

## Lab 4: Distribution of Sample Means and Hypothesis Testing

Over the past three labs you have become relatively familiar with data from the 2013 Pan-Canadian Assessment Program (PCAP) developed by the Council of Ministers of Education, Canada. You ran some descriptive statistics on students' demographic and performance data, as well as standardized the performance data to better understand those scores. In this lab, we will focus on using the data to answer some research questions!

### 1. Setting up

Quantitative methods are typically used for confirmatory analyses. This means, after we decide on a research question, the next step is to set up our hypothesis (usually this hypothesis is based on literature).

Some literature has suggested that students' socioeconomic statuses (SES) are linked to their achievement performances. Specifically, they suggest that higher SES are strongly related to higher achievement scores. We will now test this theory.

Although PCAP 2013 did not explicitly ask students to rate their SES, it did ask questions (e.g., SQ08. About how many books are there in your home?) that could be inferred as an observable measure of SES. The research question that you will answer is: Do students who are at a higher SES, as measured by the number of books in their home, perform better on a science achievement test? (Note: this is a directional question.)

### Questions:

1. Please state, using one brief sentence, the null hypothesis and alternative hypothesis.

H<sub>0</sub>: \_\_\_\_\_

H<sub>1</sub>: \_\_\_\_\_

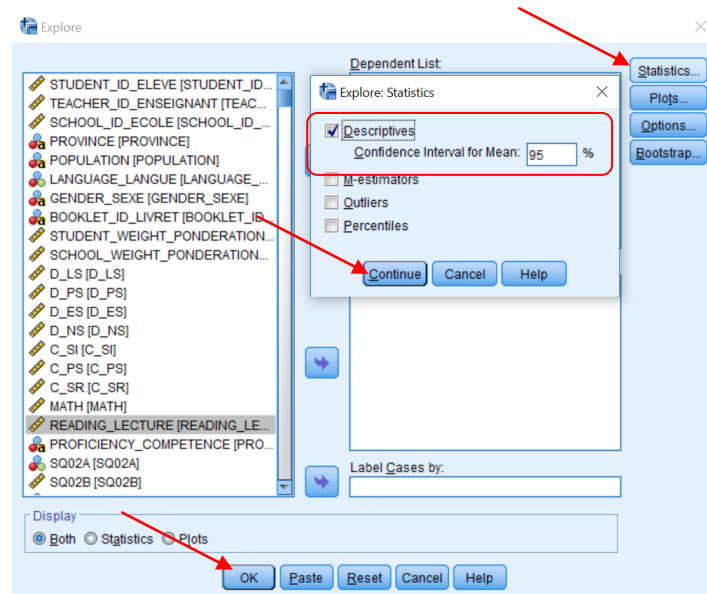
2. Please indicate your dependent and independent variables.

### 2. Understanding our data

There are many ways to statistically analyze the data to look for differences. In this lab, we will focus on using lower and upper bounds of the 95% confidence interval for means. The first step is to understand our data based on the SES groups. Please create a table of the descriptive statistics (i.e., sample size, mean, variance, 95% confidence interval [lower and upper bounds], minimum, maximum, range, and interquartile range) for each SES group (e.g., 0 to 10 books, 11 to 25 books). To do so, please select Analyze → Descriptive Statistics → Explore → move the dependent variable into the Dependent List box and move the

independent variable into the Factor List box → click Statistics → ensure Descriptives is checked → ensure Confidence Interval for Mean is set to 95% → click Continue → click OK.

*\*Please ensure your student weights are used throughout the analyses.\**



The table below may be helpful as you develop your table.

Table 1.

*Title (please make a title for this table)*

Books	<i>n</i>	Mean	Variance	Minimum	Maximum	Range	Interquartile Range	95% Confidence Interval for Mean	
								Lower	Upper
0 to 10									
11 to									
25									
26 to									
100									
Etc.	Etc.	Etc.	Etc.	Etc.	Etc.	Etc.	Etc.	Etc.	Etc.

**Questions:**

3. What does the 95% confidence interval for mean represent? (You may need to use Google or your textbook for this item.)
4. Please provide the Figure of the box plots. On the Figure, please label the interquartile range, maximum, minimum, and possible outliers. (You may need to use Google to

determine the differences between minimum and outliers.)

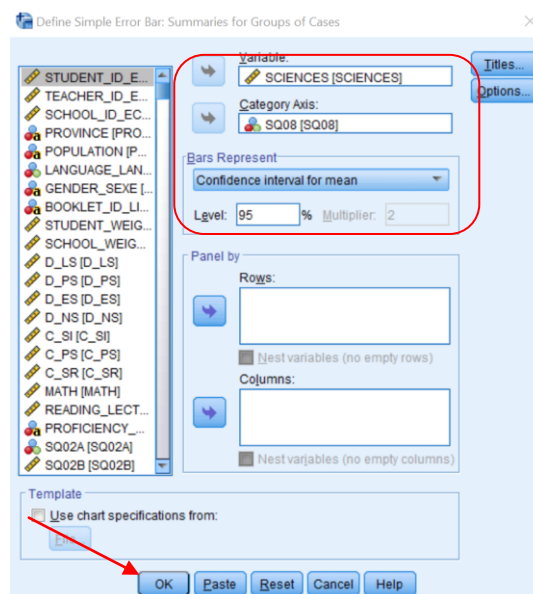
### 3. Answering the research question

Although the above Figure provides a descriptive analysis of students' achievement scores based on their SES, it still does not indicate whether there are statistically significant group differences. To investigate this difference, we need to analyze the upper and lower bounds of the 95% confidence interval.

#### Question:

- Looking at the 95% confidence interval for mean, do any of your intervals among the SES groups overlap?

The best way to determine whether the intervals overlap is to look at the actual values, such as the ones that you presented in Table 1. However, a picture is easier to communicate trends that might be evident. To do this, go to Graphs → Legacy Logs → Error Bar → select Simple → click Define → move Science into the Variable box and SQ08 into the Category Axis → ensure Confidence Interval for Mean is set to 95% → click OK.



#### Questions:

- Does your new Figure support your answer from Question 5?
- What trend do you see from this 95% confidence interval graph?
- Please answer the original research question and support your answer using empirical evidence (e.g., references to statistics from Tables and/or Figures). This response should be a **maximum** of three sentences.

### 4. Your turn

You are reading some literature that suggests students' first language is related to their performance on reading assessments. You are interested in testing this theory using the PCAP 2013 data, because the data set has information regarding students' first language (i.e., SQ04) and reading achievement (i.e., READING\_LECTURE). The research question that guides this investigation is: Do students who speak different first languages perform differently on a reading achievement test? (Note: This is a non-directional question.)

**Questions:**

Please complete questions 1, 2, 5, 7, and 8 as well as include a Table of descriptives from sections 2 and 3 above. (Your answer to question 8 may be a **bit** longer.)

You have now completed the lab.