Gov 50 October 2012

Describing Data/SPSS

Recitation Handout *TFs: Erin Baggott, Michael Gill, Allen Schmaltz*

1 What is SPSS?

It's original name was "Statistical Package for Social Sciences" and it was written by a political scientist. It is now one of the most widely used statistical software packages, especially in the private and government sectors.

1.1 Is this the only statistical package?

No. Matlab, Stata, R, SAS, etc., are all examples of other computing programs, each of which have their own special properties and capabilities. But in this class, we'll just focus on using SPSS.

2 Getting the Data loaded

Fire up SPSS. It's easiest to cancel out of the little wizard that pops up. You should be staring at what looks like a blank spreadsheet. Now, to get the our data up and running do the following:

1. Go to https://dl.dropbox.com/u/2042671/Gov50Survey.sav and save the file to the your Desktop. You should see the `Gov50Survey.sav` file on the desktop.
2. Now, back in SPSS, load the data by going to the File > Open > Data menu in the top left corner. [N.B. you should also be able to load up the data by double-clicking on the `Gov50Survey.sav`.]
3. Go to Edit > Options. Under the General tab, you should make sure that the first two settings are "Display names" and "Alphabetical."

2.1 How to explore your data

Once you load the data, you'll see what appears to be a filled-in spreadsheet; we call this the Data View. Each row of this data represents a _____, which is the _____ of this dataset.

- What is the age of respondent 21? _____
- What kind of phone does respondent 21 own? _____
- What is the gender of respondent 74? _____
- What is the gender of respondent 75? _____

2.2 How to explore your variables

Each column of the Data View is a *variable*. We can find out more about each variable by going into the Variable View, by clicking on its tab in the bottom left hand corner. We don't care about most of these columns, but a few are going to be important to us.

- Name - the name of the variable in the dataset.
- Label - a slightly longer description of the variable.
- Values - the category labels for values of the response.

Gut check:

- What is the label for the `pink` variable? _____
- What does a response of 2 represent for the `candy` variable? _____

3 Describing the data

In order to get a feel for the data, we may want to calculate some sample _____ and plot some bar charts or histograms. To do this, go to the Analyze > Descriptive Statistics > Frequencies..., which should bring up a dialog box. You'll see two lists, one on the left with all of the variables in the data and one on the right, which is empty. In order to do calculations on a variable, move it to the list on the right.

SPSS works like a restaurant; you order some output and it will give it to you after it works for a second. In the Frequencies window, you can "order" different output from the top right. Statistics... controls what kind of sample statistics you will get (mean, median, mode, standard deviation, quartiles, etc). Charts controls what kind of graphs you will get (bar plots, histograms, etc). Notice in the bottom-left corner there is an option for displaying frequency tables.

3.1 Example 1 - Other School

1. Variable name: `otherschoo`
2. Label: _____
3. Level of measurement: _____
4. Which measure of central tendency should we use?

   mean median mode
5. Which type of chart should we use?

   bar chart histogram pie chart
6. What is the mode? _____
7. What percent of the sample would have gone to Yale? _____

3.2 Example 2 - Shoe Size

1. Variable name: `shoe`
2. Label: _____
3. Level of measurement: _____

4. Which measure of central tendency should we use?

   mean median mode

5. Which type of chart should we use?

   bar chart histogram pie chart

6. What is the median? _____

7. What is strange about the distribution of our data? _____

3.3 Example 3 - Death Tax & Estate Tax

1. Variable name: `deathtax, estate tax`
2. Label: _____
3. Level of measurement: _____
4. Which measures of central tendency should we use?

   mean median mode

5. Which type of chart should we use?

   bar chart histogram pie chart

6. What is the mean and standard deviation? _____
7. What is the mode for each variable? _____
8. Do people appear to support the Death Tax and Estate Tax with equal frequency? _____

4 Histograms

Making histograms with a specified number of bins is a little tricky.

1. Go to Graphs > Chart Builder. After a warning message, you should see a window with variables on the left and some different types of charts on the bottom.
2. Double click on the simple bar chart (it has three tan bars). This should put a little histogram in your "chart preview."
3. Find your variable (`e.g., basage`) and drag it to the x-axis.
4. On the right you should see a window called Element Properties. There is a box in it called "Statistics" – here you should see a pull-down menu with "Histogram" selected. Click on the Set Parameters button below that.
5. Under Bin Sizes, click Custom and choose the number of intervals. Close this and hit Apply on the Elemental Properties window.
6. Back on the Chart Builder window, you can hit OK to plot your histogram!

5 Worksheet

1. What percentage of the sample are college students? _____
2. What percentage of the sample said they support the men wearing pink? _____
3. What was the average guess of the `temperatur` variable? _____
4. For this question use the bathroom cleaning variable (`bathroom`):
    - What is the *mean* bathroom score? _____
    - What is the *median* bathroom score? _____
    - What is the *standard deviation* of the bathroom score? _____
    - Which direction is the *skew* of the variable? _____
    - What are the *first and third quartiles* of the bathroom score? _____
    - Do you clean your bathroom more or less than your average classmate? _____
5. Create a bar plot of for the "future career plans" variable (`career`) along with a frequency table. Open a Word document and put these two items in the document.