

Sociology 601: **Homework assignment 1**, Due September 10.

Questions from Agresti and Finlay:

Chapter 1): 3, 6.

Chapter 2): 3, 5, 6, 12, 18, 20, 22.

Chapter 3): 1(a,b only), 10(a,b,c only).

Practice for midterm: Answer the following question as if you were answering it on an exam (Do it by yourself, and try to finish within 15 minutes.) Then correct your work with help from others or your text.

S1.) A schoolteacher wondered how much time his 2<sup>nd</sup> grade students spent watching cartoons. One day, he asked 10 of the students how much time they spent watching cartoons the previous day, to the nearest hour. Here are their responses, arranged in order:

$$y_1 = 0 \text{ hours}$$

$$y_2 = 0$$

$$y_3 = 0$$

$$y_4 = 1$$

$$y_5 = 1$$

$$y_6 = 2$$

$$y_7 = 2$$

$$y_8 = 3$$

$$y_9 = 3$$

$$y_{10} = 8$$

**(Please use complete sentences.)**

A.) Define a statistic of interest.

B.) Calculate the value of your statistic of interest (yes, this is jumping ahead to chapter 3 a bit)

C.) Define a “safe” population of interest that is well-represented by the sample.

D.) Define a “general” population of interest that might not be so well-represented by the sample.

E.) Name one potential source of non-sampling variability, and describe its potential effects in this case.

F.) Name a probability sampling method the teacher might have used to obtain this sample, and describe how he would have done it.

Computing assignment for homework 1: Downloading and Charting Data from the General Social Survey.

This assignment will give you practice using demographic data from the internet. You will turn in answers to the eight numbered questions.

Recent news reports have shown interest in two child families, considered the norm in the United States, and other family sizes. Non-normative family sizes could include only-child families, three-child families, larger families with four or more children. Of course, some adults have no children. Suppose you decide to look at family sizes for your term project. This assignment shows how you might begin.

A.) Go to the website for the Survey Documentation and Analysis Archive (SDA Archive) at <http://sda.berkeley.edu/>

- click on “SDA ARCHIVE”

B.) The SDA archive has data sets on elections, racial issues, health, and other topics. For this assignment, we will use the General Social Survey, or GSS.

- click on “GSS CUMULATIVE DATAFILE 1972-2006”

C.) I suggest you browse the codebook before you make any tables. If you click the buttons for “codebooks” and “standard codebook”, you can shift from the codebook to the tabulation window.

D.) For this assignment, you will use the variables for age (“age”), year (“year”), and number of children (“childs”). Browse the codebook to answer these questions. I recommend the alphabetical variable list.

- 1.) What is the range of valid answers for “childs”, and how many respondents have missing information on “childs”?
- 2.) What is the range of valid answers for “year”, and how many respondents have missing information on “year”?
- 3.) What is the range of valid answers for “age”, and how many respondents have missing information on “age”?

E.) Now that you have seen the codebook, go back to the window for the SDA frequencies and crosstabulation program for the GSS cumulative datafile 1972-2006

You need to tell GSS what variable you would like to study. The variable “childs” for number of children will be our main, or Row, variable.

There will not be a second or third variable for a crosstabulation. Leave the Column and Control fields blank.

We would like the latest data, and we would like to limit our analysis to people who have completed their childbearing (more or less). We will use the Selection Filter field to limit the analysis to the year 2004 and adults in their forties.

The GSS has begun using a weighting variable, to adjust for differential nonresponse and household size. Let’s go ahead and keep that default weighting variable for now.

Row                      childs  
Column

ControlSelection Filters

year(2006) age(40-49)

Weight

wtssall - With adjustment for Nadults for all years

- click on “Run the Table”. A new screen will appear with the tabulation. At any time, you can close the screen and go back to the main window for frequencies and cross-tabulations.
- 4.) Describe the frequency distribution for the variable “childs”. In other words, what is the count of respondents for each possible answer?
- 5.) Describe the relative frequency distribution for the variable “childs”. In other words, what is the percent of respondents for each possible answer?
- 6.) What happened to the scores for the people who gave no valid answer?

