

**PEP-LAVAL INTENSIVE GRADUATE SCHOOL IN DEVELOPMENT  
ECONOMICS**

**Understanding the Macro-Micro Linkages of Policy  
and Development  
(10-14 June 2013)**

***Learning Stata: Exercises***

**by**

**Abdelkrim Araar**

**Jean-Yves Duclos**

***Université Laval, June 2013***

### Exercise 1: *Exploring the data, producing descriptive statistics and using weights*

- 1.1 Show the number of observations in the datafile *Nigeria\_04l.dta*.
- 1.2 Show the number of surveyed households by ***strata***.
- 1.3 Estimate the population number of households by ***strata***.
- 1.4 Estimate the population number of individuals by ***strata***.
- 1.5 Estimate the proportion of the total population by ***strata*** and by ***zone***. For this, use the command `svy: tab`.
- 1.6 Estimate the mean of ***pcexp*** using the command `summary`. Comment on your results.
- 1.7 Estimate the mean of ***pcexp*** using the command `summary` and the appropriate weights.
- 1.8 Further estimate the mean of ***pcexp*** using the command `mean` and the appropriate weights.
- 1.9 Redo (8.) with the command `svy: ratio`.
- 1.10 Show the descriptive statistics of ***pcexp*** by sex of the household head.
- 1.11 Show only the averages of ***pcexp*** by ***zone***.
- 1.12 Show averages by urban and rural areas (***sector***).

### Exercise 2: *The command append*

- 2.1 Using the data *Uganda\_99l.dta* and the data *Uganda\_92l.dta*, create a new data file that contains the two data files with the variables ***hh psu strata urban region hsize sweight welfare***. Also, add the variable `year` to indicate the year of survey for each observation. Write a program that can do this.

2.2 Show the frequency of observations by year.

### Exercise 3: *The command merge*

You need to merge the files Uganda\_f1.dta and Uganda\_f2.dta. However, you are not sure that they contain the same observations or that the observations are classified in the same order. What is the surest procedure to merge these two files? Merge the two datafiles, and then save the new file as Uganda\_m99.dta.

### Exercise 4: *The command expand*

- 4.1 Use the household datafile hhexp\_94.dta to generate an individual datafile in which each household observation is replicated according to its household size.
- 4.2 Sort the data in ascending order by the variable **hhid**;
- 4.3 Generate the variable **id\_indv** that equals the number of individuals within the household.
- 4.4 Generate the variable **hh\_inc** that indicates the total income of household, assumed to be equal to its total expenditures.
- 4.5 Generate a variable **indv\_inc** that is equal to individual income. For this illustrative example, assume that:
  - a. Individual income equals total household expenditures if the household size is 1.
  - b. The individual income of the first member (**indv\_id** = 1, i.e. the household head) equals 60% of total expenditures.
  - c. The individual income of the second member (**indv\_id** = 2) equals 40% of total expenditures.
  - d. The individual income of the other members is nil.
- 4.6 Check if total household expenditures equal total individual incomes.
- 4.7 Keep the variables **hhid**, **indv\_id** and **indv\_pc**, and then save the file with the name indv\_inc\_94.dta.

### **Nigeria\_2004I.dta**

The survey data used was collected by Nigeria's National Bureau of Statistics (NBS) formerly known as the Federal Office of Statistics (FOS). They were based on National Living Standard Survey (NLSS) of households that was carried out between September 2003 and August 2004.

The sample design is a two-stage stratified sampling. At the first stage, clusters of 120 housing units called Enumeration Area (EA) were randomly selected from each State and the Federal Capital Territory (FCT, Abuja). The second stage involved random selection of 5 housing units from the selected EAs. A total of 600 households were randomly chosen in each of the States and the FCT, summing up to 22,200 households in all (FOS, 2003). However, some households did not fully complete the questionnaires. Out of the 22,200 households that were targeted, only 19,158 completed the survey.

