CEQ INSTITUTE
COMMITMENT TO EQUITY
Tulane University

Introduction to the Timor Leste Living Standards Survey (TLSLS)

Training Workshop on the
Commitment to Equity Methodology

CEQ Institute, Asian Development Bank,
and The Ministry of Finance

Dili

June, 2017

About Household Surveys

- How GDE carries them out
  - draw a random sample of households
    - stratified by populations of interest
    - clustered, for economy
  - ask them a lot of questions
    - see the questionnaire
- *Why* GDE carries them out
  - useful to establish consumption patterns to establish weights for the consumer price index
  - useful to estimate (or describe) the distribution of income and/or consumption in Timor Leste
    - poverty
    - inequality
  - keeps researchers busy

About TLSLS

- How the data are structured
  - Many files (look at subdirectories)
  - More or less follow the structure of the questionnaire
  - Some are “household-level” data
    - one record per household
  - Some are individual-level, and some are community-level
- Skip patterns and missing values
  - look at section 4, part A, for example
  - Stata does not behave well with missing values, so be careful!

About Sampling and Sampling Weights

- Not every observation in the GLSS is equally valuable
  - urban areas are "over-sampled" because it is easy to interview there
  - there could be other over-sampled groups
    - for example, a small ethnic minority
    - Oecussi
- Because it is more likely that an urban household is chosen for the sample, it gets *less* weight than a rural one
- Intuition: the rural households selected must "represent" more households than the urban households selected
- A simple example

