**NRSG 736: Handout 01**

**Example SPSS Dataset – Start to Finish:**

**Data INPUT/IMPORT and EXPORT; Initial Data Screening; Coding and Formatting**

**Dr. Melinda K. Higgins**

**SETTING UP THE CODEBOOK**

1. Begin SPSS – click “type in data”
2. A data window will open – click on the “variable” tab at the bottom
3. You can now set-up the worksheet to receive data – first you need to type in your variable names and set the data type, measurement type, a descriptive label, and any “coded” values.
4. Type in the following variable names (1st column) – variable names can not have any special characters or spaces:
   1. Row 1: SubjectID
   2. Row 2: Age
   3. Row 3: WeightPRE
   4. Row 4: WeightPOST
   5. Row 5: Height
   6. Row 6: SES
   7. Row 7: GenderSTR
   8. Row 8: GenderCoded
5. Next make sure that “numeric” is listed as the “type” for all of these except Row 7 “GenderSTR” – click in the type field and select “String.”
6. Next go to the Label column and type in the following for each:
   1. Row 1: Subject ID
   2. Row 2: Age in Years
   3. Row 3: Weight in Pounds – Before Program
   4. Row 4: Weight in Pounds – After Program
   5. Row 5: Height in Decimal Feet
   6. Row 6: Pseudo Socio-Economic-Status
   7. Row 7: Gender as a Character/Text
   8. Row 8: Gender Recoded
7. Next we will enter “values” for SES and GenderCoded
   1. For SES (Row 6) click in the “values” field and type in the following:
      1. Value = 1; Label = Low Income – then click “ADD”
      2. Value = 2; Label = Average Income – then click “ADD”
      3. Value = 3; Label = High Income – then click “ADD”
   2. For GenderCoded (Row 8) click in the “values” field and type in the following:
      1. Value = 1; Label = Male – then click “ADD”
      2. Value = 2; Label = Female – then click “ADD”
8. In the “Measure” column, click in this field for each variable and select the following for each: [\*\* THIS IS AN IMPORTANT STEP \*\* This value tells SPSS how to treat each variable.]
   1. Row 1: Subject ID **– Measure = “Ordinal”**
   2. Row 2: Age in Years **– Measure = “Scale”**
   3. Row 3: Weight in Pounds – Before Program **– Measure = “Scale”**
   4. Row 4: Weight in Pounds – After Program **– Measure = “Scale”**
   5. Row 5: Height in Decimal Feet **– Measure = “Scale”**
   6. Row 6: Socio-Economic-Status **– Measure = “Ordinal”**
   7. Row 7: Gender as a String **– Measure = “Nominal**
   8. Row 8: Gender Recoded **– Measure = “Nominal”**

**INPUTTING THE DATA:**

**TYPE IN THE DATA**

1. Now you are ready to type in data – type in the following values for each column as shown below:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **SubjectID** | **Age** | **WeightPRE** | **WeightPOST** | **Height** | **SES** | **GenderSTR** | **GenderCoded** |
| 1 | 45 | 150 | 145 | 5.60 | 2 | m | 1 |
| 2 | 50 | 167 | 166 | 5.40 | 2 | f | 2 |
| 3 | 35 | 143 | 135 | 5.60 | 2 | F | 2 |
| 4 | 44 | 216 | 201 | 5.60 | 2 | m | 1 |
| 5 | 32 | 243 | 223 | 6.00 | 2 | m | 1 |
| 6 | 48 | 165 | 145 | 5.20 | 2 | f | 2 |
| 7 | 50 | 132 | 132 | 5.30 | 2 | m | 1 |
| 8 | 51 | 110 | 108 | 5.10 | 3 | f | 2 |
| 9 | 46 | 167 | 158 | 5.50 | 2 |  |  |
| 10 | 35 | 190 | 200 | 5.80 | 1 | Male | 1 |
| 11 | 36 | 230 | 210 | 6.20 | 1 | m | 1 |
| 12 | 40 | 200 | 195 | 6.10 | 1 | f | 2 |
| 13 | 45 | 180 | 185 | 5.90 | 3 | f | 2 |
| 14 | 52 | 240 | 220 | 6.50 | 2 | m | 1 |
| 15 | 24 | 250 | 240 | 6.40 | 2 | M | 1 |
| 16 | 35 | 175 | 174 | 5.80 | 2 | F | 2 |
| 17 | 51 | 220 | 221 | 6.30 | 2 | m | 1 |
| 18 | 43 | 230 | 215 | 2.60 | 2 | m | 1 |
| 19 | 36 | 190 | 180 | 5.70 | 1 | female | 2 |
| 20 | 44 | 260 | 240 | 6.40 | 3 | male | 1 |

1. Go ahead and save your data – (within the data window) click on File / Save As – and save your data (as an SPSS datafile, \*.SAV)
2. Let’s also save this as an EXCEL formatted file – slick File / Save As – save as either \*.XLS or \*.XLSX
3. NOTICE that an SPSS outfile window is also created. Save the OUTPUT file also – click File / Save As (as an SPSS OUTPUT file \*.SPV)

**OR Open this document – CUT AND PASTE the data from the table above into the SPSS spreadsheet**

1. Once a CODEBOOK is created or even if it is not created yet, you can “CUT and PASTE” data from ANY data table (from EXCEL or a table in WORD, etc) into an SPSS data window.
2. Open the EXCEL file you just created.
3. Open a blank SPSS datafile (Click File / New / Data)
4. Highlight the cells you want to copy out of EXCEL, click COPY and then place the pointer on the 1st cell in the SPSS data window and click PASTE.
   1. If the codebook has been created, the various data types have to match for the PASTE into SPSS to work
   2. If the codebook has NOT been created, then SPSS will automatically apply formatting (variable types), but you still have to put in the Variable Labels and the “codes” (value labels).

