Bouma, Soc 286: Sociology of Health   **Handout #2- Data Analysis**

Investigating Differences in Disability by Race, Sex, Age, and Income

<u>GETTING STARTED:</u>

1) To get started, go to the website:
   http://ssdan.net/datacounts/webchip

2) You'll see a box to the left that says "Choose Dataset."   Click on the down arrow under Collection to open up possible data sets.

3) Choose "acs2010"

4) Under Dataset, click on the down arrow and scroll down to choose "EldrDisab".

5) Again, click on the down arrow and under the pop down screen, choose "EldrDisab.dat"

A) FREQUENCIES: A frequency table gives an overall sense of the distribution of a particular variable or set of variables. **To examine the frequencies of each variable:**
   1) **Click on Command**
   2) **Under the pop-down screen, click on "marginals**."

You will see a table for each variable. For example, for race/ethinicity, you will see the following table:
RaceEth

NhWhite	Black	Asian	Hispanic	AmIndian	NHOther	NHMulti
80.18%	8.4%	3.56%	6.63%	0.46%	0.08%	0.7%
30,475,611	3,191,572	1,354,623	2,520,162	173,616	29,855	264,564

Note that the first row of data gives percentages, and the second column gives the total number. Thus, in the US, in 2010, 30, 475,611 of the elderly were Non-Hispanic White, which represents 80.2% of all elderly. Generally, for comparison purposes, we talk about percentages, rather than raw numbers.

The data set contains seven variables: age, race, sex (gender), poverty status, and two measures of disability – limitations to independent living (IndLivLmt) and difficulties in self care activities of bathing and dressing (Selfcare). Examine the frequencies (marginals) for each of the variables to get a basic understanding of the people in this data set.

Now consider the variables you have:
   A) List the variables which are most likely independent variables:

   B) Which are dependent variables?

   C) Which could be either/both?
Draw out several relationships you could examine and make hypotheses. For example:
        Sex → selfcare
        Hypothesis: a higher proportion of women will report self care difficulties than men.

To investigate this hypothesis, you will need to run a cross-tab.

CROSS-TABS (Bivariate Analysis): To run cross-tabs, you will need to tell the program which variable is your dependent variable, which is your independent, and how to percentage your tables (across or down). Remember the following:

INDEPENDENT VARIABLE (X) - the variable that influences or affects another variable

DEPENDENT VARIABLE (Y) the variable that is influenced by, or depends upon, another variable

Do the following:

     A) Scroll down the side menu to the section labeled "Choose Variables" on the left.

     B) Click the down arrow next to the box labeled "Row" and select a variable. For the purposes of this assignment, always **put your dependent variable in the rows**. So, for this example, choose selfcare for your row variable by highlighting "selfcare" and then hit OK.

     C) Now you will need to choose your **column variable**. For this assignment, always **put your independent variables in the columns**. So, choose gender as your independent variable by highlighting gender as the column variable and then hit OK.

     D) Now scroll down to the section labeled "Generate Table" and click on **"percentage down."**

     E) This should bring up your bivariate, cross-tab table.

     F) To look at the effect of race on selfcare, follow the steps above, this time choosing "RaceEth" as the column variable

Your goal is to examine the various relationships in this data set and write a paper that summarizes the differences you find in disability status. There are two different forms of disability and several bivariate relationships (or cross-tabulations) that you can examine. To begin, just have fun and play with the data. Run lots of cross-tabulations and see what you learn. After you've investigated your main relationships of interest, try to figure out what story you want to tell. Your overarching theme is differences in disability status by race, sex, poverty status, and age. Use your bivariate tables to tell a story about these differences. For Monday, 3/16, come to class ready to discuss your findings. You should be able to give a basic overview of which groups have the highest levels of disability. On either Monday or Wednesday, I will teach you how to run controls.

DATA ANALYSIS PAPER

For this assignment, you are asked to write a paper which analyzes differences in disability status among the elderly in the U.S. Your paper should follow the outline of a general research paper:

Introduction: You should start with a general introduction about disability among the elderly. This should be broad, interesting paragraph introducing the context for the reader to think about issues of disability Please do NOT start by saying "This paper is a data analysis about….".

Brief Literature review: You should refer briefly to studies that help the reader understand the issue. What have you learned about race, sex, age, and income differences in disability status? You should refer to published studies referenced in our book and other sources. When you discuss research on disability, be sure to explain how disability is defined in the studies you review. Also, make sure that you properly cite your sources – ask me if you do not know how to do this properly. I know that you are not an expert on disability issues, so I fully expect that you will be summarizing and paraphrasing (and therefore citing) others who are.

Data: For the data section, you need to simply describe the population you are examining. In other words, describe the marginals or frequencies to give a very brief overview of the age, racial composition, sex composition, and income distribution among the elderly. Also describe what percent of this population is disabled, and what you mean by this. You don't need to include every single statistic (in fact, please don't – you'll put me to sleep). Instead, try to summarize the most important information succinctly. This will take you a few tries. Try reading it aloud to your roommate or a friend. The point of this exercise is for you to become comfortable using and discussing basic statistics. You want to make this comprehensible and easily understood. Please be clear that this is NOT where you describe the bivariate/cross-tabulations – this is simply where you

describe your sample by giving an overview of the frequencies. For the disability variables, make sure to define these, so that the reader understands what you are examining.