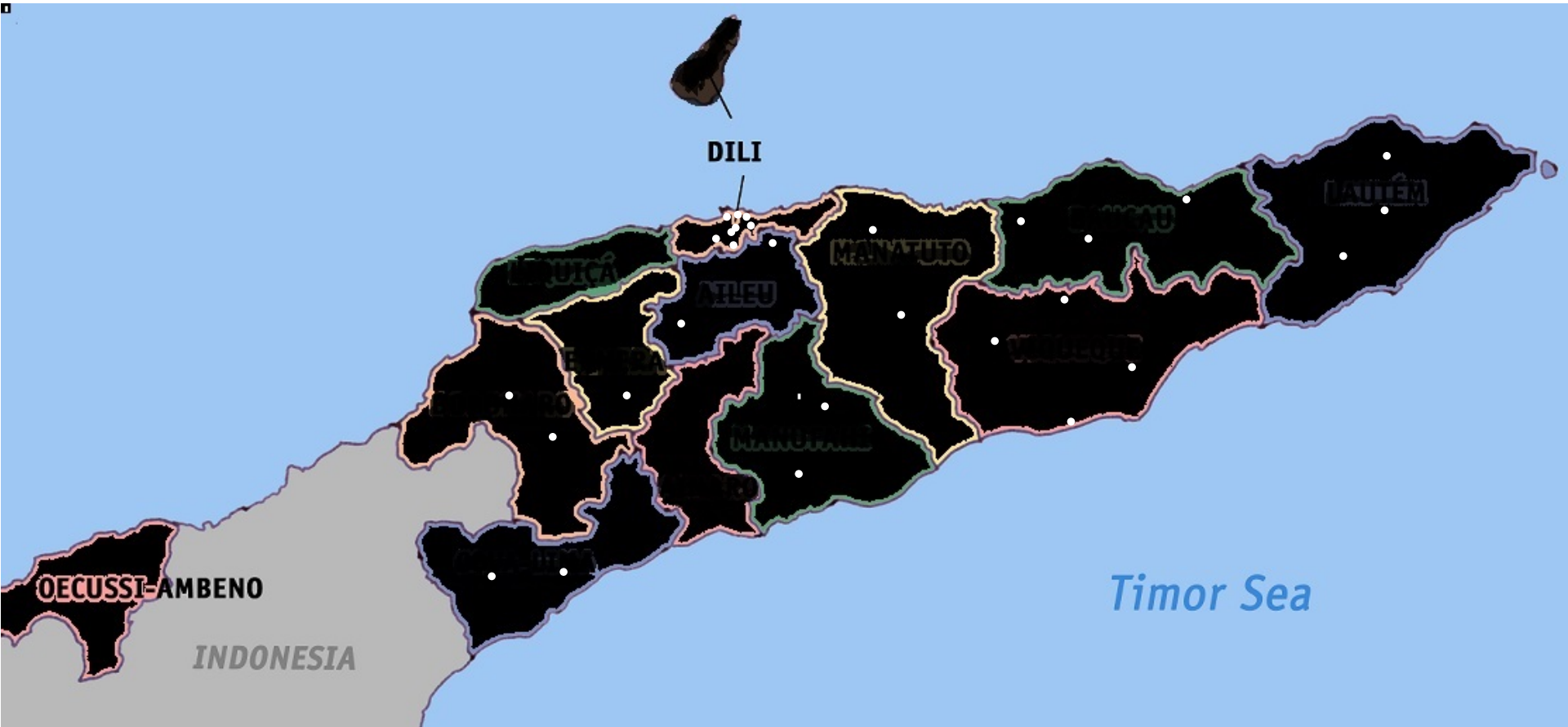
Simple Example of Sampling Weights

Person	Area	Income	Sampled?	Sample Income	Probability of Selection	Sampling Weight
1	Urban	100	Yes	100	0.5	2
2	Urban	300	No			
3	Urban	200	No			
4	Urban	400	Yes	400	0.5	2
5	Rural	50	No			
6	Rural	100	No			
7	Rural	20	No			
8	Rural	30	No			
9	Rural	40	Yes	40	0.33	3
10	Rural	60	Yes	60	0.33	3
		True Average of Income		Unweighted Estimate, Average Income	Weighted Estimate	
	Average:	130		150	130	

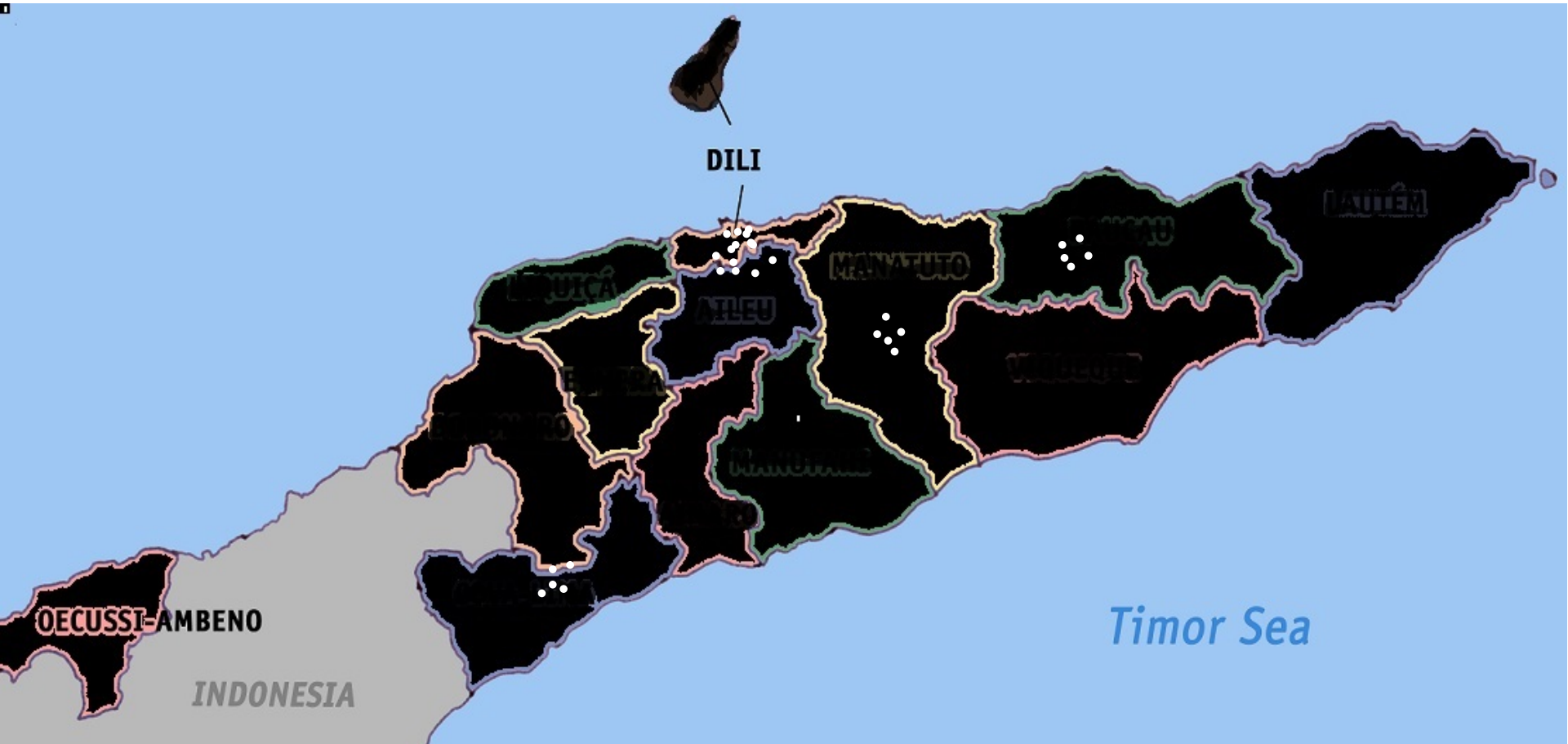
About Clustering and Stratification

- Random sampling of households vs. random sampling of clusters of households
  - see following graphics

