Lab 2: Data Management and Descriptive Statistics

This lab will use data from the 2013 Pan-Canadian Assessment Program (PCAP) developed by the Council of Ministers of Education, Canada. This assessment is administered to Grade 8 students every three years to gather information regarding students' academic achievement in mathematics, reading, and science across Canada. The first PCAP was administered in 2007. One purpose of PCAP is to determine whether students across Canada reach similar levels of performance at about the same time in their schooling despite differences in school programs from one part of the country to another.

You have been provided with the PCAP 2013 data for Alberta students and teachers. In this lab, we will focus on the students' responses. Please work through this lab to gain a strong understanding of the data set. You may need to consult with **your textbook** and **the internet** as you work through this lab. This lab is designed to scaffold your learning; this means the first few questions will provide full instructions, while subsequent questions will have less instructions. This is done purposely so that you may develop your own skills as guidance is lessened throughout the lab (e.g., Table 1 is fully outlined for you, but Table 4 requires you to develop the full table).

1. Descriptive statistics: outlier and typo examination

Typos are difficult to detect in a spreadsheet, but may make a **BIG** difference in subsequent analyses. Hence, it is important that you ensure all your data sets are correct before you start any analyses. Good analyses based on faulty data are not useful to anyone!!! One way to detect outliers and typos is to run some descriptive statistics on each variable. Please click: Analyze → Descriptive Statistics → Descriptive → Select all variables → Click Options → Ensure *minimum*, *maximum*, and *range* are checked → Continue → OK.

> **Question:**
> 1. Looking at your output for outlier examinations, were there any outliers? If yes, please list the outlier(s). What might have caused the outlier(s) to be in your data set? What did you look for in the output to indicate you had outliers, if any?

2. Understanding your data

To better understand your data, please provide some basic descriptive statistics about each of your variables. Please compute the mean, kurtosis, skewness, minimum, maximum, range, variance, and standard deviation for the science scores and each of the bullying items, by

going to Analyze → Descriptive Statistics → Descriptive → Select appropriate variables and move to Variable(s) → Options → Select appropriate analyses → Continue → OK

Questions:
2. Please answer these questions for the scale variables:
   a) How many people completed your survey? _____
   b) Minimum _____ Maximum _____
   c) Mean _____ Range _____ Standard Deviation _____ Variance _____
   d) Skewness _____ Kurtosis _____ What do the skewness and kurtosis values indicate? (You may need to use Google.)

3. What is the difference between variance and standard deviation?

When we present the data from question 2 for a paper, we need to use a table. Please fill out the table below with the responses for the scale variables:

Table 1
Title (please make a title for this table)

	n	Mean	Standard Deviation	Skewness	Kurtosis
Science Score	??	??	??	??	??
SQ13A	??	??	??	??	??
SQ13B	??	??	??	??	??
SQ13C	??	??	??	??	??
Etc.		Etc.	Etc.	Etc.	Etc.

Please ensure table is in APA format.

Now, we need to run some descriptive statistics on nominal variables. Please compute some frequency counts for these variables by clicking Analyze →Descriptive Statistics → Frequencies → select appropriate variables and move to Variable(s) → OK.

Questions:
4. Please answer this question for EACH of your nominal variables.
   a) How many people were in each category?

5. Why did we have to run two different types of analyses (i.e., descriptive statistics and frequencies) to better understand our data? NB: all *understanding your data* analyses are called *descriptive statistics*, despite SPSS giving them different names.

6. Why would it be inappropriate to calculate a mean for every variable?

Now that you have worked with your own data set (which was a sample of the Alberta data set), you will need to learn to work with larger data sets. Please download the "PCAP 2013 Data_Alberta" file for all subsequent lab questions. This data set includes all Alberta students who completed the PCAP 2013 assessment. The Alberta data set will be used for the SPSS laboratory activities, while the Canada-wide data set will be used for Learning Tasks #3 and #4.

