

Minnesota Population Center

Training and Development

IPUMS – USA Online Data Analysis

Exercise 1

OBJECTIVE: Gain an understanding of how the IPUMS dataset is structured and how it can be leveraged to explore your research interests. This exercise will use the IPUMS dataset to explore a variety of household characteristics.

Research Questions

What are the patterns of household characteristics in the U.S.?

Objectives

- Select datasets and variables of interest
- Analyze the data using sample code
- Validate data analysis work using answer key

IPUMS-USA Variables

- BLIND: Blind
- DIFFEYE: Seeing difficulty
- KITCHEN: Whether or not the household has a kitchen
- MARST: Marriage status

SDA Code to Review

| Field | Purpose |
|------------------|-----------------------------------------------------------------------|
| Row | Represents the primary variable of interest |
| Column | Divides the analysis of the variable of interest into categories |
| Control | Creates a separate chart for each category of the control |
| Selection Filter | Allows you to select cases; ex: year(2000-*) -> all years 2000-onward |

Review Answer Key (page 4)

Common Mistakes to Avoid

1 Choosing numerical instead of categorical variables for the Frequencies/Cross Tabulation Program. For these, use the Comparison of Means Program instead.

2 Forgetting to specify the years of interest

Getting Started

Step 1

- Go to <http://usa.ipums.org/usa/sda/>, and select the 1870 1% sample
- The default analysis is frequency/tabulation

...

Step 2

Research Variables of Interest

- Either browse variables under the Household and Person variables categories, or Search on the main IPUMS-USA site for variables
- When you browse for a variable, click on it, and it will appear in the Selected box. To send the variable to your input section, click the appropriate box (Row, etc)
- Row and Column are the variables of interest that you will perform the cross-tabulation on
- Filters select only specified cases
- A Control creates multiple tables for row and column variables, separated by a third categorical variable. For example, if you include the variable SEX as a control, you will get two frequency tables
- The Weight default is person weight (perwt), which extrapolates the sample to represent the entire population

Part I – Basic Frequencies

Section 1

Analyze the Data

A) What is the number and percent of blind people in 1870?

Row: blind
Weight: perwt

B) What is the number and percent of people with a vision difficulty in 2010? Return to the previous page and select the 2010 sample. _____

Row: diffeye
Weight: perwt

C) What is the universe for the variable KITCHEN? Find the variable description on the main website. _____

D) What percent of people had a kitchen in their home in 1980?
_____ In 2010? _____

Row: kitchen
Filter: gq(1-2)
Weight: perwt

E) What percent of households had a kitchen in their home in 1980?
_____ In 2010? _____

Row: kitchen
Filter: gq(1-2), pernum(1)
Weight: hhwt

F) What proportion of women and men ages 65+ are widows/widowers in 2010? _____

What proportion of the population ages 65+ are women and men in 2010? _____

Row: marst
Column: sex
Selection filter: age(65-*)
Weight: perwt
In TABLE OPTIONS, choose "Column Percentaging"

ANSWERS: Part I – Basic Frequencies

Section 1

Analyze the Data

A) What is the number and percent of blind people in 1870? 20,969 people, 0.1% of people were blind in 1870.

Row: blind
Weight: perwt

B) What is the number and percent of people with a vision difficulty in 2010? Return to the previous page and select the 2010 sample. 6,855,575 people, 2.2% of the population had difficulty seeing in 2010.

Row: diffeye
Weight: perwt

C) What is the universe for the variable KITCHEN? Find the variable description on the main website. Not group quarters

D) What percent of people had a kitchen in their home:

In 1980? 98.3% In 2010? 99.2%

Row: kitchen
Filter: gq(1-2)
Weight: perwt

E) What percent of households had a kitchen in their home:

In 1980? 98.1% In 2010? 99.0%

Row: kitchen
Filter: gq(1-2), pernum(1)
Weight: hhwt

F) What proportion of women and men ages 65+ are widows/widowers in 2010? 40.5% of women and 13.2% of men. What proportion of the population ages 65+ are women and men in 2010? 56.9% are women and 43.1% are men.

Row: marst
Column: sex
Selection filter: age(65-*)
Weight: perwt
In TABLE OPTIONS, choose "Column Percentaging"