Randomly drawn households



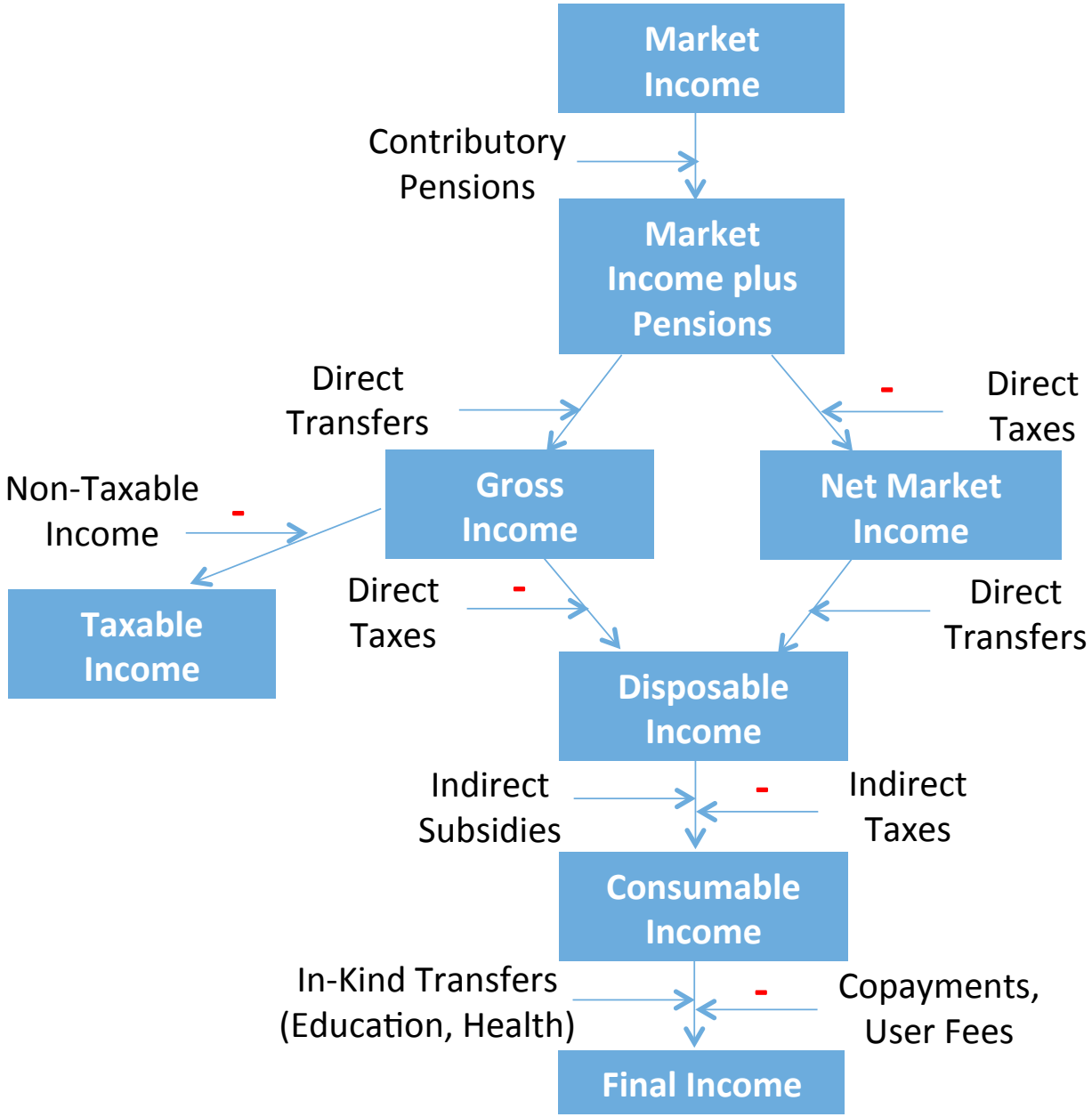
Randomly drawn *clusters* of households (census enumeration areas)



About Clustering and Stratification

- Clustering tends to increase the standard errors of things we estimate with survey data
  - households near each other are more similar to each other than they are to other households
  - reduces the variation in the sample
- Stratification can reduce the standard errors, but usually does not in TLSLS-type surveys
- Stata has very easy-to-use commands to take these things into account

What Does TLSLS Have for CEQ?



How Do We Go About the Analysis?

- We must dig up all the information we can about these income concepts and their components from TLSLS
- We must clean it up and save it in one dataset
- We can then apply pre-programmed commands to do much of the statistical work
- We must cross-check our results against administrative data whenever possible