

Psychology 310
Lab 5
Data Input and Manipulation in *SPSS*

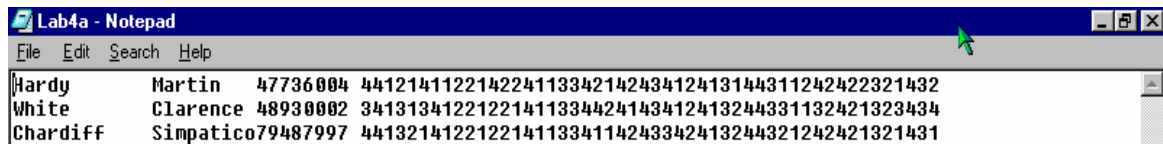
In using *SPSS* or any other data analysis program in your research, you often will face problems involving data input, recoding, and manipulation.

In this lab, we shall explore these typical situations.

1. Someone sends you data in a file format that is not compatible with *SPSS*. You need to get the data into *SPSS* so that it can be analyzed. This happens frequently when the data are in an “ASCII” plain text file, without delimiters (i.e., commas or spaces or tabs are not placed between adjacent numbers. As an example of such a file, we have LAB5A.txt, a file with last name, first name, ID, and responses to 50 questions.
2. Someone sends you data in a Microsoft Excel File.
3. You want to recode data (i.e., change some of the values) so that the data are compatible with someone else’s design specification.
4. You want to “score” data according to some specification. For example, you have raw output from a set of SCANTRON scoresheets, and you want to produce test grades for a group of students, using a key. For example, suppose LAB5A.txt is supposed to be scored with the correct responses in KEY.TXT.
5. Your data are already in *SPSS*, but must be formatted in a particular way so that they can be analyzed by another program. As an example, suppose you are instructed to take the data in LAB5A.txt, and save it in an output text file WITHOUT the ID and name information, but with all correct responses replaced by a period.
6. You want to combine “item responses” into “scale scores,” by summing items according to some rule.

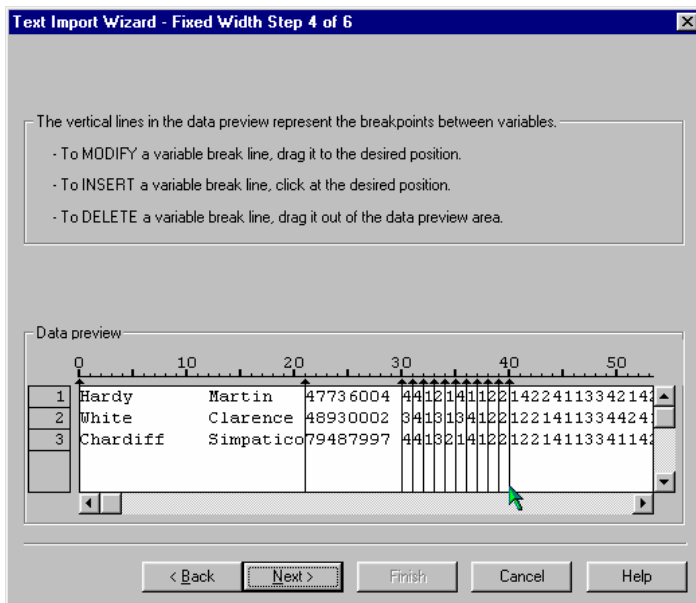
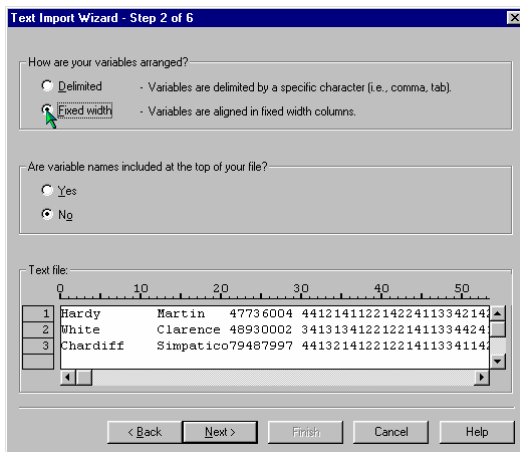
Each of these situations presents challenges. In some cases, the challenges can prove to be extremely frustrating. In today’s lab, we’ll examine some of these cases in detail, so you can learn about some typical methods for handling them.

