Module 2

HOMEWORK #2

The homework this week is mainly to have you practice some of the basic functions of SPSS. For some of you this will be super easy while for others it may be new and a bit more challenging. Regardless, I hope all of you learned something new about SPSS this week! This is also an exercise in following directions. Be sure to follow ALL directions exactly.

There are a total of 40 pts. for this homework. Each task has a specified value shown in parentheses that make up the 40 pts.

TASK #1. (5 pts)

Enter data from Table 4.1 in the Field Text (p. 112) into an SPSS file and save as “Naming Data\_ your last name”. This data was collected to explore differences in characteristics between university lecturers and students. Please do the following:

* Assign appropriate Variable names
* Assign appropriate Label names
* As a point of interest, this particular dataset uses several string variables which is quite uncommon. We don’t typically use names in a dataset for example. Instead we assign a personal identification number (PIN). Please substitute a PIN for each of the named participants. This practice is both to preserve privacy and it provides each participant a unique identifier (with the names, there could be more than one Mark, for example). We will deal with the other string variable (Job) below.
* “Job” is your independent variable of interest (lecturer or student), so create a new variable and call it “Group”. You will need to transform the “job” variable from a string variable to a coding variable (“Group”). To do this, you will need to transform “student” to a “0” and lecturer to a “1” (or something similar). Zeroes and ones are typically used for coding grouping variables. So your newly formed “Group” variable should be a list of zeroes and ones.
* Be sure to assign a value to the “0”s and “1”s. You THINK you will remember what each means but trust me, you won’t after this week! Besides, it is just good practice. [See SPSS Tip 4.3 in Field]
* Define all variables in the “Variable View”. The dependent variables (alcohol, income, number of friends, neuroticism) can be labelled according to what YOU think they mean and what you would use them for. Be sure to define the level of measurement too.
* Save your SPSS file as “Naming Data\_your last name”. It should show up as a .sav file since you created it in the Data Editor. Submit as a separate document in Canvas.

TASK #2. (10 pts)

For this task, open the Excel file called “ACL data subset” in the Module 2 folder.

* Import the data into an SPSS file.
* You might have to adjust variable names to meet SPSS naming “rules”. SPSS will let you know if the variable name is not allowable. If it is acceptable, then you are fine with the original names.
* Define all variables using the variable view of the Data Editor.
* Create labels for all variables using the following information: Subj#= Personal Identification number; Grp= intervention group; Ht (“)= height in inches; Wt(lb)=weight in lbs; Dom foot= dominant foot; ROM3L= Left mean hamstring length in degrees; ROM3R= Right mean hamstring length in degrees.
* For “Group” assign Value labels as follows: (0 is Control; 1 is Prevention Program)[NOTE: ACL dataset has these coded as (1 = Control; 2 = Prevention Program) so you’ll need to Recode as Same Variable to 0’s and 1’s first.
* For ROM3R and ROM3L, the values are in degrees of knee flexion in the active knee extension test. These were taken with a digital inclinometer (hence the degree of precision). We really don’t need quite this degree of precision, so change both of these variables in the variable view to ONE digit after the decimal point.
* Compute Height in meters and create a new variable named “Height\_m”
* Compute Weight in kg and create a new variable named “Weight\_kg”
* Compute a new variable for BMI using the formula BMI= weight (kg) / [height (m)]2
* Save as an SPSS file called “ACL data subset\_your last name” and submit as another separate document in Canvas.
* Save a copy for yourself because we may use this file again later on.

TASK #3. (25 pts)

For this task, open the file titled “Health Habits\_100” in the Module 2 folder.

1. In SPSS, go to “Analyze”, “Descriptive Statistics”, and “Frequencies”. Obtain frequency tables for “Number of injuries” and “Age Group” (two different frequency tables).

(10 pts)

* + Before you ask SPSS to analyze anything, check to make sure that the 2 variables of interest (number of injuries, and Age group) are properly defined. Do this by going to your Variable view and select the correct level of measurement for the “Measure” column. Note that in SPSS, you only have 3 options (Nominal, Ordinal, and Scale). Scale is selected when you have interval or ratio data. For Age group, people have been put in groups according to pre-defined ranges. See “Values” to see how these 3 groups were defined (now you see how important it is to define these when you enter and define variables!). Since we have placed people in age groups, we have essentially converted a continuous variable (age) into a categorical variable (age group). Since the categories are ordered, the “Measure” should be ordinal. (If this is not clear, please review Levels of Measurement on pages 8-11 of Field.
  + For Statistics, select Mean, Median and Mode.
  + For Charts, select “bar chart” since both variables are discrete.
  + Export the resulting output to a Word document that you will turn in via Canvas: (Two frequency tables, one Statistics table showing Mean, Median, Mode)
  + Now let’s try selecting cases. Under “data” in the data view, use an “If condition is satisfied” to select ONLY cases that are “overweight” in the BMI category. Now re-run the Statistics table showing Mean, Median, and Mode and the frequency tables for “Number of Injuries” and “Age Group” as before
  + How many valid cases were analyzed to generate the Mean, Median, and Mode for the selected cases? Did the new values differ much from the values generated when using the entire sample versus selected cases? Did this surprise you?
  + Don’t forget to remove the filter for selecting cases before moving on to the next Tasks!! Note: You do not need to export these new tables of selected cases to your Word document.

