

## INSTRUCTIONS:

This assignment requires you to perform and interpret a series of analyses using SPSS. For this assignment you will be using the National Longitudinal Survey of Children & Youth (1998-1999, Canada) data

This assignment is to be word processed and presented in a well-organized document. Copy-and-paste the SPSS output that you produce into a word processing file. Six percent of the overall grade for this assignment will be awarded for presentation and organization. Simply stated, your assignment must be easy to read and follow. Each analysis should be clearly numbered.

Please include an appendix at the end of your assignment where you present the syntax commands for each analysis that you perform, including 'select cases' and 'recodes'. (See the handout titled 'SPSS Commands II' for an example.) The syntax must be presented in order and with each set of commands clearly labeled. Failure to include this appendix will result in a 20% grade deduction.

**1. Produce a histogram or boxplot for each of the variables listed below (be sure to title each graph). For numerical variables, indicate the mean, the approximate mode, and describe the variability as well as the shape of the distribution. For categorical variables, describe which attributes have the greatest and least frequency. Also, for each variable, indicate the specific level of measurement (i.e., nominal, ordinal, interval or ratio).**

(a) Province of residence (CGEHD03)

(b) Age of child (CMMCQ01) 2

(c) Average number of alcoholic beverages (CHLScQ5A)

(d) Perception of neighbourhood (CSFHQ02)

(e) Social support score (CSPHS01)

(f) Average number of alcoholic beverages (CHLScQ5A) - Select for people who live in Ontario who hold a 'college or university degree' (use variable CEDSD02 to select for education) - Describe how this distribution differs from the same variable distribution for the full sample

(g) Select one numerical variable and one categorical variable of choice. For each of your chosen variables, produce a histogram or bar graph as appropriate. In each case, provide a written interpretation as instructed above. USE THE FULL SAMPLE WHEN PERFORMING THESE ANALYSES. (It is expected that your chosen variables will be different from that of other students in the class. Two assignments featuring the same two variables for this question will be treated as a potential case of academic misconduct.)

**2. EXPLORE 'child got along with other kids' (CRLCQ06) by 'aggression score' (CBECS09).**

(a) Produce the requisite 'explore' output.

(b) Describe the overall trend regarding the relationship between the two variables. As we move through categories on the 'child got along with other kids' variable, what happens to the mean aggression score? What happens to the variability?

(c) Interpret the histograms, stem & leaf plots, and box plots. Comparing categories on the 'child got along with other kids' variables, describe the variability, skewness, and medians for the 'aggression score' variable.

(d) Provide an overall explanation that summarizes the nature of the relationship between the 'child got along with other kids' and 'aggression score' variables. Does there appear to be a relationship between the two variables? If so, how strong does the relationship appear to be? Justify your answer.

**3. (A) Task: to produce a two-way Chi-square analysis that examines the relationship between a newly created categorical version of the 'months when started to read to child' variable and 'highest level of schooling obtained' (CEDPD02).**

i. RECODE the existing 'months when started to read to child' variable (CLTCQ06B) into the following three (3) categories:

- 3 months or younger
- Between 4 and 6 months
- 7 months or older

Give your new variable an appropriate name and apply the value labels (note: there will be a significant point reduction for not properly applying the value labels). **Use this newly created variable in the Chi-square analysis.**

ii. Indicate a research question for the Chi-square analysis. Identify the logical independent variable and dependent variable.

iii. State the null and alternative hypotheses (in words).

iv. Produce the two-way Chi-square analysis using SPSS. v. Interpret the percentages in the cross-tabular table. Looking at the percentages, does it appear that a relationship exists between the two variables? Explain.

vi. Interpret the Chi-square statistics (most pointedly the p-value). Can you reject the null hypothesis? (Clearly indicate why or why not.)

vii. Write a brief research conclusion that summarizes the relationship between the two variables.

**(B) Produce a two-way Chi-square analysis using two appropriate variables of your choosing. Note that you must select variables that could plausibly share an empirical relationship (think carefully about this; selecting two variables that theoretically, and from a common sense standpoint, have little or nothing to do with one another will result in zero points awarded to your answer for this question). Follow steps ii – vii listed above for Q3A.**

If you perform any data manipulation for the purposes of this analysis (e.g., selecting cases or a variable recode), indicate this and briefly explain why you did it.

(It is expected that students will select different variable combinations for this analysis. That noted, we recognize that coincidences are possible. If you produce an analysis that features the same variables as that of another student, we will scrutinize both assignments for evidence of academic misconduct.)

**4. Task: to produce a bivariate regression analysis that examines the relationship between 'emotional disorder- anxiety (4-11 yrs)' (CBECS08) and 'social support score' (CSPHS01).**

- (a) State a research question. Identify the logical independent variable and dependent variable.
- (b) State an alternative hypothesis (since we are not conducting a significance test, a null hypothesis is not required). Your hypothesis should predict the strength and direction of the relationship. Provide an explanation as to why you are predicting the relationship that you have hypothesized.
- (c) Produce the bivariate regression analysis using SPSS. Include a scatter diagram with a regression line imposed (be sure to provide a title for your graph).
- (d) State the regression equation for this analysis (start by indicating the theoretical regression equation formula). Clearly and explicitly explain what a (intercept) and b (slope) mean in this particular research context.
- (e) Interpret the correlation coefficient and  $r^2$  for this analysis. (Specifically, making reference
- (f) to these statistics, indicate the strength and direction of the relationship between the two variables). State a final conclusion regarding the relationship between these variables (in this statement, indicate whether you have support for your hypothesis).
- (g) Suppose you did find a strong, negative correlation between 'emotional disorder- anxiety (4-11 yrs)' and 'social support score' (forget about the actual result). Would you be able to infer a direct causal relationship between these two variables? Explain why or why not with reference to the three criteria for causality. Identify at least three (3) extraneous variables that might influence the relationship between 'emotional disorder- anxiety (4-11 yrs)' and 'social support score'.

**\*INCLUDE AN APPENDIX WHERE YOU PRESENT THE SPSS SYNTAX COMMANDS FOR EACH ANALYSIS\***