POS 3713: Assignment 3
Assigned: Monday, 2/21/2000
Due date: In class, Monday, 2/28/2000
Tutorial Session: Thursday, 2/24/2000, 9am-10:45am; and Friday, 2/25/2000, 9am–10:45am
Professors Mitchell and Lubell

The purpose of this assignment is to introduce you to measures of dispersion and provide practices using the normal curve and computing confidence intervals. You will use SPSS for each part of this assignment. Remember to type up your responses to each question and to either 1) include the output directly in your text, or 2) attach your printed output from SPSS.

We will be using the 1996 National Election Study data, so you should begin by opening the file NES1996Edit.sav in SPSS. To launch the program, click on the Start button, select Programs, and then click on SPSS Windows 9.0. If you saved this file in Assignment 1 on your floppy disk, simply open that saved file on your disk (Click on "File", "Open", and then find the file on your disk). Otherwise, you will need to go back to the web site (ftp://www.coss.fsu.edu) to download the data again. When you are finished with this assignment, be sure to save your working data file again by clicking on "File", and then "Save As".

Note that we will not be giving detailed instructions for basic operations that were covered in previous assignments such as renaming and labeling variables, recoding missing data, producing frequency distributions, and measures of central tendency. Remember to make sure SPSS is not using values that should be considered missing data, such as Don't Know (DK) or Not Appropriate (NA) in the calculation of measures of central tendency and dispersion. You can completely eliminate missing values by recoding them as system-missing, using the data preparation procedure discussed at the beginning of Assignment 2. However, in some cases missing values in the NES have already been defined as "user-missing". SPSS treats "user-missing" values as missing values when calculating statistics, but does not completely eliminate them from the dataset. If you run a frequency distribution for a variable, you can check which values are defined as user-missing. For example, in the frequency table below the values "8=DK" and "9=NA/RF" are user-missing. If there are values that should be considered missing values, but that are not defined as user-missing, you should recode those values into missing.

96PR:ANGRY--CLINTON AFFECT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Yes	900	52.5	52.7	52.7
	5. No	809	47.2	47.3	100.0
	Total	1709	99.7	100.0	
Missing	8. DK	4	.2		
	9. NA; RF	1	.1		
	Total	5	.3		
Total		1714	100.0		

Part A: Index of Qualitative Variation and Standard Deviation

Measures of dispersion allow you to compare how much heterogeneity there in a distribution of responses to a survey question. One can interpret the dispersion in a survey question tapping attitudes as the amount of agreement among survey respondents. Variables with higher levels of dispersion have more disagreement.

Exercise A1

Calculate the mean and Index of Qualitative Variation (IQV) for v960421: Pre. Does R consider Clinton intelligent, and v960423: Pre. Does R consider Clinton moral. According to the IQV, do respondents think Clinton is more intelligent or more moral? Which of Clinton's personality traits do people agree upon the most? Speculate on why agreement among respondents is different for each variable. You answer should include SPSS output for descriptive statistics and frequencies, and a paragraph interpreting the results.

- Rename v960421= *Intellig* and v960423 = *Moral*
- Produce a frequency distribution for *Intellig* and *Moral*
- Check to make sure missing values are coded correctly
- Produce descriptive statistics for *Intellig* and *Moral*
- Calculate the Index of Qualitative Variation using the formula on p.92 of Healey

Exercise A2

Produce the range, variance, and standard deviation for v960272: Bill Clinton Feeling Thermometer (named *ClinTher* from previous exercises) and v960281: Hillary Clinton feeling thermometer (named *HillTher* from previous exercises). What are the technical/mathematical differences between the three measures of dispersion? According to the measures of dispersion, do people have more similar feelings for Bill or Hillary; in other words, do people agree more about Bill or Hillary. Explain your interpretation fully with reference to the measures of dispersion.

- If you haven't renamed the variables to *HillTher* and *ClinTher* from previous exercises, rename them now
- Check to make sure missing values are coded correctly
- Select "Analyze", "Descriptive statistics", then "Descriptives" to open the "Descriptives" window
- Move *ClinTher* and *HillTher* into the variables text box
- To produce range and variance, click on the "Options" button in the descriptive statistics window, place checkmarks in the boxes next to range and variance, and then click "Continue"
- Click "OK" in the "Descriptives" window

Part B: Using the Normal Curve

The properties of the normal curve allow you to calculate the probability that a score will fall into a particular range within the distribution of scores. Remember, to accomplish this task you must first convert the score of interest into a Z-score. Then, you can use the Normal Curve Table in Appendix A of Healey to find the correct probabilities.

Exercise B1

Using the standard deviations and means produced in exercise A2 and the procedures described in Chapter 5 of Healey, calculate the following probabilities **for both** *ClinTher* and *HillTher* (the probability should be expressed as a proportion, not a percentage):

- Probability that the score on the feeling thermometer will be greater than 75; e.g., prob (*ClinTher* > 75) = ? and prob (*HillTher* > 75) = ?
- Probability that the score on the feeling thermometer will be less than 25; e.g., prob (*ClinTher* < 25)= ? and prob (*HillTher* < 25) = ?
- Probability that the score on the feeling thermometer will be greater than 30 and less than 70; e.g. Prob (30< *ClinTher* < 70)= ? and Prob (30< *HillTher* < 70)= ?

Exercise B2

Using the SPSS procedures described below, calculate the 95% confidence intervals for *HillTher* and *ClinTher*. The information relevant for this question is reported in the first three rows of the table. The first row provides the mean and standard deviation of the variable. The second and third rows tell you the lower and upper bounds for the 95% confidence level. You can ignore the remaining information in the Descriptives table. Use the information in the table to answer the following questions.

1) Describe in words how to interpret the 95% confidence intervals for *HillTher* and *ClinTher*.
2) Calculate the 95% Confidence intervals for *HillTher* and *ClinTher* assuming the sample size increases by 1000, but the mean and SD stay the same (you need to do this by hand using formula 7.2 in Healey, Chapter 7, page 161. Remember this formula is used when you are calculating confidence intervals from a sample as opposed to a population). How does the increase in sample size increase the accuracy of your estimate of the population mean?

- Click on Analyze, Descriptive Statistics, Explore. Move the variables *HillTher* and *ClinTher* into the "Dependent List" box.
- Click on OK. This will produce a descriptives table with the information you need. Note that if the number is reported as 3.35E-02, this means you move the decimal point two places to the left; the number is actually 0.035. Or if the number was 5.43E-03, you would move the decimal point three places to the left; the number is .00543.