## 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  $\rightarrow$  Descriptive Statistics  $\rightarrow$  Descriptive  $\rightarrow$  Select all variables  $\rightarrow$  Click Options  $\rightarrow$  Ensure *minimum*, *maximum*, and *range* are checked  $\rightarrow$  Continue  $\rightarrow$  OK.

## **Question:**

 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  $\rightarrow$  Descriptive Statistics  $\rightarrow$  Descriptive  $\rightarrow$  Select appropriate variables and move to Variable(s)  $\rightarrow$  Options  $\rightarrow$  Select appropriate analyses  $\rightarrow$  Continue  $\rightarrow$  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:

····· (1	<b>J</b>				
	n	Mean	Standard	Skewness	Kurtosis
			Deviation		
Science Score	??	??	??	??	??
SQ13A	??	??	??	??	??
SQ13B	??	??	??	??	??
SQ13C	??	??	??	??	??
Etc.		Etc.	Etc.	Etc.	Etc.

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

\*\*\*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  $\rightarrow$ Descriptive Statistics  $\rightarrow$  Frequencies  $\rightarrow$  select appropriate variables and move to Variable(s)  $\rightarrow$  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.

File	Edit	View	Data	Transform	Analyze	Direct I	larketing g	raphs <u>U</u> ti	lities Add- <u>o</u>	ns <u>W</u> indo
				) <b>E</b>	<b>~</b> 📱	i 📥	=	1	<b></b>	- 4
				Name			Type	Width	Decimals	
		STUDE	INT_ID_	ELEVE			Numeric	5	0	STUDENT
1	2	TEACH	IER_ID	ENSEIGNA	NT		Numeric	5	0	TEACHER
	3	SCHO	OL_ID_E	ECOLE			Numeric	4	0	SCHOOL_
4	1	PROV	NCE				String	6	0	PROVINC
	5	POPU	ATION				String	9	0	POPULAT
(	5	LANG	JAGE_L	ANGUE			Numeric	3	0	LANGUAG
	7	GEND	GENDER_SEXE				String	3	0	GENDER
1	3	BOOKLET_ID_LIVRET				String	3	0	BOOKLET	
	)	STUDENT_WEIGHT_PONDERATION_ELEVE				Numeric	6	2	STUDENT	
1	0	SCHO	SCHOOL_WEIGHT_PONDERATION_ECOLE				Numeric	12	0	SCHOOL
1	1	SCIEN	SCIENCES				Numeric	6	2	SCIENCES
1	2	D_LS	D_LS			Numeric	6	2	D_LS	
1	3	D_PS	D_PS			Numeric	6	2	D_PS	
1	4	D_ES	D_ES			Numeric	6	2	D_ES	
1	5	D_NS	D_NS			Numeric	6	2	D_NS	
1	6	C_SI			Numeric	6	2	C_SI		
1	7	C_PS	C_PS			Numeric	6	2	C_PS	
1	8	C_SR			Numeric	6	2	C_SR		
1	9	MATH			Numeric	6	2	MATH		
2	0	READ	READING_LECTURE			Numeric	6	2	READING	
2	1	PROFI	PROFICIENCY COMPETENCE			String	3	0	PROFICIE	
2	2	SQ02A	SQ02A			Numeric	2	0	SQ02A	
2	3	SQ02E	SQ02B				Numeric	4	0	SQ02B
2	4	SQ03					Numeric	1	0	SQ03
							All second as		-	

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.

	(	Variable(s):	Ontions
STUDENT_ID_ELEVE (STUDENT_ID_ELEVE)	-	SCIENCES [SCIENCES]	Options
TEACHER_ID_ENSEIGNANT [TEACHER_ID_ENSE		nath [math]	Style
SCHOOL_ID_ECOLE [SCHOOL_ID_ECOLE]		READING_LECTURE [READING_LECTURE]	Bootstra
LANGUAGE_LANGUE [LANGUAGE_LANGUE]			
STUDENT_WEIGHT_PONDERATION_ELEVE [STU			
SCHOOL_WEIGHT_PONDERATION_ECOLE [SCH			
D_LS [D_LS]			
D_PS [D_PS]			
D_ES [D_ES]			
D_NS [D_NS]			
🖉 C_SI [C_SI]			
C_PS [C_PS]			
C_SR [C_SR]	4		
💫 SQ02A [SQ02A]			
🔗 SQ02B [SQ02B]			
💫 SQ03 [SQ03]			
💫 SQ04 [SQ04]			
🗞 SQ05 [SQ05]			
💫 SQ06A [SQ06A]			
💫 SQ06B [SQ06B]			
🗞 SQ06C [SQ06C]			
🗞 SQ06D [SQ06D]			
💫 SQ06E [SQ06E]			
💫 SQ06F [SQ06F]	-		
<u> 2007 120071</u>		L	
Save standardized values as variables			
OK	Paste	Keset Cancel Help	

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.



STUDENT_ID_ELEVE [STUDENT_L.  TEACHER_ID_ENSEIGNANT [TEAC  SCHOOL_ID_ECOLE [SCHOOL_D  LANGUAGE_LANGUE [LANGUAGE  SCHOOL_WEIGHT_PONDERATION  SCHOOL_WEIGHT_PONDERATION  SCHOOL_WEIGHT_PONDERATION  D_LS [D_LS]  D_PS [D_PS]  D_S [D_S]  D_S [D_S [D_S]  D_S [D_S]  D_S [D_S [D_S [D_S]  D_S [D_S [D_S [D_S]  D_S [D_S [D_S [D_S]  D_S [D_S [D_S [D_S [D_S [D_S [D_S [D_S [	height Cases	<b>X</b>
OS [D_NS]         OS [D_NS]           OS [C_S]         C_S [C_S]           OS [C_S]         C_S [C_S]           MATH [MATH]         Current Status: Do not weight cases           OK         Paste           Reset         Cancel           Help	STUDENT_JD_ELEVE [STUDENT_J           TEACHER_ID_ENSEIGNANT [TEAC           SCHOOL_D_ECOLE [SCHOOL_ID           SCHOOL_D_ECOLE [SCHOOL_ID           SCHOOL_D_ECOLE [SCHOOL_ID           SCHOOL_D_ECOLE [SCHOOL_ID           SCHOOL_SIGNERS[SCHOOL_ID           SCIENCES [SCIENCES]           D_LS [D_LS]           D_PS [D_PS]           D_SS [D_S]           C_SS [C_SI]           C_SR [C_SR]           MATH [MATH]           READING_LECTURE [READING_LE           S002A [S002A]	Do not weight cases         Weight cases by         Frequency Variable:            STUDENT_WEIGHT_PONDERATION          Current Status: Do not weight cases         Reset       Cancel         Heip

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.

$\boldsymbol{c}$ ·	CDCAD	A 1 ·	1	•	$\mathbf{C}_{1}$	<b>TI</b> 7 • 1
I omnarison	OT PLAP	Achievement	wnon	using	NTUAONT	WPIGHTS
Comparison			which	using	Sinachi	neignis
1	./					0

1 0		0	0		
	No Student Weights Used		Student Weights Used $(n=??)$		
	( <i>n</i>	=??)			
_	Mean	Standard	Mean	Standard	
		Deviation		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.