F.) Now you are going to move the information from the frequency distribution from SDA to excel. You can cut and paste the frequency distributions from SDA to excel, or if you like to go low tech, simply type in the categories numbers, and percentages in excel.

When you are finished entering the data, you should have a set of rows and columns that look like this (The numbers shown are from GSS 2004 – yours will differ):

	D	E	F	G	H
1	number of children	frequency		number of children	relative frequency
2	None	113		none	18.4
3	one	99		one	16.2
4	two	197		two	32.1
5	Three	135		three	22
6	Four	36		four	5.8
7	Five	21		five	3.4
8	Six	6		six	1
9	Seven	2		seven	0.3
10	Eight	5		eight	0.8

(Note, the top row will look like a single line in excel, and the words “number of children” might be partially hidden. This is okay. Also, you might have put the variables in different columns than DEFG and H. This is okay.)

G.) Use the following instructions to make a one-way bar chart of the distribution of completed family size in the United States.

- a.) Use a mouse (click ‘n’ drag) or keystrokes to highlight a rectangle of cells that includes a column of categories and a column of relative frequencies for the variable “childs”. In the example shown above, you would highlight from G2 to H10. (We don’t need to graph the row for “total”.)
- b.) Use the mouse to click on the menu button with a red, blue, and yellow chart, called the chart wizard.
- c.) In step 1, click on chart type column, and the first sub-chart type, then next. (This is probably the default chart.)

- d.) In step 2, click on data range series in columns. (Again, this is probably the default chart. You might click on rows first to see what rows would look like, then go back to columns.) Then click next.
- e.) In step 3 in the Titles menu, provide an informative title and a label for the y axis. (“Homework 1” is not an informative title. Because this is a relative frequency distribution, the y axis is for percent.) For this chart, you do not need to label the x-axis, but you will often label both axes.
- f.) In the Legend menu, turn off the “show legend” indicator. Do not click next, but instead click the button for the Data labels menu.
- g.) In the Data Labels menu, click the button to turn on the “Value” for each bar. Then click next.
- h.) In step 4, just click finish.

To see what the bar chart looks like at this point, pull down the file menu and click on print preview. You may notice that some things need fixing. Click on close.

H.) Use the following instructions to change fonts, formats, and otherwise improve your chart.

- a.) Double click on a part of the plot area without any bars on it, or else single click and press CTRL-1. Hopefully, the format plot area window will open. Click on none for border and none for area, then click OK to return to the chart.
- b.) Double click on the x-axis to get the format axis window to open. In the font menu, click off the auto-scale and select 10 for size. Click OK to return to the chart.
- c.) Double click on the y-axis to get the format axis window to open. In the font menu, click off the auto-scale and select 10 for size. In the scale menu, click off the auto for major unit and set its value to 10. Click OK to return to the chart.
- d.) Double click on one of the bars to get the format data series window to open. Under the “patterns” tab, select a shade of gray for area. Under the “options” tab, set both “overlap” and “gap width” to zero (thus creating a relative frequency *histogram*). Click OK to return to the chart.
- e.) Double click on the chart title to open the format chart title window. In the font menu, click auto scale to off and set size to 16. Click OK to return to the chart.
- f.) Double click on the y-axis title to open the format axis title window. In the font menu, click auto scale to off and set size to 10. Click OK to return to the chart.
- g.) Double click on one of the data labels (numbers) to open the format data labels window. In the font menu, click auto scale to off and set size to 8. In the Number menu, set category to “number” and set the number of decimal places to “0”. Click OK to return to the chart.

To see what the histogram looks like again, pull down the file menu and click on print preview.

7.) Print your chart.

8.) Write a paragraph that will inform the reader of a key finding from your chart. Be sure to describe the numbers you are comparing and some possible social implications or explanation for what you found.