**OR Open this data from an Excel file (“IMPORTING” from EXCEL)**

1. File / Open / Data
2. Files of Type – choose Excel (\*.xlx, \*.xlsx, \*.xlsm)
3. Select the dataset
4. Click Open
5. Choose the TAB/Worksheet you want to import [you can select among several TABs/WORKSHEETs – but you can only IMPORT ONE worksheet at a time]
6. Keep selected the “Read variable names from first row of data”
7. Go back and fix the codebook (Steps 6, 7 and 8 above)

**NOTES:**

* You can also IMPORT data formatted for SAS, Stata, S-plus, comma delimited formats, tab delimited formats
* You can also EXPORT (Save As) data formatted for each of these as well

**C:\Users\mkhiggi.EMORYUNIVAD\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\0Z3SUGEM\MM900395755[1].gif** “**LOOK” AT THE DATA**

1. Analyze / Descriptive Statistics / Frequencies
   1. Select all of the variable (except for SubjectID) and move to the right (🡪) into “Variable(s)” list
   2. Keep “Display frequency tables” selected
   3. Click “Statistics” – select Mean, Median, Standard Deviation, Minimum, Maximum, click continue
   4. Click Charts – select Histogram and “Show Normal Curve”, click continue
   5. Click OK
   6. *[or click Paste – puts “code” into a Syntax window]*
2. Review the output, look for unusual values, outliers, and typos
   1. What do you notice about the variable GenderSTR?
3. Assumptions of Normality – click on Analyze / Descriptive Statistics / Explore
   1. Select “Age” and “WeightPRE” and “WeightPOST” and “Height” – place into the “Dependent List”
   2. Leave “Factor List” blank for now
   3. Label cases by - put “SubjectID” into the list
   4. Click on Statistics Box – keep default “Descriptives” selected and also Select “Percentiles” – you also have the option to explore additional outliers here
   5. Click the Plots Box – unclick “Stem-and-Leaf” and click “Histogram” and also click “Normality Plots with Tests” – click “continue”
   6. Click “OK”
   7. Look at Histograms and Boxplots and look for outliers – notice the Height for SubjectID “18”
   8. Run through the sequence above again – click on Analyze / Descriptive Statistics / Explore
   9. Leave all of the options the same, but this time put “GenderCoded” into the “Factor List” box
   10. Click “OK”
   11. Notice the boxplots are now “grouped” by Gender – side by side boxplots of each variable by Gender
4. Fix the Height value for SubjectID “18” – this was mistyped – change the value from 2.60 to 6.20.
5. If you wish, go back and re-run your descriptive statistics and plots for Height
6. OPTIONAL NOTE: You can also get quick descriptive stats through Analyze / Descriptive Statistics / Descriptives
   1. Put in the variables of interest
   2. Click on OPTIONS and choose the statistics you wish
   3. You have the option at the bottom of the window to select “Save Standardized Values as Variables” – this is where you can CREATE Z-SCORES

**DATA CALCULATIONS and IMPACT OF DATA STRUCTURE/FORMATTING**

1. Let’s calculate BMI
   1. Transform / Compute Variable
   2. Target Variable “BMI\_PRE”
   3. Numeric Expression (WeightPRE\*703) / ((Height\*12)\*(Height\*12))
   4. Click PASTE – this opens a SPSS SYNTAX window and pastes in the following CODE

COMPUTE BMI\_PRE=(WeightPRE\*703)/((Height\*12)\*(Height\*12)).

EXECUTE.

* 1. Go ahead and calculate BMI\_POST – you can go back through the steps (a – c) above OR can simply cut and paste the SPSS SYNTAX and change the variable names as needed

COMPUTE BMI\_POST=(WeightPOST\*703)/((Height\*12)\*(Height\*12)).

EXECUTE.

**RESTRUCTURE DATA – “Stack” Time Points (Time 1 “PREP and Time 2 “POST”)**

1. File / Save As – save the data file under a different file name indicating that the data has been restructured
2. For now, let’s delete the 2 BMI variables you just created – in data view select the variable column, right slick select delete – or in variable view, select the 2 variables and click delete
3. Data / Restructure /
   1. Click Restructure Selected Variables into Cases – click Next
   2. (for now) select One [we’re only going to “stack” the 2 Weight measurements]
   3. Case Group Identification – click on the pick list and choose “Use selected variable”
   4. Put SubjectID into the Variable box
   5. Change the variable in the Target Variable to “Weight” (instead of default trans1)
   6. Put in the box below “WeightPRE” and “WeightPOST”
   7. Click Next
   8. How many index variables? – select One – click Next
   9. You can either accept the default “Index1” or type in a variable name you like – maybe “Time” – click Next
   10. Handling of Variables Not Selected – use the default of “Keep and Treat as Fixed Variable(s)”
   11. System Missing – use the default – “Create a case in new data file”
   12. You also have the option to create a COUNT variable – we’ll skip for now, but this can be helpful for counting up the number of time points for a given individual (i.e. tracking attrition, number of appointments kept, etc)
   13. Click Next
   14. Go ahead and click OK to restructure the data now – you also have the option to PASTE the syntax for this restructuring procedure
4. Save the data

**REVISIT – DATA CALCULATIONS and IMPACT OF DATA STRUCTURE/FORMATTING**

1. Now let’s create BMI using the same approach as above – **but now we only have to do it once!**

COMPUTE BMI=(Weight\*703)/((Height\*12)\*(Height\*12)).

EXECUTE.