SOC 364 RESEARCH REPORT #3 (of 3)

General Instructions

- The exercise is open book, open notes. You may discuss this assignment with anyone _before you begin your computer work_, after which you may only ask questions of the instructor. (Feel free to ask _any_ question, but recognize there are some I can't answer.)

- Write "I have completed this assignment on my own, without assistance from anyone other than the instructor" at the top, and sign your name to pledge observance of this restriction.

- There is no time limit other than the due date and time. Late penalties _will_ be enforced.

- Leave yourself plenty of time for both the computer work and answering the questions. The computer analysis will take anywhere from an hour to several afternoons, depending on how well you plan ahead. It may take several more hours to write your answers.

- Save your work early and often, keeping in mind the guidelines to which you have agreed: You are responsible for computer problems, which do not excuse you from deadlines.

Managing Your SPSS Output

- Submit your work in hard copy (i.e. printed, not emailed and not on diskette).

- Delete large sections of irrelevant output, though err on the side of caution.

- Type your answers directly into the output file, using INSERT – TEXT; or copy-and-paste the output you want to use into a word processing file where you're typing your answers

- Intersperse text among the tables to which it refers, rather than putting everything at the bottom, but do not retype the tables. Organizing your findings is part of your assignment.

Answering the Questions

- Be certain to answer all questions listed; points are allotted for everything asked!

- Make certain that you have produced all the output you need before typing your answers.

- Explain each step, formula, and decisions clearly, and interpret each result.

- Answer in prose form (sentences and paragraphs, not just numbers or short phrases) and try to come as close to English explanations of the meaning of each number as possible.

- Do not simply repeat statistics from the output, and do not simply put a question and a number into a sentence; interpret the results of each exercise. Use them to tell a story.

- Do not forget that all numbers are measured in specific units. Be clear what the unit of analysis is, and mention it wherever appropriate. (For example, not just "8" for EDUC, but "eight years of education" – and not just "2" for SEX, but "female".)

- Do not answer questions for variables that do not apply to that particular question.

Getting Started

- For the first report, you _described_ data (using univariate techniques). For the second report, you made _estimates and inferences_ (using ideas about probability distributions). For this third report, you will attempt to _explain relationships_ between variables (using bivariate techniques appropriate to each level of measurement).

- This third report has more questions than the first two, but should take you less time, for three reasons: First, you should now be comfortable working in SPSS at a slightly faster pace. Second, SPSS now does almost all the work for you, so you won't need to calculate much – just get the output, and "write up" the results. Third, some of the questions are the same on every report, so you have had experience answering some of them twice already.

- You will need six variables – two nominal, two ordinal, and two interval – from any data set. If none of the data sets available are to your liking, tell me and I'll find you something else. If there's a topic you'd like to study and don't have data on it, tell me and I'll help you find it.
  - In a more advanced course, such as 497, you would have gained the skills and the responsibility to acquire data on your own. For this course, the focus is on using the data, and I'd rather you not get distracted with trying to find or collect your own.)
- Be certain to resolve any missing values.
  - Note: You might choose to study the missing cases, as a nominal variation – for instance, those who refused to give an opinion about a certain topic, or those who declined to give their age – but make sure there are enough cases in that category for a bivariate analysis.
  - Also, if you *still* aren't sure what to do about missing values,see slides 8, 14, and (especially!) 15 of the lecture on Indices; and http://www.csun.edu/~egodard/364/faqs.shtml#missings.

I: Introduction

Define your population of interest, and define the available sample. What is the sample size, and what is the population size? Comment on any strengths or weaknesses of the sample for your study, including the sample size, any biases you might suspect, any advantages or disadvantages of the sampling procedure, and anything you would change about that procedure.

