

PSC 101(11): Scope and Methods of Political Science
Data Analysis Assignment I
Due: February 15, 2005

[General Instructions: Question 2 is worth 10%; the others are worth 30% each. For questions 1 and 2, paste your tables, figures, or graphs into a word processing document.]

1. **Interpreting Cross-tabulations:** [*Stata command: tab rowvar colvar; Stata pulldown: statistics/summaries/tables, & tests/tables/two way tables with measures of association/check either within column or within row relative frequency box*]

Identify two variables from the NES 2000 dataset linked on the PSC 101 webpage. The codebook is also linked on the page, and you can flip through the codebook to guide your choice of variables. Both variables should take on between 2 and 5 values, inclusive. Therefore, the size of the table you generate can range from 4 cells (2 x 2) to 25 cells (5 x 5). Generate the table with either row or column percentages, paste it into your assignment write-up, and answer the following questions:

- a. Briefly describe the meaning of the two variables as described in the codebook. (For example: The row variable describes the respondent's party identification; the column variable describes the respondents gender.)
 - b. Choosing an appropriate measure of central tendency, what are the central tendencies of the row and column variables?
 - c. For one of the rows (if you show row percentages) or one of the columns (if you show column percentages), describe, in words, exactly what the row (or column) percentages mean for each of the percentages in the row (or column).
 - d. Does there appear to be a relationship between the row and column variables, in the sense that knowing the value of one variable gives you information about the value of the other variable? Why or why not?
2. **One-way graph:** Using the same dataset, graph the Clinton Thermometer score (v000359) as a histogram and paste it into your document. [*Stata command: hist v000359; Stata pulldown: graphics/easy graphs/histogram*] Describe the distribution, in words. This question asks respondents to place President Clinton on a "feeling thermometer" ranging from 0 (bad) to 100 (good). Do you notice anything noteworthy about the types of responses that people tend to give? If so, what?
3. **Central Tendency and Dispersion:** Enter the following numbers into your statistical package, giving you 10 observations: 10, 14, 23, 17, 32, 18, 11, 18, 6, 22. Given these data, answer the following questions:
 - a. What are the mean, median, and standard deviation for the variable?
 - b. You can add one observation to the data but want to keep the mean unchanged. What number do you add?
 - c. You can add one observation to the data but want to keep the median unchanged. What number do you add?
 - d. What happens to the standard deviation in parts (b) and (c)? Why?
 - e. If you wanted to *increase* the standard deviation in the data by adding an observation, roughly where should add the observation? Why?

4. **Working with the Normal distribution:** You are given a set of data with mean 100 and standard deviation 20. Assume that the data are normally distributed, and answer the following three questions. Parts b and c will require the use of a normal table such as Appendix D of Johnson and Reynolds (p. 484).
- a. What is the probability that an observation will be greater than 100?
 - b. What is the probability that an observation will be greater than 115?
 - c. What is the probability that an observation will be less than 90?