

Data Analysis Project- Handout 1

Investigating the Effect of Race, Sex, Age, and Poverty Level on Disability in the US

For the past couple weeks, we have been discussing disability in the United States. As we've learned, disability can be defined in multiple ways. In this exercise, we will examine two forms of disability noted by the Census Bureau, difficulty in self-care and difficulty in independent living. We will examine how these abilities differ by race, sex, poverty level, and age. The data come from the 2010 American Community Survey (ACS).

LEARNING GOALS:

A. Substantive: Students will be able to:

Discuss and write about how various forms of disability differ by race, sex, income, and age by examining 2010 American Community Survey data. The data shown in the tables for this worksheet come from <http://ssdan.net/datacounts/webchip>, acs2010, EldrDisab.

B. Methodological/Quantitative Skills: Students will acquire the following skills, related to data analysis:

1. Interpretation: Students will be able to:

- Read and report basic frequencies from a large data set
- Describe bivariate tables, both orally and in writing

2. Representation: Students will be able to:

- Take raw data in tabular form and create properly formatted and labeled tables

3. Application/Analysis: Students will be able to:

- Manipulate variables in a large data set using a basic statistical package
- Identify independent and dependent variables
- Form hypotheses about the relationship between two variables
- Analyze relationships between two variables as presented in bivariate table.
- Learn how to introduce and explain the effect of control variables.

4. Communication: Students will be able to:

- Present hypotheses and findings for the relationship between two variables in informal presentation and in a paper that tells a story about differences in disability by race, age, income, and sex, using numbers as evidence

5. Confidence: Students will feel more comfortable

- reading and discussing data from a table
- writing about and using percentages to make an argument

SAMPLE: The data, included as part of the American Community Survey, are from the 2008-10 3-year PUMS (Public-Use Microdata Sample) file. About 1 in 40 (~3 million) households is selected each year into an ACS sample, and the 3-year PUMS combines 3 years of those samples. Not every responding household makes it into the PUMS, which is a subsample, but those that do are weighted to represent the population in the end year of the release. That is to say that the 2008-10 ACS PUMS data are weighted to fairly closely align with the 2010 population estimates. Thus these data should fairly accurately represent all elderly aged 65 and above in the year 2010.

VARIABLES: For this exercise, we will be examining the following variables: race, sex, poverty level, age, ability to care of oneself, and ability to live independently. Although there are many ways in which each could be conceptualized, the following definitions are those used by the U.S. Census Bureau:

SELFCARE DIFFICULTY (Variable name: *selfcare*) Having difficulty bathing or dressing

INDEPENDENT LIVING DIFFICULTY: (Variable Name: *IndLvLmt*): Because of a physical, mental, or emotional problem, having difficulty doing errands alone such as visiting a doctor’s office or shopping.

SEX (variable name: *Gender*) - individual’s self-identification as either male or female.

AGE (of elderly): - age divided into the following categories: 65-74, 75-84, 85+

RACE/ETHNICITY (Variable name; *RaceEth*) – individual’s self-identification as:

- **Non-Hispanic White (NHwhite)** – all persons who indicated their race as white and not of Hispanic origin.
- **Black** – all persons who indicated their race as black.
- **Asian (or Pacific Islander)** – includes all persons who indicated their race or ethnicity as Chinese, Filipino, Japanese, Asian Indian, Korean, Vietnamese, Cambodian, Hmong, Laotian, Thai, or other Asian as well as Hawaiian, Samoan, Guamanian or other Pacific Islander.
- **Hispanic**– persons of white or “other” races who identified themselves as Mexican, Puerto Rican, Cuban, or Other Spanish/Hispanic. This category can refer to ancestry, nationality group, lineage, or country of birth of the person’s parents or ancestors before their arrival in the U.S.
- **American Indian (AmIndian)** – all persons who classified themselves as American Indian, Eskimo or Aleut.
- **Non-Hispanic Multiracial (NHMulti)** – any Non-Hispanic persons who identified as more than 1 race
- **Non-Hispanic Other (NHOther)** – any Non-Hispanic persons of a single race that was not white, black, American Indian/Alaskan Native (AIAN), Asian or Pacific Islander

POVERTY LEVEL: *these are based on poverty thresholds published by Census Bureau each year and are determined by age, number of people in the household, and total household income. To give you an idea, I’ve included the general amounts for elderly living alone.*