Results: This is the heart of your paper. First, begin by briefly stating your hypotheses. What did you expect to find? Then explain whether and how the two forms of disability differ by race, sex, age, and income status. Who had the highest levels of disability and who had the lowest? Did your findings match what you'd learned from your literature review? Were your hypotheses correct? The goal here is to tell a brief, cohesive story, using data as your evidence. Your main focus is on how two different forms of disability may differ across race, age, sex, or income. For one set of bivariate tables, please also run a set of controls and explain your findings.

Tables: Your paper should include properly labeled and formatted tables. Here is an example:

Table 1: Difficulty in Independent Living by Poverty Status, 2010

Difficulty in Living Independently	Poverty	Near Poor	Middle	Comfortable	TOTAL
Yes	25.7%	23.7%	15.6%	10.1%	16.0%
No	74.3%	76.3%	84.4%	89.9%	84.0%
TOTAL	3,564,646	4,195,306	20,825,453	9,424,598	38,010,003

Source: wgtd 2008-10 ACS,SSDAN/U-Mich

Tables should always be numbered, titled, and clearly labeled (and please number your tables in the order in which you refer to them). **You should have totals at the bottom**.   To get theses, run the same cross-tab and then instead of choosing "percent down," choose "frequencies." You also need to have the source at the bottom of your tables. You can embed the tables into your paper or simply put them at the end of the paper. Just make sure that they are numbered and properly referenced in your paper. In other words, whenever you begin discussing the findings in a table, you should always begin by referencing the table. For example, "As seen in Table 1, ….." Do NOT simply add "see table 1" at the end of your description.

Remember the basic rules for reading tables:
- start with general statement without numbers (e.g. "A higher proportion of women are disabled than men.")
- back up claim with stats – and follow same direction of statement. For example, if you say "More women are disabled than men," then start with the higher proportion of women, not the lower proportion of men.
- when using stats, always say ___% of ___ (e.g. "10% of men"). Do NOT say "Men are 10%."
- when showing similarities across groups, use 'and"
- when showing differences, use "compared to" "while only" etc.
- have a summary statement at the end.
- Write as if you care about what you found – these numbers mean something!

Useful phrases:
- A higher proportion of … fall into the (lower income/higher income) category
- a disproportionately high/low number of … fall into the …
- … are overrepresented/underrepresented among those with disabilities

Conclusion: Discuss some of the reasons you think we find these differences. Why do we find these patterns? You might want to refer back to your earlier literature review. Did your findings confirm what you'd learned from other researchers?

Bibliography: Make sure to include the full citation for the sources you discuss in you brief literature review. To cite the data set, you can use the following:

Frey, William H. "2010 Elderly by disability, gender, race/ethnicity, and poverty, 2008-10 08 ACS, SSDAN/University of Michigan. 20108. http://ssdan.net/cgi-bin/webchip/webchip3-beta.cgi?webchip=datasets/pc/acs2010/EldrDisab.dat [date you got your information from the website].

FORMATING: Please use 1-inch margins, 12-point font, double-space, and number all pages.

Context Statement: This project was designed to introduce you to and give you practice with the following skills: read and report basic frequencies in a table, interpret numbers in a bivariate table, convert raw data into formatted tables, identify independent and dependent variables, analyze the relationship between two variables, tell a story using numbers. For your paper, please also include a "Context Statement." In a couple paragraphs, please describe your experience writing a paper based on data analysis. What did you learn from this experience? What did you find most challenging? Has this helped you feel more confident about working with numbers? In what ways would you like more help or practice? You may comment directly on the skills listed above or discuss any other issues relevant to this assignment.

DUE: Peer Review, Mon, 3/23. Bring typed draft to class – this counts for 10% of the grade of the paper. Please bring a printed copy so that a classmate can work directly on your paper (I will check these in class.) **Final Draft due Wed, 3/25 (I will not accept emailed copies).**

Rubric for Data Analysis Papers

	Grade	Strengths – what to focus on	Areas to improve:/typical problems
Peer Review 10			
Intro 5		Clear, lays out main points	Needs to be broader
Literature review 10		Relevant to paper, defines disability, discusses literature that reviews differences in disability status; brings in outside sources; properly cites sources	Info not relevant Didn't bring in sources Didn't cite properly (if improper citing or plagiarism, **paper fails**)
Data and Results sections:			
Data (describe frequencies of sample) 10		Clearly describes frequencies (marginals)	Didn't adequately describe frequencies
Differences in disability status by age, race, sex, and income 35		Uses language properly Tells a story using data from cross-tabs as evidence; tests at least one control variable	
Tables 15		Properly numbered, labeled, have totals and source at bottom, etc.	
Conclusion 5		Sums up findings, refers back to relevant lit, offers some reasons why we see patterns	
Works Cited 5			If improper citing/plagiarism – **paper fails**
Context Statement 5			