SPSS Assignment 2
Using Percents, Means, Medians, Range, Standard Deviation, and/or Interquartile Range to Describe the Characteristics of A Sample

Due Date: Wednesday, October 23rd by 4pm (if I am not in my office, you should leave your assignment with a member of the Sociology Dept. office staff on the 4th floor of Brewster Wing B).

In this assignment, you will use SPSS to analyze the data in one of the datasets on your disk. Keep in mind that you have three to choose from:
1. studentsurvey.sav - the full dataset from the surveys we conducted of students.
2. healthdataset.sav - a subset of the Community, Crime, and Health Survey which includes demographic information on the respondents as well as information about their health, exercise, access and use of health care, etc.
3. crimedataset.sav - a subset of the Community, Crime, and Health Survey which includes demographic information on the respondents as well as information about crimes they have experiences, perceptions of crime in their neighborhood, whether they own a weapon, and whether they have committed a crime in the last 12 months.

I suggest that you decide which dataset most closely matches your interests and work with that dataset on this and other assignments to follow.

Your goals for this homework assignment are to write up a description of your sample in terms of key demographic variables available in the dataset (see below) and then compare your sample to the population from which it was drawn. In order to meet these goals, you will need to 1) assign missing values for all the required demographic variables, 2) recode or compute new variables based on old variables where necessary or desired, and 3) generate frequencies and/or measures of central tendency, measures of range. See instructions attached on how to do these things in SPSS.

Your homework assignment should include the following sections with subtitles. You should include your name at the top of the assignment. Your assignment must be typed, double-spaced, with one-inch margins, and in a standard 11-12 point font.

Methods: A description of the methods used to collect the data you are using, the population, what kind of sample you have, how the sample was generated/selected, and the sample size. Much of this is in the codebooks I gave you on your disk and/or the survey instructions for the student survey. You may paraphrase from these sources.

Variables: You should describe the demographic questions asked of each respondent which generated the data you are using to describe your sample (see below). In your description of your variables, you should note the range of the responses given by the respondents on each Ordinal, Interval, and Ratio variable.

Description of Sample: In paragraph form, you should describe your sample on the demographic characteristics required (see below), using appropriate statistics given the level of measurement of each question. When you use a mean or median, you should provide appropriate evidence for why this is the appropriate statistic. By the end of this description, we should know what the typical respondent was like and how much variation there was in the responses for each question.

You should turn in your homework assignment with a copy of the output used to generate the statistics reported. You may make use of tables if you prefer. However, a table cannot substitute for an actual write-up of your results. In other words, you must “translate” that table into written form. See the course website for an example of a write-up of this kind.
Demographic Information You Will Need to Include:

*If you are using the Student Survey Data:*
1. a description of the sex of the respondents in the sample
2. a description of the representation of minority respondents in the sample*
3. a description of the representation of international students in the sample
4. a description of the age distribution of the sample
5. a description of the sample in terms of year in school
6. a description of the sample in terms of in-state or out-of-state residency
7. a description of relationship statuses of students in the sample (ie. do they have a boyfriend/girlfriend, spouses, etc)

*If you are using the Crime or Health subset data:*
1. a description of the sex of the respondents in the sample
2. a description of the representation of minority respondents in the sample*
3. a description of the age distribution of the sample
4. a description of the education of respondents in the sample
5. a description of the incomes of respondents in the sample
6. a description of the number of dependent children among the households of people in the sample
7. a description of the number of adults among the households of people in the sample

*Note: For the purpose of this study, we will define anyone who says they are white and not hispanic as a non-minority group member. Anyone who identifies as a race other than white and/or identifies as hispanic will be identified as a minority group member. You may notice that none of the datasets include a minority variable. They include variables that identify the person’s race and whether they are hispanic. You will have to create a variable called minority. See instructions below.*
Things You May Need to Be Able to Do in SPSS:

Open A Dataset
There are three possible ways to open a dataset in SPSS:

1. In Windows Explorer, doubleclick on the dataset you want to use. Remember that SPSS saves datasets with a .sav extension. So, anything on your disks with a .sav extension is a dataset. This should include:
   - studentsurvey.sav
   - healthdataset.sav
   - crimedataset.sav

2. If you doubleclick on the SPSS icon, you should get a dialogue box that asks you what you would like to do. Instead of choosing “Type in data” like you did on the last assignment, on this assignment you want to choose “Open an Existing Data Source” and then click on the “More Files” option under this command (if your dataset does not show up in the list below). Change to the directory on which your datasets are saved (a:\; if you are working off the disk I gave you). Then doubleclick on the dataset you want to use.

3. If you are in SPSS, click on File, then Open, then Data, and choose the directory on which your data is saved, and the dataset you want to open.

