



## IPUMS Data Training Exercise:

### IPUMS International (No stats software needed) SDA Online Data Analysis



#### Learning Goals

- Understand how IPUMS International dataset is structured
- Select datasets and variables of interest from IPUMS International database in SDA (the online analysis tool)
- Analyze the cell phone ownership in South Africa, fertility in China, and the relationship between income and migration in India using sample code
- Validate data analysis work using the answer key
- Understand how IPUMS data can be leveraged to explore research interests

## Exercise Research Questions and Variables

In this exercise, you will gain basic familiarity with the IPUMS International data exploration and online analysis system to answer the following 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?

You will use sample code and the variables CELL, URBAN, CHBORN, CHBORNF/CHBORNM, MGCAUSE, and INCWAGE to analyze the data.

## Register as an IPUMS International User

Go to <http://international.ipums.org>, click on User Registration and Login and Apply for access. On the login screen, enter email address and password and submit your application. Please note that IPUMS International user applications are reviewed by IPUMS staff, and a final decision may take 2-5 business days.

## Getting Started with SDA

### SDA code to review

Code	Purpose
Row	Represents the primary variable of interest
Column	Divides the analysis of the variable of interest in to 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

- Forgetting to exclude missing data categories.
- Forgetting to specify the years of interest.
- Failing to update the weight. The default value for weight is the person weight (PERWT), which is not appropriate for household level analysis.



- Choosing a numerical rather than categorical variable for the Frequencies/Cross Tabulation Program. For continuous variables, use the Comparison of Means Program instead.
- Not updating column or row percentages (column percentages are default).

## Select samples

- Navigate to the IPUMS International homepage and click on "Analyze Data Online" On the left-hand menu below the "Data" header.
- The next page summarizes information about SDA and allows you to select single-year samples or multi-year samples for countries.
- Note that you will need to log in to use SDA.

## Browsing and adding variables

- You may browse variables under the Household and Person variable categories on the left-hand menu of SDA, or search for them on the IPUMS International website and enter the variable names directly as inputs into SDA.
  - Note that some variables available through the IPUMS International extract system are not available for analysis in SDA.
- When you browse variables within SDA, click on the plus symbol to the left of a variable topical group to see the available variables.
  - Click on a variable name, and it will appear in the "Selected" box at the top of the left-hand SDA variable browsing area.
  - To insert the selected variable as an input for your SDA analysis, click on the appropriate input location (e.g., row, column).



# Analyze the Data

## South Africa

From the IPUMS International SDA page, select the multi-year South Africa sample. This analysis will focus on the following variables:

- CELL: Cellular phone availability
- URBAN; Urban/rural status
- SEX: Sex

For each variable, search through the tabbed sections of the variable description to answer each question.

1. What are the codes and universe for CELL? Is this a person variable or a household variable? \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_
2. What are the respective shares of households for which a cell phone was available?
  - a. Rural households across all samples? \_\_\_\_\_
  - b. Urban households? \_\_\_\_\_

```
ROW: cell
COLUMN: urban
WEIGHT: wthh
SELECTION FILTER(S): cell(1-2), urban(1-2)
```

3. 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
CONTROL: year
```



```
SELECTION FILTER(S): cell(1-2), urban(1-2)
```

```
WEIGHT: wthh
```

## China

Navigate back to the IPUMS International SDA page, and select the multi-year China sample. This analysis will focus on the following variables:

- CHBORN: Children ever born to the individual
- CHOBORNF/CHBORNM: Children ever born to the individual, female/male
- SEX: sex

For each variable, search through the tabbed sections of the variable description to answer each question.

4. What is the universe for CHBORN? Why would a frequency table by CHBORN and SEX not be informative? \_\_\_\_\_
5. What is the code in CHBORN for NIU (not in Universe)? Should they be excluded from a comparison of means? \_\_\_\_\_

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

6. Compare the means of CHBORN between 1982 and 1990.  
\_\_\_\_\_
7. What was the ratio of male children ever born to female children ever born in 1990?  
\_\_\_\_\_

```
DEPENDENT: chborn
```

```
ROW: year
```

```
SELECTION FILTER(S): sex(2), chborn(*-30)
```

```
WEIGHT: perwt
```



DEPENDENT: chbornf, chbornm

ROW: year

SELECTION FILTER(S): chbornf(\*-30), chbornm(\*-30), sex(2)

WEIGHT: perwt

## India

Navigate back to the IPUMS International SDA page, and select the single-year India sample for 1983. This analysis will focus on the following variables:

- INCWAGE: Wage and salary income per week
- MGCAUSE: Cause of migration

For each variable, search through the tabbed sections of the variable description to answer each question.

8. What are the reasons for migration in India in 1983?

- a. Most common reason? \_\_\_\_\_
- b. Least common reason? \_\_\_\_\_

ROW: mgcause

SELECTION FILTER(S): mgcause(10-69)

WEIGHT: perwt

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

- a. Natural disaster? \_\_\_\_\_
- b. Social and political problems? \_\_\_\_\_
- c. Job relocation? \_\_\_\_\_

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

DEPENDENT: incwage

ROW: mgcause

SELECTION FILTER(S): mgcause(10-69), incwage(\*-9999997)

WEIGHT: perwt



# Answers

## South Africa

1. 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 household variable
2. What are the respective shares of households for which a cell phone was available?
  - a. Rural households across all samples: Yes: 33%; No: 67%
  - b. Urban households: Yes: 56.7%; No: 43.3%
3. 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.

## China

4. What is the universe for CHBORN? Why would a frequency by CHBORN and SEX not be informative?  
Females ages 15 to 64. There are no men in the universe of this question.
5. 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.
6. Compare the means of CHBORN between 1982 and 1990.  
1982: Average of 2.61 children ever born; 1990: Average of 2.1
7. 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.

## India

8. What are the reasons for migration in India in 1983?
  - a. Least common reason: Marriage or union
  - b. Least common reason: Natural disaster



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

- a. Natural disaster: 14.86 rupees
- b. Social and political problems: 23.38 rupees
- c. Job relocations: 166.78 rupees

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.