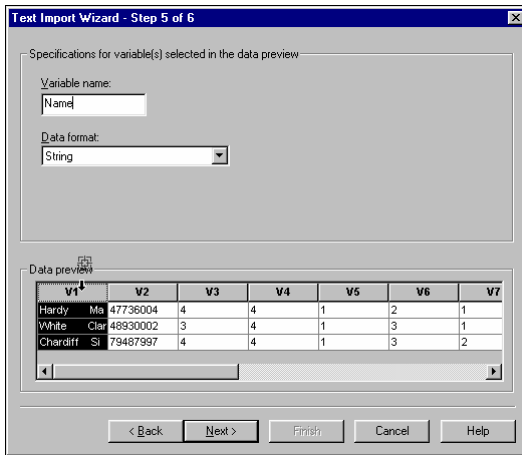
- 1. Someone sends you data in a file format that is not compatible with *SPSS*. You need to get the data into *SPSS* so that it can be analyzed. This happens frequently when the data are in an “ASCII” plain text file without delimiters (i.e., commas or spaces or tabs are not placed between adjacent numbers). As an example of such a file, we have LAB5A.txt, a file with last name, first name, ID, and responses to 50 questions.**



SPSS does allow you to input this data using the pull-down menus, but it’s a long and tedious process. You have to select *File* → *Open* → *Data*, then change the *Files of Type* box to “Text(*.txt)” in the *Open File* window, and open up “LAB5A.txt”. Then follow the instructions

in the *Text Import Wizard*. Try this yourself on the first few variables (up to item 10). The key points you'll need to know are for Steps 2, 4, and 5 in the *Text Import Wizard*. In Step 2, because you have no delimiters, you'll need to click the radio button next to *Fixed width*. In Step 4, you'll need to place breakpoints between the variables. Do this by using the mouse to drag the existing vertical arrows to where you want breakpoints. To add new breakpoints, left click on the places where want to add them. In Step 5, click on each variable in the *Data preview* box and change its name in the *Variable name* box above.



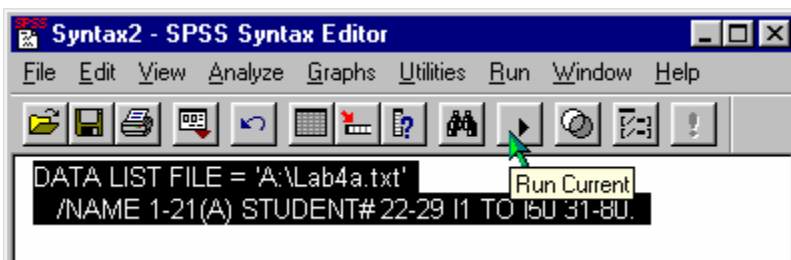


A faster way to input the data is to write a program in *SPSS* “command syntax”. Command syntax is a versatile tool in *SPSS* that allows you to perform all the same tasks you can perform with the pull-down menus plus a range of other tasks. To write syntax, you need to open a *Syntax* window by selecting *File* → *New* → *Syntax*. In this window, type

```
DATA LIST FILE = 'LAB5A.TXT'
/ NAME 1-21(A) STUDENT# 22-29 I1 TO I50 31-80.
```

This tells *SPSS* that the first 21 columns of the file to be read in comprise a string variable (that’s what the “A” in parentheses means) called “NAME”, that columns 22 to 29 comprise a numeric variable called “STUDENT#”, and that columns 31 to 80 comprise fifty numeric variables (one for each item on the exam) called “I1”, “I2”, all the way up to “I50”.

To run the syntax, simply highlight the text and click on the *Play* button (it looks like a little arrowhead pointing right) from the toolbar. You could also select *Run* → *All* from the pull-down menu in the *Syntax* window.



After you do this, you will possibly see a data file open up with the variable names. However, you may also see no data. Note that, at the bottom of the Data window, there is a note, “Transformations pending.” This means that your syntax has not yet executed. To execute the pending transformations, choose *Transform* → *Run Pending Transformations* from the menu system.

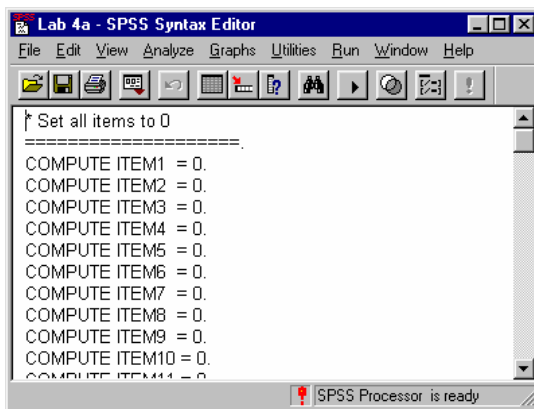
For the rest of this lab, we will use a combination of syntax and pull-down menus. As we discussed in an earlier class, anything you do using the pull down-menus can be converted into

syntax by clicking on the *Paste* button in any active dialog box rather than the *OK* button. Do this for the remaining exercises of this lab.

