

Minnesota Population Center

Training and Development

# IPUMS – Int.l 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 cell phone ownership, fertility, and migration patterns.

## Research Questions

What are the patterns of cell phone ownership in South Africa? What is the trend in fertility in China? What is the relationship between income and migration in India?

## Objectives

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

## IPUMS-I Variables

- CELL: Cellular phone availability
- URBAN: Urban/rural status
- CHBORN: Children ever born to the individual
- CHBORNF/CHBORNM: Children ever born to the individual, female/male
- MGCAUSE: Cause of migration
- INCWAGE: Wage and salary income per week

## 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 6)

## 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

#### Select a Sample

- Go to <https://international.ipums.org/international/>, and on the left-hand side of the page, select "Analyze Data Online"
- You will have the option to analyze single year or multi-year samples of one country, or datasets of all countries within a region for all available years
- Choose the multi-year sample for South Africa
- The default analysis is frequencies and cross-tabulation

...

### Step 2

#### Research Variables of Interest

- Search on the main IPUMS-I site for variable descriptions, codes, universe, etc, or browse under Household and Person variable categories
- **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 (wtper), which extrapolates the sample to represent the entire population

## Part I - South Africa Cell Phone Ownership

### Section 1

#### Analyze the Data

A) What are the codes and universe for CELL? Is this a person variable or a household variable? \_\_\_\_\_

B) What are the respective shares of households for which a cell phone was available?

i. Rural households across all samples? \_\_\_\_\_

ii. Urban households? \_\_\_\_\_

Row: cell

Column: urban

Weight: wthh

Selection Filters: cell(1-2), urban(1-2)

C) Compare the share of rural or urban households where a cell phone was available in 1996 to 2007. Comment on the change over time.

---

Row: cell

Column: urban

Weight: wthh

Control: year

Selection Filters: cell(1-2), urban(1-2)

## Part II – Fertility in China

Choose the multi-year sample for China. In the upper-left corner, hover over “Analysis” and then select “Comparison of Means”.

### Section 1

### Analyze the Data

A) What is the universe for CHBORN? Why would a frequency table by CHBORN and SEX not be informative?  
\_\_\_\_\_

B) What is the code in CHBORN for NIU (Not in Universe)? Should they be excluded from a comparison of means? \_\_\_\_\_

C) Compare the means of CHBORN between 1982 and 1990.  
\_\_\_\_\_

D) What was the ratio of male children ever born to female

```
Dependent: chborn  
Row: year  
Selection filter: sex(2), chborn(*-30)  
Weight: wtpcr
```

children ever born in 1990? \_\_\_\_\_

*Either return to the main SDA page and choose the China 1990 sample, or create a year (1990) filter.*

```
Dependent: chbornf, chbornm  
Row: year  
Filter: chbornf(*-30), chbornm(*-30), sex(2)  
Weight: wtpcr
```

## Part III – Migration and Income in India

Choose the sample for India 1983. Use the Frequencies and Crosstabulations first.

### Section 1

Analyze  
the Data

A) What are the reasons for migration in India in 1983?

Most common reason? \_\_\_\_\_

Least common reason? \_\_\_\_\_

Row: mgcause

Selection filter: mgcause(10-69)

Weight: wtper

B) Now change to Comparison of Means. What was the average income for migrants who moved due to:

i. Natural disaster? \_\_\_\_\_

ii. Social and political problems? \_\_\_\_\_

iii. Job relocation? \_\_\_\_\_

Why might this be? \_\_\_\_\_

Dependent: incwage

Row: mgcause

Selection filter: mgcause(10-69), incwage(\*-9999997)

Weight: wtper

...

Complete!

Validate  
Your  
Answers

## ANSWERS: Part I - South Africa Cell Phone Ownership

### Section 1

#### Analyze the Data

- A) What are the codes and universe for CELL? Is this a person variable or a household variable? Codes: 0 NIU (not in universe); 1 Yes; 2 No; 9 Unknown; Universe: South Africa 1996: Private households; South Africa 2001: Non-homeless households; South Africa 2007: Private household; CELL is a household variable
- B) What are the respective shares of households for which a cell phone was available?
- Rural households across all samples? Yes: 33%, No: 67%
  - Urban households? Yes: 56.7% No: 43.3%

Row: cell  
Column: urban  
Weight: wthh  
Selection Filters: cell(1-2), urban(1-2)

- C) Compare the share of rural or urban households where a cell phone was available in 1996 to 2007. Comment on the change over time. 1996: Yes: 3.9% No: 96.1%; 2007: Yes: 71.6% No: 28.4% The availability of cell phones to rural households in South Africa is converging on the availability of cell phones in urban households.

Row: cell  
Column: urban  
Weight: wthh  
Control: year  
Selection Filters: cell(1-2), urban(1-2)

## ANSWERS: Part II – Fertility in China

Choose the multi-year sample for China. In the upper-left corner, hover over “Analysis” and then select “Comparison of Means”.

### Section 1

### Analyze the Data

A) What is the universe for CHBORN? Why would a frequency table by CHBORN and SEX not be informative? Females ages 15 to 64. There are no men in the universe of this question.

B) What is the code in CHBORN for NIU (Not in Universe)? Should they be excluded from a comparison of means?

The code for NIU is 99, and 98 is Unknown. If these aren't excluded, they will be included in the means of children ever born and bias the estimates upwards.

C) Compare the means of CHBORN between 1982 and 1990.

1982: Average of 2.61 children ever born; 1990: Average of 2.1

```
Dependent: chborn  
Row: year  
Selection filter: sex(2), chborn(*-30)  
Weight: wtper
```

D) What was the ratio of male children ever born to female children ever born in 1990? The mean number of male children ever born was 1.09, and mean number of female children born was 1.01, for a ratio of  $1.09/1.01 = 1.079$  boys to each girl.

*Either return to the main SDA page and choose the China 1990 sample, or create a year (1990) filter.*

```
Dependent: chbornf, chbornm  
Row: year  
Filter: chbornf(*-30), chbornm(*-30), sex(2)  
Weight: wtper
```

## ANSWERS: Part III – Migration and Income in India

Choose the sample for India 1983. Use the Frequencies and Crosstabulations first.

### Section 1

#### Analyze the Data

A) What are the reasons for migration in India in 1983?

Most common reason? Marriage or union

Least common reason? Natural disaster

Row: mgcause

Selection filter: mgcause(10-69)

Weight: wtper

B) Now change to Comparison of Means. What was the average income for migrants who moved due to:

i. Natural disaster? 14.86 rupees

ii. Social and political problems? 23.38

iii. Job relocation? 166.78

Why might this be? The income for migrants who moved due to natural disasters are likely to earn much less than average in the past week because they may have lost everything or did not have enough capital to rebuild and stay in the same place. Families that are slightly wealthier than average might be able to afford to move due to security. Job relocators are paid the most because the pay must be higher to exceed the costs of moving.

Dependent: incwage

Row: mgcause

Selection filter: mgcause(10-69), incwage(\*-9999997)

Weight: wtper