3. Understanding the variables in the data set

Before you start any analyses, it is important to understand your data set. Please download and open the data set entitled PCAP 2013 Data_Alberta and select the Variable View tab on the bottom left.
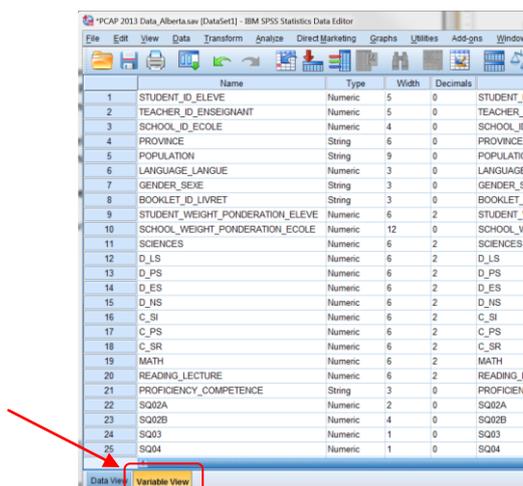
Looking at the *Name* of the variables is often not very useful because they are typically abbreviated. Hence, we often use a codebook to decipher between all the different variables in a data set. Please consult the "PCAP 2013 Merged Codebook" file to better understand the variables collected and their names in your SPSS file.

Questions:

7. Please indicate what each of these SPSS variable names represent:
a) STUDENT_WEIGHT_PONDERATION_ELEVE
b) SCIENCES
c) C_SI
d) SQ13A
e) SQ26B
f) SQ31B

8. In many of the questionnaire items (e.g., SQ03 and SQ04), students' responses were entered as "-9" or "-8" in SPSS. What do these values indicate?

9. Why is it important to use a value to represent a missing response instead of leaving that cell blank?

4. Weighting your participants

A *Student Weight* variable has been developed for the PCAP 2013 data. It is important that you use this variable throughout the lab.

Questions:

10. What is the purpose of a weighted variable?

11. Please calculate students' mean science, math, and reading scores using:
   a) no student weights. Select Analyze → Descriptive Statistics → Descriptives → select SCIENCES, MATH, READING_LECTURE and move over to the Variable(s) box → select OK.
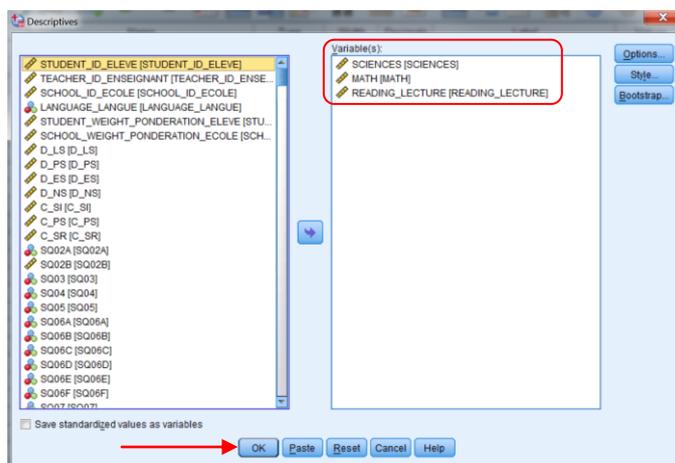
   b) student weights. First, you need to put the student weight onto your data set. Select Data → Weight Cases → select Weight cases by → select STUDENT_WEIGHT_PONDERATION_ELEVE and move over to the Frequency Variable box → select OK.
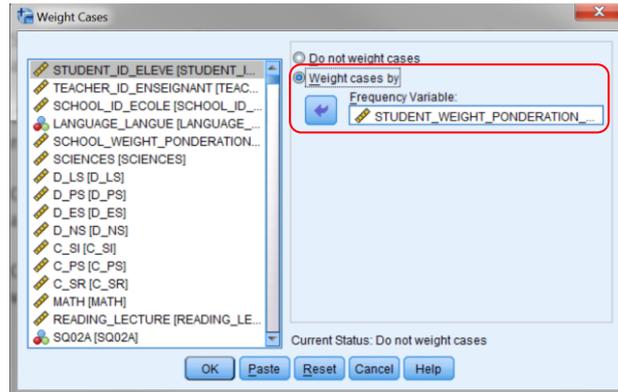
Now run the same descriptive analysis as in **3. Understanding the variables in the data set.**

c) Please fill in the following table (i.e., fill in every '??') with the descriptive statistics calculated above:

Table 1.

Comparison of PCAP Achievement when using Student Weights

	No Student Weights Used (*n=??*)		Student Weights Used (*n=??*)	
	Mean	Standard Deviation	Mean	Standard Deviation
Science	??	??	??	??
Math	??	??	??	??
Reading	??	??	??	??

d) What are the main differences between using and not using the student weights?

12. Why is it important to ensure this student weight variable is used during all our analyses of the PCAP data?

13. Please compare the mean of the science scores calculated from the data set you entered (i.e., SPSS lab 1) and the Alberta data set (using the weights). What is the difference? Why is there a difference between these two values?

You have now completed the lab.