- You want to “score” data according to some specification. For example, you have raw output from a set of SCANTRON scoresheets, and you want to produce test grades for a group of students, using a key.**

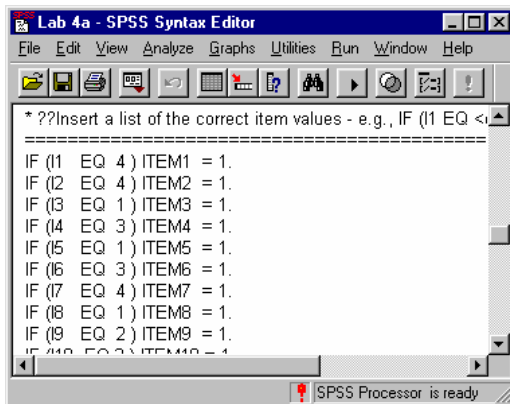
To save time, the key for this example is already entered into a syntax file entitled “Lab5a.sps”. Double-click on this file to load it into an SPSS syntax window, or, alternatively, use *File->Open->Syntax* and select the file from the dialog.

Note that there are three components to this syntax. The first component involves creating 50 variables (ITEM1 to ITEM50) and setting them all to equal zero.



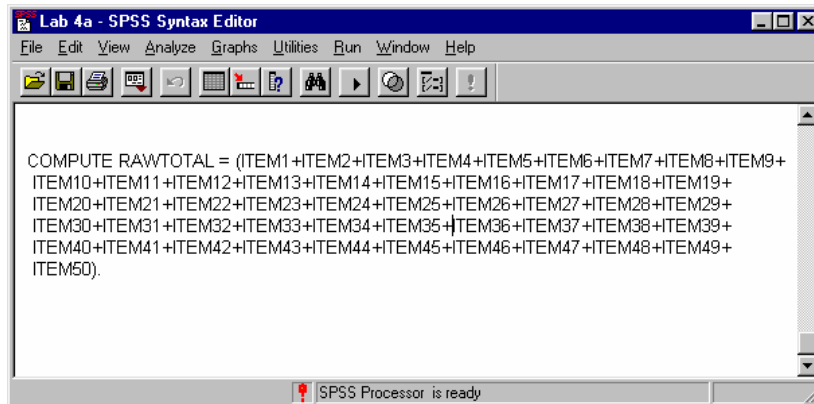
```
Lab 4a - SPSS Syntax Editor
File Edit View Analyze Graphs Utilities Run Window Help
[Icons]
↑ Set all items to 0
-----
COMPUTE ITEM1 = 0.
COMPUTE ITEM2 = 0.
COMPUTE ITEM3 = 0.
COMPUTE ITEM4 = 0.
COMPUTE ITEM5 = 0.
COMPUTE ITEM6 = 0.
COMPUTE ITEM7 = 0.
COMPUTE ITEM8 = 0.
COMPUTE ITEM9 = 0.
COMPUTE ITEM10 = 0.
COMPUTE ITEM11 = 0.
SPSS Processor is ready
```

These variables will take on two possible values, “0” if the person answered incorrectly on the corresponding question (I1 to I50) and “1” if he or she answered correctly. The second component to this syntax contains the key. For each question, if the person’s response matches the correct response [e.g., IF (I1 EQ 4)], the value of the corresponding ITEM will be set to 1. Note that although this may have been a multiple-choice test with options represented by letters, the key uses numbers instead (e.g., “4” instead of “D”).



```
Lab 4a - SPSS Syntax Editor
File Edit View Analyze Graphs Utilities Run Window Help
[Icons]
* ??Insert a list of the correct item values - e.g., IF (I1 EQ <...
-----
IF (I1 EQ 4) ITEM1 = 1.
IF (I2 EQ 4) ITEM2 = 1.
IF (I3 EQ 1) ITEM3 = 1.
IF (I4 EQ 3) ITEM4 = 1.
IF (I5 EQ 1) ITEM5 = 1.
IF (I6 EQ 3) ITEM6 = 1.
IF (I7 EQ 4) ITEM7 = 1.
IF (I8 EQ 1) ITEM8 = 1.
IF (I9 EQ 2) ITEM9 = 1.
IF (I10 EQ 2) ITEM10 = 1.
SPSS Processor is ready
```

The third component to the syntax simply calculates totals the number of correct responses for each person. Try to write the syntax to calculate each person's percentage on the exam. (Hint: If you are not sure how to do this, use the pull-down menus and then click the *Paste* button instead of the *OK* button).