It is to be noted that there is no official absolute poverty line in Nigeria. Usually, the relative approach is used to estimate the poverty line in Nigerian studies (poverty line equals to two third of average standard of living). In this study, we additionally use the World Bank poverty line that is US\$1 per day by adult equivalent.

|                       |  |
|-----------------------|--|
| <b><i>case_id</i></b> | Household identifier   |
| <b><i>state</i></b>   | Stratum  |
| <b><i>psu</i></b>     | Primary sampling units                                       |
| <b><i>sector</i></b>  | Rural/Urban areas  |
| <b><i>pid</i></b>     | Person id  |
| <b><i>hhsize</i></b>  | household size   |
| <b><i>zone</i></b>    | Zone   |
| <b><i>pcexpdr</i></b> | Per capita expenditure in regionally deflated current prices |
| <b><i>sweight</i></b> | Sampling weights   |
| <b><i>sex</i></b>     | Sex  |

***age***            Age years

***occupation***   Occupation group

***ed\_level***      Educational groups for highest level attained

***agr\_occ***        Agricult Occupation?

**Uganda\_1999I.dta**

The Uganda National Household Surveys (UNHS) of 1999 is a nationally representative survey, with sample selection using two-stage stratified random sampling..Uganda\_1999I.dta contains the following variables:

***hh***                Household identifier

***hsize***           household size

***district***        District

***psu***             Primary sampling unit(enumeration area)

***sweight***        Sampling weight

***urban***           Urban dummy

                      0 rural  
                      1 urban

***region***          region

                      1 central  
                      2 eastern  
                      3 northern  
                      4 western

***Strata***          Stratum: region/place of residence

                      10 central rural  
                      11 central urban  
                      20 eastern rural  
                      21 eastern urban  
                      30 northern rural  
                      31 northern urban  
                      40 western rural  
                      41 western urban

***equiv***            household size adjusted for adult equiv. scale

***nwelfare***        consumption aggregate per adult equivalent in 1997/98 prices

***spline*** absolute poverty line in 1997/98 prices

***poor99*** poor dummy

0 non-poor  
1 poor

***welfare*** Monetary welfare indicator

***sex*** Sex of household head

1 Male  
2 Female

***age*** Age of household head

***marital*** Marital status of household head

0 Undefined  
1 Unmarried  
2 Married  
3 Cohabiting  
4 Divorced/separated  
5 Widowed

***activity*** Main activity-status of household head

1 Too young or old  
2 Disabled  
3 Student  
4 Employer  
5 Own account worker  
6 Unpaid family worker  
7 Gov't employee  
8 Private employee  
9 Employed  
10 Political, social, religious worker  
11 Att. Domestic duties  
19 Others

**Uganda\_1992I.dta**

The Uganda National Household Surveys (UNHS) of 1999 is a nationally representative survey, with sample selection using two-stage stratified random sampling..Uganda\_1999I.dta contains the following variables:

|                        |  |
|------------------------|--|
| <b><i>hh</i></b>       | Household identifier   |
| <b><i>hsize</i></b>    | household size   |
| <b><i>district</i></b> | District   |
| <b><i>exdis</i></b>    | Districts not covered in 1999/00 dummy   |
| <b><i>psu</i></b>      | Primary sampling unit(enumeration area)  |
| <b><i>sweight</i></b>  | Sampling weight  |
| <b><i>urban</i></b>    | Urban dummy  |
|                        | 0 rural<br>1 urban   |
| <b><i>region</i></b>   | region   |
|                        | 1 central<br>2 eastern<br>3 northern<br>4 western  |
| <b><i>Strata</i></b>   | Stratum: region/place of residence   |
|                        | 10 central rural<br>11 central urban<br>20 eastern rural<br>21 eastern urban<br>30 northern rural<br>31 northern urban<br>40 western rural<br>41 western urban |
| <b><i>equiv</i></b>    | household size adjusted for adult equiv. scale   |
| <b><i>nwelfare</i></b> | consumption aggregate per adult equivalent in 1997/98 prices   |
| <b><i>spline</i></b>   | absolute poverty line in 1997/98 prices  |
| <b><i>poor92</i></b>   | poor dummy   |
|                        | 0 non-poor<br>1 poor   |
| <b><i>welfare</i></b>  | Monetary welfare indicator   |