II: Nominal Association

a) For your two nominal variables, state an argument about how they might be related. Make a case that there is a causal relationship between them (specifically stating which is the independent variable and which is the dependent variable, even though you will be using a symmetric measure) and state expectations about the form of that relationship.

b) For each nominal variable, identify its variable name, variable label, operational definition (including value labels, if appropriate), and level of measurement, noting any changes introduced in recoding. Provide a concise (brief but complete) univariate analysis of each variable. Pay special attention to missing values. (If you have two variables with many missing cases, you may not have enough cases which are valid for *both* variables.)

c) Briefly (a few sentences) describe the pattern and size of any relationship observed in the crosstabulation, using a comparison of modal percentages.

d) Conduct a statistical test for the hypothesis concerning your two nominal variables, using an alpha of 0.05. List all steps taken and all assumptions made in testing the null hypothesis, including statements of both hypotheses (in Ho and Ha notation *and* prose explanations), *interpretations* of the test statistic and the p-value, and a sound and complete decision regarding both hypotheses. (A *general* interpretation of the test statistic is satisfactory). You do *not* have to compute the test statistic by hand, nor do you need to address percentage comparisons.

e) Assess the strength of the relationship between the two variables, using an appropriate measure. Also, provide an interpretation of this measure in terms of prediction errors.

III: Ordinal Association

a) For your two ordinal variables, state an argument about how they might be related. Make a case that there is a causal relationship between them (specifically stating which is the independent variable and which is the dependent variable) and state expectations about the form of that relationship.

b) For each ordinal variable, identify its variable name, variable label, operational definition (including value labels, if appropriate), and level of measurement, noting any changes introduced in recoding. Provide a concise (brief but complete) univariate analysis of each variable. Pay special

attention to missing values. (If you have two variables with many missing cases, you may not have enough cases which are valid for *both* variables.)

c) Briefly (a few sentences) describe the pattern and size of any relationship observed in the crosstabulation, using a comparison of modal percentages.

d) Assess whether there might be a dependent relationship between these two variables. You do *not* have to compute the test statistic by hand (nor, for *this* question, do you need to address percentage comparisons) nor do you need to list all steps and assumptions involved, nor do you have to specify hypotheses ... but you *must* report the significance level of the test statistic, interpret that value, and make a conclusion about dependence.

e) Assess the strength and direction of the relationship between the two variables, using gamma to address this association. Also provide an interpretation of this measure in terms of prediction errors.

IV: Interval Covariation

a) For your two interval variables, state an argument about how they might be related. Make a case that there is a causal relationship between them (specifically stating which is the independent variable and which is the dependent variable) and state expectations about the form of that relationship.

b) For each variable, identify its variable name, variable label, operational definition, and level of measurement, noting changes introduced in any recoding. Provide a concise (brief but complete) univariate analysis of each variable, accounting for any missing values. (If you have two variables with many missing cases, you may not have enough cases which are valid for *both* variables.)

c) Make a scatter plot of the relationship between these two variables. Give a general description of the plot – does it suggest that a relationship exists, and if so what type does it suggest? (Make certain to make all possible inferences.) Are there any outliers evident in the diagram?

d) Report the parameters ("y-intercept" & "slope") of the regression equation, explain their meanings in general terms, and give an interpretation of the particular statistics calculated from your data.

e) *Extra credit:* What is the value of the standard error of the estimate? Give a general interpretation of this statistic, and tell how it assesses the efficiency of your estimator of the slope.

f) Calculate (by hand) predicted values of the dependent variable for two values of the independent variable. Interpret these two predicted values, label them on the scatter plot, and plot the regressions line between these two points.

g) *Extra credit:* Which two statistics demonstrate the extent to which the dependent variable is affected by the independent variable? Give their values and interpretations, and say how they differ in meaning.

h) Is the relationship statistically significant? How do you know? What is the null hypothesis? Give the value of the appropriate test statistic and its significance level, and interpret both. (You need not conduct a full hypothesis test – answer only these questions.)

i) Interpret the correlation coefficient for the association of these two variables. How does this statistic differ in meaning from the regression coefficient (for the "slope")?

j) *Extra credit:* What is the difference between the Sum of Squares of Regression and the Sum of Squares of the Error (in words, not just the statistical difference)? What is the sum of these two, and what does that sum assess?

k) Report the PRE statistic that describes how strongly your values of your independent variable predict values of your dependent variable. Provide both a general interpretation of this measure *and* an interpretation of your observed (calculated by SPSS) statistic.