1. Now go to “Graphs” and “Chart Builder” to create a bar graph of “MVPAAvg” for the 3 different age groups. (go to your variable view and check the Label for MVPAAvg to see what this variable really is). (5 pts.)
   * Once again, before you ask SPSS to build a chart, check to make sure that the 2 variables you are interested in are properly defined. If you classify Age Group as a Scale Measure for example (rather than ordinal), it will not work when you ask SPSS to create a bar graph of MVPAAvg for the Age Groups (try it!). MVPAAvg is number of minutes per week of moderate to vigorous physical activity so it is a continuous variable and should be a “Scale” measure.
   * Another good thing to do before analyzing or building a graph is to get a simple boxplot of your data (in this case MVPAAvg). We will explore boxplots more later but for now, you should know what the box plot represents and what each line of the boxplot represents. Perhaps more useful though is that it allows us to look at all of our data at once AND to spot outliers. One outlier in particular jumps out at us in the resulting box plot. Participant #46 reported almost 1500 minutes per week of MVPA. Does this represent a reporting error, recording error or just an overactive individual?! We will learn how to deal with these later in Chapter 6, but for now, we will just leave it. It will get averaged in with the rest of the data.
   * Now go ahead and build your bar graph with Age Group on the x-axis and MVPAAvg on the y-axis.
   * Edit your chart by adding a title, and changing its appearance as you like (color, lines, etc). Just play with things to familiarize yourself with what SPSS can do!
   * Once you are happy with your chart, export it to the same Word document you created earlier (above).
2. For the final part of this task, answer the following questions relating to your graphs/tables in #3A and 3B. (10 pts.) Answer the questions in regards to outputs generated using the full dataset and not the ones where you selected cases.
3. Looking at your Frequencies output, how many people reported a valid age group and number of injuries? 220 total
4. From that same output, we see that the mean, median and mode for age group are all pretty similar, which is characteristic of a normal distribution. Highlight the *values* for mean, median and mode in the table then copy the table into your Word document.
5. How many participants were in the “Older” group? What percentage of all participants does the Older age group represent? Older group participants= 69 people, which is 31.2% of all the total participants.
6. How many participants reported no injuries? What percent of the sample do these participants represent? 188 participants reported no injuries, which is 85.1% of the total sample.
7. Which age group was least active (participated in the least amount of physical activity)?

Were you surprised? Young people (20-39) participated in the least amount of physical activity. No, I was not really surprised by this finding as this is the age group where many individuals are experiencing a lot of changes. The people of this age group have not gotten to the point in their lives where things are generally more constant (ie: job, salary, family, responsibilities, housing, etc.).

By the way, this is real data (at least part of the data) that was used to study injuries in women aged 20 to 80 who were physically active compared to those who were not active.

In summary:

You will submit 3 files for Homework #2:

1. Naming Data file for Task 1 (.sav)
2. ACL data set for Task 2 (.sav)
3. Word file (this document) with charts for Tasks #3A and 3B, and answers to Questions 1-5 above for Task #3C. (.doc or .docx)

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistics** | | | |
|  | | AGE\_Group | Number of injuries sustained |
| N | Valid | 220 | 220 |
| Missing | 1 | 1 |
| Mean | | 2.16 | .17 |
| Median | | 2.00 | .00 |
| Mode | | 2 | 0 |

**Frequency Table**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **AGE\_Group** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Young (20-39) | 33 | 14.9 | 15.0 | 15.0 |
| Middle (40-59) | 118 | 53.4 | 53.6 | 68.6 |
| Older (60+) | 69 | 31.2 | 31.4 | 100.0 |
| Total | 220 | 99.5 | 100.0 |  |
| Missing | System | 1 | .5 |  |  |
| Total | | 221 | 100.0 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Number of injuries sustained** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 0 | 188 | 85.1 | 85.5 | 85.5 |
| 1 | 26 | 11.8 | 11.8 | 97.3 |
| 2 | 6 | 2.7 | 2.7 | 100.0 |
| Total | 220 | 99.5 | 100.0 |  |
| Missing | System | 1 | .5 |  |  |
| Total | | 221 | 100.0 |  |  |

**Bar Chart**

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated



Chart, bar chart

Description automatically generated