**Stata hands-on: Data Cleaning, Sierra Leone Integrated Household Survey 2011**

**Part 1. Document and Understand the Survey**

Introduction to the Sierra Leone Integrated Household Survey, 2011

**Questions**:

* How many households are in the survey?
* What is the level of representativeness of the survey?
* What is the structure of the data sets?

**Expenditures**

Before constructing the consumption aggregate, we need to get to know the questionnaire and variables. Take stock of all the variables. Which sections in the questionnaire ask about consumption? Which variables capture food and non-food expenditure and consumption?

Table 1. Notes on consumption related variables in the SLIHS2011

|  | Section in Questionnaire, page number | Units and Time reference | Notes | Stata file and variables |
| --- | --- | --- | --- | --- |
| Food |  |  |  |  |
|  |  |  |  |
| Non-Food |  |  |  |  |
| Education |  |  |  |  |
| Health |  |  |  |  |
| Housing |  |  |  |  |

**Questions**:

* What is the data collection period of the survey?
* How many food categories are there?
* How many non-food categories are there?
* What does 2nd – 3rd – 4th , 5th – 6th mean?
* What are the monetary units and time reference?

**Variables for adjustments**

**Prices** are need for spatial deflations. Ideally, costs are adjusted to reflect differences in regional cost of living.

|  |  |  |
| --- | --- | --- |
|  | Notes | Stata file and variables |
| Price |  |  |
| Quantity |  |  |

**Questions**:

* When was the survey conducted, which months?
* Are there other sources of price data if there is no price and quantity data in the questionnaire?

Consumption is measured in per capita or adult equivalency units, variables on **household composition** is needed.

|  |  |  |
| --- | --- | --- |
|  | Notes | Stata file and variables |
| Household roster |  |  |
| Age |  |  |
| Gender |  |  |
| Relation to the head of household |  |  |

**Questions**:

* Does household size include non-family members?