No matter how you go about opening your data, you should see the data in front of you, with columns that include variable names at the top. These variable names should correspond to the variable names in your codebook (or on the student survey if that is what you are using). You may find it useful to have a copy of the codebook or survey printed out to refer to. (Note: If you have not already done so, I would suggest that you save your dataset on a different disk or with a different name so that you always have a “clean” copy of the data somewhere.)

Assigning Missing Values
Keep in mind when you are having the computer calculate measures of central tendency (mean, median, mode) that you need to have told the computer ahead of time what values should be considered missing values. If you don’t it will count a missing value as a valid response. This could really throw off your results. For example, in the crime subset, if the computer thought that 97 was a valid response for the question on number of adults in the household, it would figure out a mean number of adults thinking that some people had 97 adults in their household (BIG PROBLEM!). For this reason, you should make sure that all missing values have been assigned for the variables you will use. To check this, run frequencies on your variables (see instructions below) and make sure any invalid answer is listed in the missing rows. If you need to assign a value as missing, go into the variable view and assign missing values like we did for the first SPSS assignment (ie. click in the missing values column for the variables you will use and fill in the numbers which should be considered missing or not valid).

Running Frequencies
If you want to know how many or what percentage of people said what on a particular variable or variables, click on Analyze, then Descriptive Statistics, then Frequencies. You will now be in the Frequencies dialogue box. On the right hand side will be a box with a list of all the variables in the dataset. Double click on each variable you want frequencies for and it will move into the box on the right hand side. Do this until you have all the variables you want information on in the right hand side box. Then click on OK. You will now get an output file with the information you asked for.

Generating Measures of Central Tendency
If you want to generate the means, medians, and/or modes for particular variables, click on Analyze, Descriptive Statistics, and Explore. You will now be in the Explore dialogue box. On the right hand side will be a box with a list of all the variables in the dataset. Click on each variable you want information on and then click on the arrow next to the Dependent Variables box. Do this until you have all the variables you want
information on in the Dependent Variables box. Then click on OK. You will now get an output file with the information you asked for (and then some).

**Computing a New Variable (ex. Creating a Minority Variable)**
You will need to compute a variable for minority status no matter which data you are using. You may also want to compute other new variables that recode some of the original variables in new ways.

To compute a new variable, first make sure you are in the data view window (not output or variable view). Then click on Transform, then Recode, then Into Different Variables. (Note: Although you can recode an existing variable, we generally avoid doing that so that we don’t screw up our original variables.) You will now be in the Recode Into Different Variables dialogue box.

**Minority Example for Student Dataset:**
In this case we want to start by recoding the race variable, so doubleclick on the RACE variable in the left-hand list of variables so that it moves into the right hand empty box (You will see “race ->?” now in that box.) Now click into the Output variable box and give your new variable a name. In this case we could call it MINORITY, for minority versus not minority. Click in the Label box and fill in a variable label to remind you of what this new variable is (ex. Respondent is minority or not). Now, click on the CHANGE button and you should see “race -> minority” in the box to the left. Now we need to tell the computer what the values of this MINORITY variable should be based on the original values in the RACE variable. Click on the Old and New Values button. In this case, we can click on the Range: _______ Through _______ option on the left side of the box under “Old Value” and fill in 2 in the empty box before Through, and 5 in the empty box after Through. Then click in the New Value box on the Top Right and type 1. Finally click on the Add button. You should now see “2 thru 5 -> 1” in the box on the right side. This tells the computer that if the person had a race coded 2, 3, 4, or 5, they should be coded as 1 in the MINORITY variable. Now, go back to the left side and click on the Value button and fill in1 in the blank. Click in the New values box and type in 0, then click on the Add button. Now, in addition to “2 thru 5 -> 1”, you should also have “1 -> 0” in the box to the right. Now you can click on Continue, and then OK in the original dialogue box. You should now see that you have a new column called Minority at the end of your dataset, which has values 0 and 1.

Now we want to bring in the people who identified as hispanic in the next question. Go back to Transform and now choose Recode into Same Variable. Doubleclick on the MINORITY variable in the left-hand list of variables so that it moves into the right hand empty box. Now click on the If button. Click the button next to If the Case Satisfies Condition (this will turn the back ground in the list of variables to white). Now doubleclick on HISPANIC variable in the list so that it moves in to the empty white box. After the word hispanic type = 1. You should now see “hispanic=1” in the white box. (Note: This is telling the computer to do whatever you tell it next to anyone who is coded 1 for hispanic.) Now click on the Continue button. You will be back in the first dialogue box. Now click on Old and New Values. Under Old Value click Value button and then fill in 0. Now fill in 1 in the New Values box. Don’t forget to click on the Add button. You should now see 0->1 in the Old ->New Values Box on the right of the box. Now click on Continue and then OK.

