Data Setup

38th Summer Seminar

Amonthep “Beet” Chawla

Objectives

Familiarize with the data set by

- Knowing important variables in the data set,
- Understanding the codebook and
- Realizing if the data set has any problems.

2

Methods

1. Transform Raw Data into the STATA Format
2. Extract Variables
3. Data Problems (more on Friday by SH LEE)

3

1. Transform Raw Data into the STATA Format

- Raw data may not be recorded in the STATA format. If the data set is recorded in other statistical program, you can use the program “STAT TRANSFERS” to change the format of the data set.
- However, if you don’t have that, you can create dictionary files to transform raw data into the STATA format.
- Please refer to your codebook to know what variables you have in order to write a dictionary file.

4

Dictionary

- Raw Data “1001010548….006504…”

The first 10 digits are household ID

household income

- Command in STATA
  “infile using “dictionary.txt””
- Dictionary file *(dictionary.txt)*
  - *dictionary using “rawdata.dat"{*
  - *double id %10f "id“*
  - *_column(71) mthcuri %6f "monthly current income"*

5

2. Extract Variables

- To create Consumption and Labor Income for the National Transfer Flow Accounts, we need to use several variables from the surveys.
- There are two groups of variables: households and individuals.
- Please write down variable names from your data set so that you can remember them.

6

Consumption

- Education*
- Health*
- Housing (service from owner-occupied housing)
- Durable
- Tobacco
- Alcohol
- Total household consumption*

7

Non-consumption

- Income tax
- Social security tax
- Other cash payment to the government
- Interest payment on debt
- Transfer expense to other households*

8

Income

- Earnings*
- Income from unincorporated enterprises*
- Property income (i.e., dividends, rent, interest)*
- Non-labor income
- Transfer income from other households*
- Social security benefits
- Other cash transfer received from the government

9

Characteristics

- Dummy for school attendance (currently attending school or not)*
- Education level for school attendance (if currently attending school, what is the level of education: primary, secondary or higher education)*
- Age of individuals*
- Dummy for household ship (is the person a head of the household?)*
- Health characteristics (number of times visiting hospitals or any information regarding to health situation)
- Weights (stratification of households)*

10

3. Data Problems (more on Friday by SH LEE)

- Change from two or more household heads to only one head
- Missing values (i.e. replace those who do not report income with 0)
- Detect households that have too high income/expenditures (i.e. using scattered plots)
- Summarize data

11

Data Problems (Cont.)

- Change "." to "0" for wage using *recode wage (.=0)*
- Using the following command to see scattered plots for wage
  - qnorm wage, grid
  - stem wage, prune

12

Next

- **We will discuss about “Aggregate Control” on June 4.**
- **Please bring National Income Accounts for your country.**

13