp\hs01.sav"**  **/ keep id gender all.**  **get file "c:\temp\hs01.sav".**  **display variables.** |

## 2.2  Adding variable and value labels

Adding variable labels is a very useful data management strategy, and we encourage you to take the time to do this when you input a data set or receive a data file.

|  |  |
| --- | --- |
| * **Click** on "Variable View" tab * (in the lower left corner)for **schtyp**, * type in the label "The type of school the * student attended." in the Label column. * In the Values column, click on the right of * the box for **schtyp** to open the dialog box * type 1 in the Value box and "public" in the * Label box, and then click on Add. Do the * for the next value label. | \* adding variable and value labels to schtyp.  **variable labels schtyp "the type of school the student attended.".**  **value labels schtyp 1 "public" 2 "private".**  **display dictionary**  **/var = schtyp.**  **list schtyp**  **/cases from 1 to 10.** |

## 2.3  Changing a string variable to a numeric variable

If we click on the “Variable View” tab, we can see that the variable **prgtype** is a string variable, and this may cause some difficulty when we are using this variable in analyses, so let’s create a numeric version of this variable.

|  |  |
| --- | --- |
| * **Transform** * **Automatic recode...** * select **prgtype** * type in a name for the new variable * (**prog**) and **click** on the "Add New Name" * Button | \* changing prgtype from a string  \* to a numeric variable (called prog).  **Autorecod variables = prgtype**  **/into prog**  **/print.** |

We will add a variable label to the variable that we just created.

|  |  |
| --- | --- |
| * Add a variable label to **prog**. | \* adding the variable label.  **variable labels prog "The type of program**  **in which the student was enrolled.".** |

## 2.4  Renaming variables

Renaming variables is easy.  We can rename the variable **gender** to **female**, and then we will add a variable label and values labels.

|  |  |
| --- | --- |
| * **Click** on "Variable View" tab * (in the lower left corner). Change **gender** to **female**, * type in the label "The gender of the student.", and add * the value labels. | \* renaming the variable gender to female and adding  \* a variable label and value labels.  **rename variables (gender = female).**  **variable labels female "The gender of the student.".**  **value labels female 1 "female" 0 "male".**  **display dictionary**  **/var = female.**  **list female**  **/cases from 1 to 10.** |

## 2.5  Recoding values

Suppose that we would like to recode some values of a variable.  For example, we might want to change the 5s to missing.  If you like, you can use the **frequencies** command before and after the recoding to see the changes. You may also want to include some reminders of this change.  We can create a document for this purpose.

|  |  |
| --- | --- |
| * **Transform** * **Recode** * **Into same variable...** * select **race** * **click** "Old and New Values" and type in the old value (5) and the * new value, in this case, click on the "System-missing" radio button, * and then **click** on **"**Add**"**, then **"**Continue**"**, then **"**OK**"** * **Utilities** * **Data File Comments** * **Type** "The variable gender was renamed to female. * Values of race coded as 5 were recoded to be missing." | \* recoding race = 5 to missing.  **frequencies var = race.**  **recode race (5 = sysmis).**  **frequencies var = race.**  \* adding notes to the data set and viewing the notes.  **document The variable gender was renamed to female.**  **document Values of race coded as 5 were recoded to be missing.**  **display document.** |

## 2.6  Creating a new variable

There are many ways that you can create a new variable.  One way is to use a numeric expression.  For example, let’s create a variable called **total** that will be the sum of the reading, writing and math scores.

|  |  |
| --- | --- |
| * **Transform** * **Compute...** * type in the name of the new variable, **total**, * (called the "Target Variable") * and the numeric expression that will create the variable: * read + write + math * **Analyze** * **Descriptive Statistics** * **Descriptives...** * select **total** | \* creating a variable that is a total  \* of some of the test scores.  **compute total = read + write + math.**  **summarize var = total.** |

It might make more sense to add the social studies score to the total rather than the math score, so let’s change that.

|  |  |
| --- | --- |
| * **Transform** * **Compute...** * type in the name of the new variable, **total**, * (called the "Target Variable") * and the numeric expression that will create * the variable: read + write + socst | \* creating a variable that is a total  \* of the reading writing and social  \* studies test scores.  **compute total = read + write + socst.**  **variable labels total "the total of the reading, writing and social studies scores.".** |

Now let’s summarize the variable that we have just created.

|  |  |
| --- | --- |
| * **Analyze** * **Descriptive Statistics** * **Descriptives...** * select **total** | \* creating a variable that is a total  \* of some of the test scores.  **summarize var = total.**  **display dictionary**  **/var = total.** |

We will recode **total**to become **grade**as shown below.