- **Poverty:** living at <100% of 2010 poverty threshold (<\$10,458 for person 65 or above; living alone)
- **Near Poor:** living at 100-199% of the poverty threshold (\$10,459-20, 915; 65+; living alone)
- **Middle:** living at 200-399% of the poverty threshold (\$20,916 -\$41,831; 65+; living alone)
- **Comfortable:** living at 400% or greater of the poverty threshold (>\$41,832; 65+; living alone)

FREQUENCIES: A frequency table gives an overall sense of the distribution of a particular variable or set of variables. Here are the frequencies for the variable RaceEth (race) for all elderly in 2010.

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 row gives the total number. Thus, in the U.S., 30,475,608 elderly aged 65 or older in 2010 were Non-Hispanic white, which represents 80.2% of the elderly. Generally, for comparison purposes, we talk about percentages, rather than raw numbers. We see that about four-fifths of US elderly in 2010 were White, about 8% were Black, a little less than 7% Hispanic, about 4% Asian, and less than 1% were Native American, Multiracial, or of an unlisted race.

For fun, make a guess about the frequencies for the other variables listed above.

Now examine the frequencies for age, sex, poverty level, and disability below:

Gender

Male	Female
43.45	56.55
16,515,670	21,494,333

PovLevel

Poverty	NearPoor	Middle	Comf
9.38%	11.04%	54.79%	24.8%
3,564,646	4,195,306	20,825,453	9,424,598

AgeEldr

65-74	75-84	85+
54.96%	32.92%	12.12%
20,889,961	12,511,581	4,608,461

IndLvLmt [Because of a physical, mental, or emotional problem, having difficulty doing errands alone such as visiting a doctor’s office or shopping.]

Yes	No
16.05%	83.95%
6,099,843	31,910,160

Selfcare: [Having difficulty bathing or dressing]

Yes	No
8.48%	91.52%
3,221,925	34,788,078

Given the above frequency tables, please answer the following:

- What percentage of elderly are men? _____
- What percentage of elderly live in poverty? _____
- What percentage live between 100-200% of the poverty line (near poverty)?: _____ are considered “comfortable? (Live at 400% or more of the poverty line?) _____
- What percentage of elderly appear to be “disabled” in terms of not being able to dress or bathe themselves _____ What % are unable to run errands for themselves? _____

Now describe these frequencies in easily understood English. In other words, how would you describe the elderly in 2010?

Now you get to be the sociologists. You can see that about 16% of the elderly report independent living limitations, and 8.5% report difficulty bathing or dressing themselves. Who are they? Do they differ across sex or race? Are they the youngest or oldest? The poorest or wealthiest? Begin by making hypotheses about what you expect to find:

HYPOTHESES:

- 1) **SEX/GENDER:** _____ will have higher levels of (independent living difficulty or difficulty bathing/dressing) than _____.
- 2) **RACE:** _____ will have the highest levels of disability (decide how to define this) and _____ will have the lowest.
- 3) **POVERTY LEVEL:** People in the _____ income level will report the lowest levels of disability, and people in the _____ income group will report the highest levels of disability.

In order to investigate your hypotheses, you will need to do cross-tabulations, also called bivariate tables. In a cross-tabulation, you are simply exploring the association between two variables. Before running these, think about what you really want to know. It makes sense to say that a person's sex affects his or her disability status. It does not make sense to say that disability status affects a person's sex.

The following two definitions are important:

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

You can write the relationship between the two as $X \rightarrow Y$. In this case, we are interested in how sex influences disability. Let's specifically examine whether sex affects independent living difficulties. Another, less causal way of describing this is to simply say you want to examine differences in independent living difficulties for men and women. Sex would be the independent variable (X), and independent living difficulties would be the dependent variable (Y). In other words, to some extent, independent living difficulties depend on sex:

Independent Dependent
Variable \rightarrow variable

$X \rightarrow Y$

sex \rightarrow Disability Status (or more specifically, ability to live independently)

Useful language for describing the relationship between two variables:

We are looking at the relationship between sex and independent living limitations.

We are examining the effect of sex on independent living limitations.

