



## IPUMS Data Training Exercise: CPS Online Data Analysis (SDA Exercise 1)



### Learning goals

- Gain an understanding of how the IPUMS dataset is structured and how it can be leveraged to explore your research interests.
- Select samples and variables of interest
- Analyze the data using online SDA coding
- Validate data analysis work using answer key

### Summary

This exercise will use the IPUMS dataset to explore general statistics about public housing recipients, and also changes in age of marriage of women over time. Research questions include: “what is the general income status and frequency of public housing recipients in the United States?” “what are the differences in income between non-citizens and citizens?” “has women's average age of marriage changed in the past 50 years, and why?”

## IPUMS Variables

- PUBHOUS: Public housing status
- HHINCOME: Household income
- INCTOT: Total personal income
- EMPSTAT: Employment status
- CITIZEN: Citizen status
- EDUC: Educational attainment
- AGEMARR: Age when first married

## SDA Field Descriptions

- 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

## 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

### a. Select Samples

- Go to <https://cps.ipums.org/cps/sda.shtml>, and select "Analyze all ASEC samples"
- Note that to analyze specific years, you will need to create a selection filter. For example: year(2005)
- The default analysis is frequency/crosstabulation

### b. Choose Variables of Interest

- Either browse variables under the Household and Person variables categories, or Search on the main IPUMS-CPS 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)
- The Weight default is person weight (sdawt), which extrapolates the sample to represent the entire population.



## Analyze the Sample

### c. Part I: Basic Frequencies

1. What is the description and universe for PUBHOUS?

---

---

---

2. On the website, find the codes page for the PUBHOUS variable and write down the code value, and what category each code represents.

---

---

---

3. What is the number and percent of people who lived in public housing in 1980?

---

Row: pubhous

Weight: sdawt

Selection Filters: year(1980)

4. What is the number and percent of people who lived in public housing in 2011?

---

Row: pubhous

Weight: sdawt

Selection Filters: year(2011)

5. What percent of households lived in public housing in 2011?

---

Row: pubhous

Weight: hsdawt

Selection Filters: year(2011), pernum(1)



d. Part II: Relationships in the Data

In the upper-left corner, hover over "Analysis" and select "Comparison of Means".

6. What are the average household incomes for each category of PUBHOUS in 2011?

---

---

---

Dependent: hhincome

Row: pubhous

Filter: year(2011), pernum(1)

Weight: hsdawt

7. Controlling for education, do citizens earn more personal income on average than non-citizens in the United States in 2011?

---

---

---

Dependent: inctot

Row: citizen

Column: educ

Control: educ

Filter: year(2011), citizen(2-3), educ(2-\*)

Weight: sdawt

e. Part III: Relationships in the Data

Continue analysis under "Comparison of Means".

8. What is the code for NIU (Not in Universe) for AGEMARR?

---



9. What was the average age of marriage for women who are at work and for women who are unemployed, new worker in 1965?

a. Does this make sense? \_\_\_\_\_

```
Dependent: agemarr
```

```
Row: empstat
```

```
Selection filter: year(1965), sex(2), agemarr(1-*)
```

```
Weight: sdawt
```



# ANSWERS

## f. Part I: Basic Frequencies

1. What is the description and universe for PUBHOUS? Description: "PUBHOUS indicates whether the house, apartment, or mobile home is part of a government housing project for people with low incomes, commonly known as a "public housing project"." Universe: Interviewed households and group quarters whose members are not buying and do not own their house.
2. On the website, find the codes page for the PUBHOUS variable and write down the code value, and what category each code represents. 0 = NIU, 1 = No, 2 = Yes.
3. What is the number and percent of people who lived in public housing in 1980? 4,687,197 people, 2.1% of the population lived in public housing in 1980
4. What is the number and percent of people who lived in public housing in 2011? 7,716,254 people, 2.5% of the population lived in public housing in 2011
5. What percent of households lived in public housing in 2011? 3.0% of households lived in public housing in 2011

## g. Part II: Relationships in the Data

In the upper-left corner, hover over "Analysis" and select "Comparison of Means".

6. What are the average household incomes for each category of PUBHOUS in 2011? NIU: \$80,221; No: \$45,208; Yes: \$15,801
7. Controlling for education, do citizens earn more personal income on average than non-citizens in the United States in 2011? Yes, for every level of education, citizens earn more than non-citizens

## h. Part III: Relationships in the Data

Continue analysis under "Comparison of Means".

8. What is the code for NIU (Not in Universe) for AGEMARR? 00. Because this variable is numerical, the NIU values must be excluded otherwise the means will be biased towards zero.



9. What was the average age of marriage for women who are at work and for women who are unemployed, new worker in 1965? At work: married at 21.86 yrs;  
Unemployed, new worker: married at 18.26 years
- a. Does this make sense? This makes sense because women may delay marriage if they enter the work force and are self-sufficient