|  |  |
| --- | --- |
| * **Transform** * **Recode...** * **Into different variables...** * select **total** as the "input -> output variable" * type **grade** as the output variable * **change** * **click** "old and new values" * **click** on "range: lowest thru"and type 80 * as the value and type 0 as the new value. * **click** Add. * **click** on "range" and continue to enter * the values according the table below. * For the last category, * **click** on "range: * thru highest" * recoding **total** into **grade**: * lowest - 80 = 0 * 80 - 110 = 1 * 110 - 140 = 2 * 140 - 170 = 3 * 170 - highest = 4 | \* assigning some letter grades to these test scores.  **recode total (0 thru 80=0) (80 thru 110 =1) (110 thru 140=2)**  **(140 thru 170=3) (170 thru 300=4) into grade.**  **execute.**  **value labels grade 0 "f" 1 "d" 2 "c" 3 "b" 4 "a".**  **variable labels grade**  **"these are the combined grades of reading, writing and social studies scores.".**  **display dictionary**  **/var = grade.**  **list read write socst grade**  **/cases from 1 to 10.** |

Let’s label the data set itself so that we will remember what the data are. We can also add some notes to the data set.

|  |  |
| --- | --- |
| * **Syntax must be used to label data.** * **Utilities** * **Data File Comments** * The variable gender was renamed to female; * The values of race coded as 5 were recoded to be missing. | **file label "High School and Beyond".**  **document The variable gender was renamed to female;**  **The values of race coded as 5 were recoded to be missing.**  **display document.** |

Finally, let’s make z-scores of some of our variables.  There are at least two way that you could do this. If you remember the formula for creating z-scores and you know the mean of the variable, you can use the transform -> compute function as we did before.  Another way to create the z-scores is shown below.

|  |  |
| --- | --- |
| * **Analyze** * **Descriptive Statistics** * **Descriptives...** * select **read** * **click** on the box "Save standardized values as variables" | \* there is another way to create variables  \* in SPSS that uses special functions.  **descriptives var = read**  **/save.**  **summarize var = zread.**  **list read zread**  **/cases from 1 to 10.** |

## 2.7  Using functions

SPSS has many functions that you can use to create new variables.  First we will create a new variable that contains the mean of **read** for each level of **ses**.

|  |  |
| --- | --- |
| * **Data** * **Aggregate** * select **ses** as Break Variable(s) * select **read** as Summaries of Variable(s) * **click** on Name and Label and type **rmean** as name | **aggregate**  **/break = ses**  **/rmean = mean(read).** |

Next, we will create a new variable that contains the mean of several variables.  Please note that there will be a mean for observation 9 even though it has a missing value for **science**.

|  |  |
| --- | --- |
| * **Transform** * **Compute Variable** * type **row\_mean** in Target Variable box * select Statistical from Function group box * double click on mean from Functions and Special Variables box * double click or type **read, write, math, science** in the Numeric * Expression box | **compute row\_mean = mean(read, write, math, science).** |

Before we leave this unit, let’s save the data set.

|  |  |
| --- | --- |
| * **File** * **Save As...** * **hs1** | **save outfile "c:\temp\hs1.sav".** |

## 3.0 Syntax version

\* open the data file.

get file "c:\temp\hs0.sav".

\* ordering the variables in a way that makes sense.

save outfile = "c:\temp\hs1.sav"

/ keep id gender all.

get file c:\temp\hs1.sav".

display variables.

\* adding variable and value labels to schtyp.

variable labels schtyp "the type of school the student attended.".

value labels schtyp 1 "public" 2 "private".

display dictionary

/var = schtyp.

list schtyp

/cases from 1 to 10.

\* changing prgtype from a string to a numeric variable (called prog).

autorecode variables = prgtype

/into prog

/print.

\* adding the variable label.

variable labels prog "The type of program

in which the student was enrolled.".

rename variables (gender = female).

variable labels female "The gender of the student.".

value labels female 1 "female" 0 "male".

display dictionary

 /var = female.

list female

 /cases from 1 to 10.

\* recoding race = 5 to missing.

frequencies var = race.

recode race (5 = sysmis).

frequencies var = race.

\* adding notes to the data set and viewing the notes.

document the variable gender was renamed to female.

document values of race coded as 5 were recoded to be missing.

display document.

\* creating a variable that is a total of some of the

\* test scores.

compute total = read + write + math.

summarize var = total.

\* creating a variable that is a total of the reading

\* writing and social studies test scores.

compute total = read + write + socst.

variable labels total "the total of the reading, writing and social studies scores.".

\* creating a variable that is a total of some of the test scores.

summarize var = total.

display dictionary

/var = total.

\* assigning some letter grades to these test scores.

recode total (0 thru 80=0) (80 thru 110 =1) (110 thru 140=2)

(140 thru 170=3) (170 thru 300=4) into grade.

execute.

value labels grade 0 "f" 1 "d" 2 "c" 3 "b" 4 "a".

variable labels grade "these are the combined grades of reading, writing and social studies scores.".

display dictionary

/var = grade.

list read write socst grade

/cases from 1 to 10.

file label "High School and Beyond".

document The variable gender was renamed to female;

The values of race coded as 5 were recoded to be missing.

display document.

\* there is another way to create variables

\* in SPSS that uses special functions.

descriptives var = read

/save.

summarize var = zread.

list read zread

/cases from 1 to 10.

aggregate

/break = ses

/rmean = mean(read).

save outfile "c:\temp\hs1.sav".