UNIVERSITY OF CALGARY
WERKLUND SCHOOL OF EDUCATION

Lab 9: Multiple Regression

In the previous lab, you worked on simple regression in which only one variable (i.e., bullying) was used to predict another (i.e., science achievement). In this lab, we will be investigating the use of two variables (i.e., bullying and learning environment) to predict science achievement; this is called ***multiple regression***. Theoretically, adding another or more predictors should improve a prediction. An example of this would be: using age to predict one's health versus using age, weight, blood pressure, and cholesterol level to predict one's health.

1. Correlation and internal consistency

In lab 8, you completed correlation and Cronbach alpha for the bullying items; in this lab you will complete the same analysis for the five learning environment items (i.e., SQ12A to SQ12E).

Questions:

1. What does the correlation table from the SPSS output tell you about the items?

2. Looking at the Cronbach's Alpha if Item Deleted column of the Item-Total Statistics table for the learning environment items, what is the best possible alpha value that could be achieved? Which item(s) would you recommend we combine in our mean to achieve this value?

3. If you recommend removing items in question 2, please look at the item you plan to remove (e.g., look at the student questionnaire or codebook) and speculate why the specific item(s) would reduce the internal consistency of the set of items?

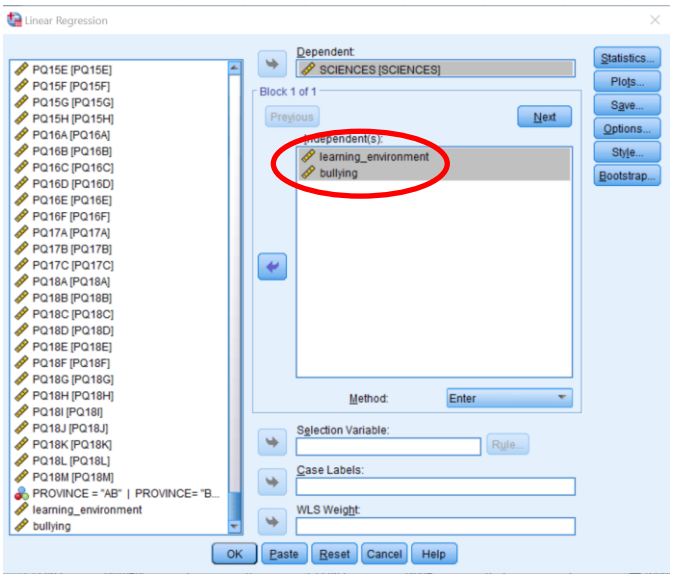
4. Please consider your results from both the correlation and internal consistency analyses to recommend which item(s) you would use to calculate a mean for the learning environment variable.

Using your results from question 4, please calculate the mean for the learning environment variable.

2. Multiple regression

A multiple regression uses the same procedure as simple regression, except for adding one

more predictor (attractiveness) into the Independence(s) box. To begin this analysis, please compute the means for both bullying and learning environment using your recommendations for items to include as stated in questions 3 and 4. Instructions on how to compute the mean of a set of items are located in lab 8 on page 3.



Questions:

5. Using the Model Summary and ANOVA tables, is the overall model (using both bullying and learning environment to predict science achievement) significant in predicting achievement? Please answer using the following sentence "bullying and learning environment together are [able/not able] to predict science achievement significantly using statistical analyses, r^2=[insert your R Square value], $F(df_{regression}, df_{residual})$=[insert your F-value], p[</>]0.05."

6. Now we need to look at each predictor variable used to see if both predictors are necessary (i.e., if bullying alone is able to predict science achievement just as well as using both predictors, then using bullying alone is preferred; simple models are preferred to complex models). Using the Coefficients table, please indicate whether each predictor variable is significant in predicting science achievement. Please answer using the following sentences: "Bullying is [able/not able] to predict science achievement significantly using statistical analyses, b_1=[insert your B value], t=[insert your t-value], p[</>]0.05. Learning environment is [able/not able] to predict science achievement significantly using statistical analyses, b_1=[insert your B value], t=[insert your t-value], p[</>]0.05."

7. Now we need to interpret the results. Please select [increased/decreased] and fill in the blank with the appropriate values (you may need to use Google for this item):

i) Controlling for learning environment, when bullying is increased by a unit, science achievement will be [increased/decreased] by _____units.
ii) Controlling for bullying, when the learning environment as reported by students is increased by a unit, achievement will be [increased/decreased] by _____units.

3. Chi-square test

When working with categorical variables, we need to use non-parametric tests because they often do not fit the assumptions (i.e., normality and homogeneity of variance). In this analysis, we will focus on the teacher survey. Specifically, we hope to answer the research question: Is there a difference between the number of science-related subjects taken by the teacher (i.e., TQ04) and their views of being a specialist as a science educator (i.e., TQ05A and TQ05B)? Therefore, we have two nominal variables: number of science classes taken in postsecondary (1–2/3–5/6–9/10+) and teachers' views of being a specialist as a science educator (education and experience yes/no).

To analyze the Pearson chi-square please click Analyze → Descriptive Statistics → Crosstabs → transfer each of the variables into the Row(s) and Column(s) boxes *(Note: it does not matter which set of variables goes into which box)* → click Display clustered bar charts → Statistics → click Chi-square, Phi, and Cramer's V → Continue → Cells → check Observed from the Counts box, and Row, Column and Total from the Percentages box → Continue → click OK.

Questions:

8. Using the Crosstabulation table, please indicate how many participants are in each of the eight categories (include percentages). (Hint, please show results for both TQ05A and TQ05B in the same table.)

9. Using the Pearson chi-square values, please indicate whether there is significant association between the variables using the sentence structure "There is [a/no] significant association between the number of science-related classes taken by a teacher and their view of being a specialist by [education/experience], χ([insert your df])=[insert your chi-square value], p[</>]0.05." You will need two sentences for this item since you have two items for TQ05.

10. What do Phi and Cramer's V values indicate (you may need to use Google for this item).

11. What do YOUR Phi and Cramer's V values (i.e., 0.775 and 0.376) indicate about the relationship between the variables? Often, reporting the graphs makes it easier for non-statistically savvy readers to understand the relationship between the variables.

This is the last lab for our course.