Now go ahead and score the data by running the syntax.

3. You want to combine “item responses” into “scale scores,” by summing items according to some rule.

Suppose the data in Lab5A.txt are actually responses on the **Good Riddance to Actual Discourse** in **SCHOOL** questionnaire. Calculate the three individuals' scores on the 6 scales of this test. Scores on Scales 1 to 5 are found by taking the *mean* of the responses on the items that make up the scales (see below). Scores on Scale 6 are obtained by *summing* the scores from Scales 1 to 5.

Scale 1 (YME) – Items 1,6,11,16,21,26,31,36,41,46

Scale 2 (SCULSAPAN) – Items 2,7,12,17,22,27,32,37,42,47

Scale 3 (IMOK) – Items 3,8,13,18,23,28,33,38,43,48

Scale 4 (TYMONMYHANZ) – Items 4,9,14,19,24,29,34,39,44,49

Scale 5 (MYTASGUD) – Items 5,10,15,20,25,30,35,40,45,50

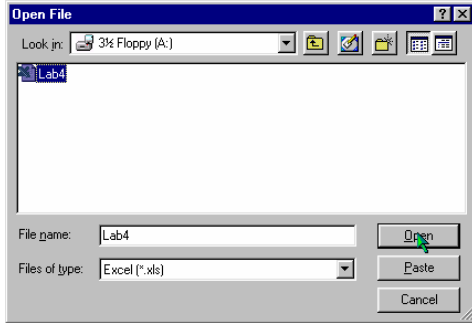
Scale 6 (TOTAL) – Scales 1,2,3,4,5

(Hint: There is a shorthand way of writing the COMPUTE syntax for these scales. For mean scores, type COMPUTE <new variable name> = MEAN (<variable 1>, <variable 2>, <variable 3>,...); for summed scores, replace MEAN with SUM.)

NOTE: Save your file into a file called MyData5.sav, and hand this in to the TA.

4. Someone sends you data in a Microsoft Excel File.

Open the file, Lab5.xls, from SPSS by selecting *File* → *Open* → *Data*. Then in the *Open File* window, change the *Files of Type* box to “Excel (*.xls)”, select the file “Lab5.xls”, and click *OK*.

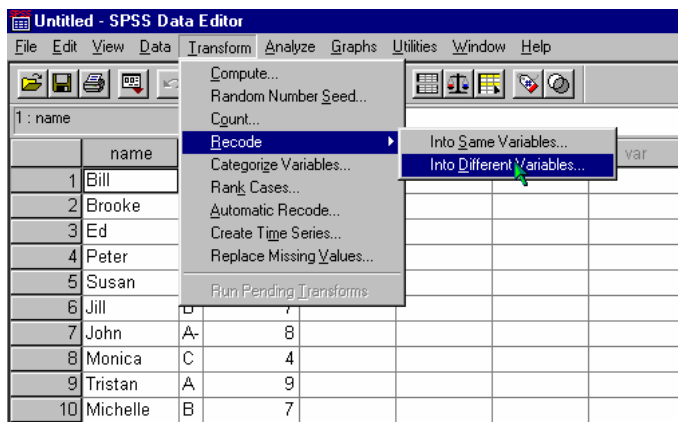


Excel will ask you if variable names are stored in the first row of the file. They are, so simply click *OK* in the *Opening Excel Data Source* window that opens up. You will see a file with several student names, letter grades, and number grades on a stanine system.

