

SOC364 RESEARCH REPORT #2 (of 3)

This page is the generic stuff...

General Instructions

- This assignment 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.)
- There is no time limit other than the due date. Late penalties will be enforced.
- 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.
- Answer all questions listed, in the order given. Points are allotted for everything asked!
- Leave yourself plenty of time for both the computer work and answering the questions below. The computer analysis will take anywhere from an hour to several afternoons, depending on how well you plan ahead, how erratic your typing is, etc. It may take several more hours to write your answers to all the questions. Remember that you can copy any (or all) output into a Word document, which you can then open and edit on other computers (even without SPSS).
- Save your data early and often, and keep in mind the general guidelines to which you agreed: You are responsible for computer problems, which do not excuse you from deadlines.

Managing Your SPSS Output

- Hand in both the computer output and *typed* answers in hard copy (i.e. printed, not emailed).
- 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 your text among the tables to which it refers, where possible, rather than putting everything at the bottom, but do not retype the tables. Organizing your findings is part of your assignment.
- Answer questions 1 and 2 at the top of your file, question 3-5 under the output relevant to each variable, and question 6 at the bottom of your file.

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*.
- 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.
- Do not simply put statistics from the output into a sentence. Use them to tell a story.
- Answer in prose form (i.e. sentences and paragraphs, not just numbers or short phrases) and try to come as close to the English explanations of the meaning of each number as possible.
- 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:

Before answering the questions below, you will need to formulate research hypotheses. You should plan to have your values approved by me, in lab or via email, at least one week before the report is due.

You will need both an hypothesis (some research which suggests what some value “ought” to be, what it is elsewhere, or what it might be) *and* some way of testing it (that is, data for appropriate variables.)

- Read through the entire assignment first, so that you know what you’ll be doing.
- Go to the web or the library to find a hypothesized value for each variable. *Do this soon!!*
- Keep track of what the values measure and for whom, and the source of those values. If the source is in print, photocopy the page with the value; if you get the value from a web page, print that page and copy down the URL (web address).
- The hypothesis for your nominal variable should generate a one-tailed test; the hypothesis for your interval variable should generate a two-tailed test.
- If you will have fewer than 50 valid cases for your nominal variable, be sure to formulate a research hypothesis for your nominal variable that does not specify a *hypothesized* proportion less than 0.1 or greater than 0.9, since those would require a sample size greater than 50.
- You may use formulas we have discussed to make certain that you have a sample size sufficient for the test you wish to conduct. You are responsible for any problems that result from choosing a test too demanding for your sample size.

Research Instructions:

You will analyze data from the National Health and Social Life Survey (NHSL.SAV).

- Use the file that you created for HW3, which you were instructed to save at that time. If you did not save it, or have not yet completed HW3, you will need to go back and (re)create that file.
- Note that you will all have different data, since you all have different samples. You thus better all have different results and answers, since you are all to be working *on your own!*
- Open your HW3.SAV file (or whatever you called the file where you saved your systematic sample of 300 NHSLS cases).
- Use FILE - SAVE AS to create a new file called REPORT2.SAV.

Pick two variables, one nominal and one interval:

- These variables need not measure or indicate the same concept, although they may if you like.
- You are free to combine variables (using COMPUTE) to make an index, but do not have to.
- Be certain that all variables you use have appropriate variable and value labels.
- Account for and resolve (that is, find and recode) any missing values. (Note: That does *not* mean, get output that shows cases that have already been labeled as missing and then call me to remind you that those are missing cases. Those are missing cases. You are asked here to take care of any unresolved missing *values*, which means to look at the frequency table and see if there are values that are *not yet identified* as missing but which *need* to be so SPSS won’t include them in analysis and calculations.)

Quick Summary:

- Intro and Sample size test
- Hypothesis Test *and* Confidence Interval, each for *both* a nominal *and* an interval variable
- Questions 3-6 are *univariate*, *not* about relationships between the two variables.