Try to come up with 3-4 more statements for describing the above bivariate relationship:

Below is the bivariate table (cross-tabulation) of sex and ability to live independently:

Table 1: Percentage of Men and Women with Independent Living Limitations, 2010

Independent Living Limitations?	Male	Female	TOTAL
Yes	11.8%	19.3%	16.0%
No	88.2%	80.7%	84.0%
TOTAL	100%	100%	100%
	16,515,670	21,494,333	38,010,003

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

What is the take-home message of this table? Write a one-sentence general statement to describe this table:

Now write a couple sentences using statistics from the table to support the statement you wrote above.

Here's my attempt at describing the table:

From Table 1, we can see that a higher percentage of women have difficulty living independently than men. For example, 19.3% of women compared to only 11.8% of men report that, because of a physical, mental, or emotional problem, they having difficulty doing errands alone such as visiting a doctor's office or shopping.

Note how I have described the data. I started with a broad, generalized statement: "a higher percentage of women have difficulty living independently than men." Then I used specific statistics from the table to make my case. Note that I **percentaged down the rows**, and I **compared across the columns**. In other words compare percentages **across men and women**, not across the yesses and nos.

Some basic rules for reading tables:

- start with general statement without numbers (e.g. "A higher proportion of women are disabled than men.")
- Always reference the table at the beginning of your description (e.g. "As we see in Table 1"). Do NOT wait till the end of your paragraph to say "see Table 1".
- back up your 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 percentage, 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!

In the next bivariate table, we will examine racial differences in poverty. In this relationship, which is the independent and which is the dependent variable?

Independent variable: _____

Dependent variable: _____

Table 1: Income/Poverty Status by Race-Ethnicity among the Elderly, 2010

	NhWhite	Black	Asian	Hispanic	AmIndian	NHOther	NHMulti	TOTAL
Poverty	7.3%	19.4%	12.8%	19.0%	20.0%	10.2%	12.6%	9.4%
NearPoor (100-199% poverty threshold)	10.1%	15.9%	10.6%	16.1%	16.3%	11.0%	13.2%	11.0%
Middle (200-399% Poverty threshold)	55.9%	49.9%	46.2%	52.1%	50.2%	56.5%	53.7%	54.8%
Comfortable (>400% above poverty threshold)	26.7%	14.8%	30.4%	12.7%	13.5%	22.3%	20.5%	24.8%
TOTAL	100%	100%	100%	100%	100%	100%	100%	

wgtd 2008-10 ACS,SSDAN/U-Mich

True-False: Answer the following for the elderly population in Table 1:

1. ___ 26.7% of those with “comfortable incomes” are Non-Hispanic Whites.
2. ___ 19% of Hispanic elderly have incomes that are below the official poverty line.
3. ___ 11% of the elderly qualify as “near poor” (between 1 and 2 times the poverty threshold).
4. ___ 39% of Hispanic and Native American elderly live in poverty.
5. ___ 64.7% of African American elderly have incomes 200% or more above the poverty threshold.
6. ___ In 2010, about a quarter of all elderly in the U.S. lived at four times or above the official poverty line.
7. ___ 12.8% of the elderly in poverty are Asian.
8. ___ Table 1 compares race/ethnicity and poverty levels.
9. ___ Elderly whites in poverty are 7.3%. [correct this sentence]

Table 2: Difficulties in Bathing or Dressing among the Elderly, by Race-Ethnicity, 2010

	NhWhite	Black	Asian	Hispanic	AmIndian	NHOther	NHMulti	TOTAL
Yes	7.8%	12.1%	8.4%	11.8%	12.6%	8.6%	11.2%	8.5%
No	92.2%	87.9%	91.6%	88.2%	87.4%	91.4%	88.8%	91.5%
TOTAL	100%	100%	100%	100%	100%	100%	100%	

wgtd 2008-10 ACS,SSDAN/U-Mich

Please write 4 T/F statements to describe the table above:

Now let's examine the influence of poverty on independent living.

Table 3: 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	100%	100%	100%	100%	

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

When interpreting this table, you are want to compare ACROSS the categories (e.g. the poor to the near poor), not down the categories. Below, write a brief description of differences in independent living ability by poverty level. (You might want to rename the categories for income/poverty status, if that makes it easier to discuss. If you do so, then lay that out clearly.) Remember to start with a general statement (with no numbers). Then use the statistics to support your opening claim.