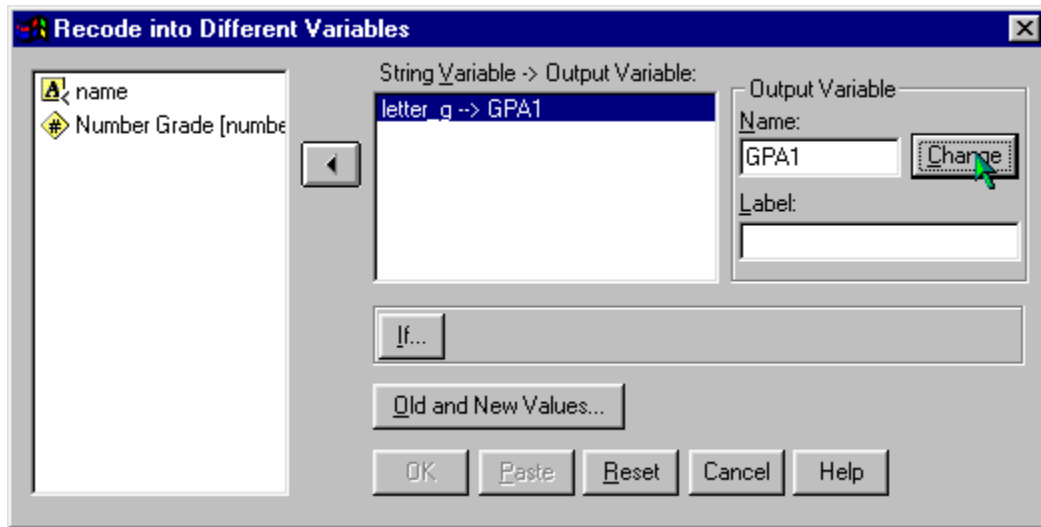
5. You want to recode data (i.e., change some of the values) so that the data are compatible with someone else’s design specification.

There are a number of different ways to recode data. One way involves using the COMPUTE command. Use this to convert the number grades, which are in a stanine metric (a nine point scale used by some universities and colleges), into the more commonly used 4-point GPA system (call the new variable “GPA”). Try doing this by writing syntax. (Hint: If you are not sure how to do this, use the pull-down menus and then click the *Paste* button instead of the *OK* button).

Now change the letter grades into number grades (using the 4-point system). But this time, use the RECODE command. Select *Transform* → *Recode* → *Into Different Variables* (Selecting *Into Same Variables* changes the existing variable).

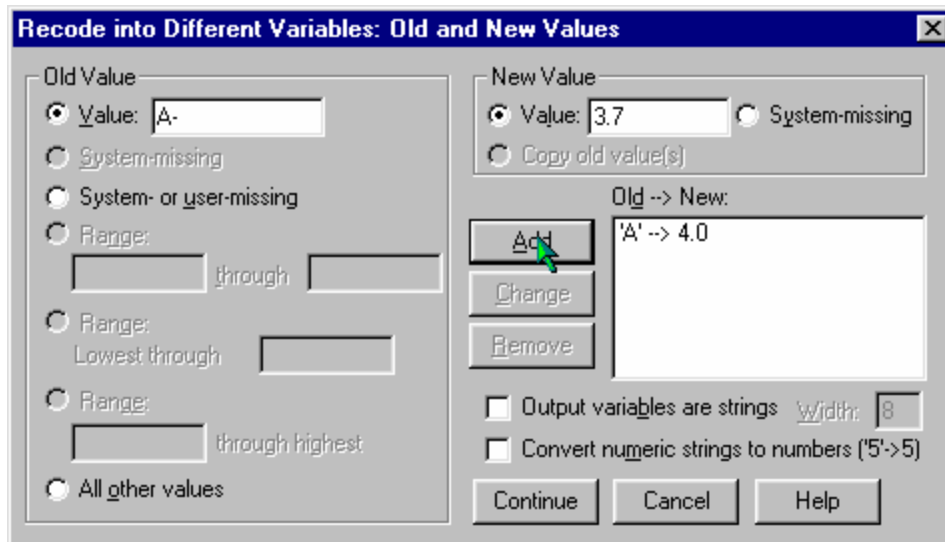


Select the *Input Variable*, Letter Grade, and name the *Output Variable*, the click the *Change* button.



Then click the *Old and New Values* button. In the new window that pops up, enter in the old and new values. Each time you enter in a pair of these values, you need to click on the *Add* button to add them to the list. Use the following scale to recode the values:

<u>Letter Grade</u>	<u>GPA</u>
A	4.0
A-	3.7
B+	3.3
B	3.0
B-	2.7
C+	2.3
C	2.0
C-	1.7
D+	1.3
D	1.0
F	0.0-0.9



When you have entered all the pairs, click *Continue* and then *Paste*.

Now try writing the syntax to recode the GPA variable back into the stanine system using the values you calculated for the first part of this question (i.e., when you computed GPA from Number Grade).

HAND-IN: Your syntax, output, and data files (and don't forget to run your syntax before handing it in). Call the first data file MyData5.sav, and the second file MyData5a.sav.

Final note: The syntax you've generated in sections 1 and 2 is almost identical to the scoring syntax for most Psych SCANTRON exams (I think the NAME variable for SCANTRON scoring syntax only goes from 1-20). Hang onto it – it may prove useful at a later date.