Double check your work by running frequencies on the original RACE variable, the original Hispanic variable and the new MINORITY variable. The number of minority students should add up to at least the number of students designated as categories 2 thru 5 on the RACE variable.

**Minority Example for Crime and Health Datasets:**
First we want to recode race, so doubleclick on the RACE variable in the left-hand list of variables so that it moves into the right hand empty box (You will see “race ->?” now in that box.) Now click into the Output variable box and give your new variable a name. In this case we could call it MINORITY, for minority versus not minority. Click in the Label box and fill in a variable label to remind you of what this new variable is (ex.
1. The default setting in SPSS is to generate these lists of variables by variable label (not name) and in the order of the dataset. I find it easier to find variables I'm interested in if I change this so that the list shows up in alphabetical order by variable name (ie. the eight letter name we gave each variable). If you want to do this, get out of frequencies and click on Edit, then Options. This will take you into a series of dialogue boxes. The top box should be the General box. On the right hand side, top of the General box, you should see two things under Variable Lists. Select “Display Names” and “Alphabetical”. Then click on OK. The bad part is that it will only become active the next time you open SPSS, so close SPSS down and then Start over. Now all your variable lists should be in alphabetical order by variable name.

Respondent is minority or not). Now, click on the CHANGE button and you should see “race -> minority” in the box to the left. Now we need to tell the computer what the values of this MINORITY variable should be based on the original values in the RACE variable. Click on the Old and New Values button. In this case, we can click on the Range: ______ Through ______ option on the left side of the box under “Old Value” and fill in 2 in the empty box before Through, and 6 in the empty box after Through. Then click in the New Value box on the Top Right and type in 1. Finally click on the Add button. You should now see “2 thru 6 -> 1” in the box on the right side. This tells the computer that if the person had a race coded 2, 3, 4, or 5, they should be coded as 1 in the MINORITY variable. Now, go back to the left side and click on the Value button and fill in 1 in the blank. Click in the New values box and type in 0, then click on the Add button. Now, in addition to “2 thru 6 -> 1”, you should also have “1 -> 0” in the box to the right. Now you can click on Continue, and then OK in the original dialogue box.

Now we want to bring in the hispanics who identified as hispanic in the next question. Go back to Transform and now choose Recode into Same Variable. Doubleclick on the MINORITY variable in the left-hand list of variables so that it moves into the right hand empty box. Now click on the If button. Click the button next to If the Case Satisfies Condition (this will turn the back ground in the list of variables to white). Now doubleclick on HISPANIC variable in the list so that it moves in to the empty white box. After the word hispanic type =1. You should now see ‘hispanic=1” in the white box. (Note: This is telling the computer to do whatever you tell it next to anyone who is coded 1 for hispanic.) Now click on the Continue button. You will be back in the first dialogue box. Now click on Old and New Values. Click in the System or User-missing button and then fill in 1 in the New Values box. Don’t forget to click on the Add button. You should now see MISSING->1 in the Old ->New Values Box on the right of the box. Now click on Continue and OK.

Double check your work by running frequencies on the original RACE variable, the original HISPANIC variable and the new MINORITY variable. The number of minority respondents should add up to at least the number of respondents designated as categories 2 thru 6 on the RACE variable.

Printing Output
When you have all the Output you need, get into the Output Viewer window (either by choosing it from the bottom of your screen or by going to Window at the top of the screen and choosing Output. Then click on the Printer button (or File, then Print). (Note: If you have extra analyses within your output that you don’t need to hand in for this assignment - ex. you played around with something, you can click on that part of the output in the outline to the left of your output viewer window and then hit the delete button.)

If you work on the assignment in stages and want to keep adding to your output file before you print it, you can save your output file (like any other file). It will save with a .spo extension. All .spo files are SPSS output files. Then the next time you get on the computer to work on the assignment, after you have opened the data file, go up to File, Open, Output, and open your output. Then the computer will continue to put the results of any subsequent data analysis in that output file.

NOTES:
1. The default setting in SPSS is to generate these lists of variables by variable label (not name) and in the order of the dataset. I find it easier to find variables I’m interested in if I change this so that the list shows up in alphabetical order by variable name (ie. the eight letter name we gave each variable). If you want to do this, get out of frequencies and click on Edit, then Options. This will take you into a series of dialogue boxes. The top box should be the General box. On the right hand side, top of the General box, you should see two things under Variable Lists. Select “Display Names” and “Alphabetical”. Then click on OK. The bad part is that it will only become active the next time you open SPSS, so close SPSS down and then Start over. Now all your variable lists should be in alphabetical order by variable name.