QUESTION 1 (Introduction):

- Briefly introduce your report, explicitly stating your research question and relevant concepts. You need not treat this as a short paper, but may wish to cite work read in other classes or from sociology (or other *academic*) journals.
- Define your population of interest, and define the available sample. What is the sample size, and what is the population size? (Note: The population of interest is *not* the full NLHLS dataset. There would be no reason to make inferences to a set of respondents we already know, and don't particularly care, about.) 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.
- For each variable that you selected, identify its variable name, variable label, operational definition (including value labels, if appropriate), relationship to the research question (i.e. concepts) chosen, 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.

QUESTION 2 (Determination of Sample Size):

- For the interval variable used in questions 5 and 6, what is the *minimum* sample size you would need to come within a specified range of error 99% of the time?
- Describe each number plugged into the formula, and explain what the sample size n that you computed represents. Be sure to specify a margin of error beforehand, and to be explicit about justifying why you chose it. Assume that your sample statistics are reasonable estimates for the unknown in this future-oriented calculation; i.e. use the sample standard deviation since you do not have the population's.

QUESTION 3 (Nominal Hypothesis Test):

- Conduct a one-sided statistical test for the hypothesis concerning your nominal variable, using the 0.05 level of significance.
- List all steps and assumptions you would need in testing the null hypothesis, including statements of both hypotheses, show computations, interpret the test statistic and the p -value, make the decision and fully indicate why you made the decision, and indicate what this decision says about your research question or how it might affect your addressing that question.
- Be sure to attach a copy of the data used to formulate your hypothesis and state the source of the hypothesized value, the sample studied by that source, and the population of interest to that source and assessed by that sample. Include in your citation (if this information is not included on your printed or photocopies attachment) the title, year of publication, publisher, author if relevant, page of table if relevant, and table number if relevant.

QUESTION 4 (Nominal Confidence Interval):

- For the nominal variables used in question 3, report the sample proportion for a category of the variable and then construct a 95% confidence interval for this category.
- Describe each number you plugged into the computation, and interpret the confidence interval.
- Be certain to clearly justify use of the category you select.

QUESTION 5 (Interval Hypothesis Test):

- Conduct a two-sided statistical test of the hypothesis concerning your interval variable, using an alpha of 0.01
- List all steps and assumptions you would need in testing the null hypothesis, including statements of both hypotheses, show computations, interpret the test statistic and the p-value, make the decision and fully indicate why you made the decision, and indicate what this decision says about your research question or how it might affect your addressing that question.
- Be sure to attach a copy of the data used to formulate your hypothesis and state the source of the hypothesized value, the sample studied by that source, and the population of interest to that source and assessed by that sample. Include in your citation (if this information is not included on your printed or photocopies attachment) the title, year of publication, publisher, author if relevant, page of table if relevant, and table number if relevant.

QUESTION 6 (Interval Confidence Interval):

- Construct a 99% confidence interval around the sample mean of the variable used in question 5.
- How is this interval consistent with the result of your hypothesis test, or is it not? How do you explain that inconsistency? Why *should* they be consistent, and why *might* they contradict? If they contradict, do you have reason to consider one more persuasive than the other?

QUESTION 7 (Subsample Confidence Intervals):

- Construct *separate* confidence intervals for the interval variable, one for each of two groups (or categories) of your nominal variable. (If your nominal variable has more than two categories, you need not analyze the others.)
- You will need to use Select Cases twice. (Be sure to reselect “All Cases” in between, and to use Filter and *not* Delete.)
- Be sure to justify your choice of test categories: why would the distribution (center or spread) of the interval variable differ between these two groups? If you find that it does not differ, account for that.
- *Be sure to begin with univariate analyses* (shape, central tendency, and dispersion) of the separate distributions for each of the groups you will analyze.
- Are there any unanalyzed categories of this variable that might differ as well? If so, how & why?

QUESTION 8 (Difference of Means):

- Test for a difference of means for your interval variable between the two chosen nominal categories. Include all steps, and make explicit conclusions both about variation in the distributions and about what sort of relationship you believe exists between the variables. (Warning: we are not yet analyzing correlation, association, or causation, so *do not* make those sorts of claims.)

Fall 2009: Mystery Measurement #3 is.... 27. See the MM page on the website for updates!