

Training Course

An Introduction to Calibration Techniques in Sample Surveys

Jointly Organized by the World Bank and the Asian Development Bank

May 23-26 2016, ADB Headquarters, 6 ADB Avenue,
Mandaluyong City, Metro Manila, Philippines

Course Programme

DAY 1

9:00 – 12:30

Unfreezing

- Introduction of facilitators
- Introduction of participants
- Objectives of the training and outline of the course

What is Calibration? Why Should National Statistical Offices Care about It?

- Overview
- Toy Monte Carlo simulation: how Calibration can improve efficiency
- Toy Monte Carlo simulation: how Calibration can reduce nonresponse bias

An Introduction to R

- The R Project website and the CRAN
- How to download and install R under Windows
- How to find and use add-on packages of interest
- A quick reading guide to the R help system
- Overview of R basics
 - ✓ Data structures (vectors, matrices, arrays, data frames, factors, lists)
 - ✓ Objects manipulation (assignment, indexing, subsetting, replacing subsets, removing objects)
 - ✓ Data import/export
- Hints to: manipulating data frames (sort, append rows, add columns, merge, ...), vectorization, recycling rule, basic functions and the 'apply' family (sapply, tapply, lapply, mapply)

(11:15 – 11:30 coffee break)

Lunch Break, 12:30 – 13:30

13:30 – 17:00

Exercise Session

DAY 2

9:00 – 12:30

Unfreezing

- Questions and comments on topics covered by previous lessons: clarification

Introduction to the ReGenesees system

- Main statistical functions
- Software architecture
- Software design principles and user interaction
- An Object-Oriented overview of ReGenesees
- How to find, download and install ReGenesees
- Working with ReGenesees
 - ✓ Command-line interface and GUI
 - ✓ System resources: documentation, practical examples, sample data

Survey data and sampling designs

- How to persistently bind survey data to sampling design metadata: *survey design* objects
- How to handle complex sampling designs
 - ✓ Multistage, stratified, clustered, sampling designs
 - ✓ Sampling with equal or unequal probabilities, with or without replacement
 - ✓ Self-representing and non-self-representing strata
- How to manipulate survey design objects:
 - ✓ Accessing sample data and weights
 - ✓ Adding new variables and merging new data
 - ✓ Collapsing strata to handle 'lonely' PSUs

(11:15 – 11:30 coffee break)

Lunch Break, 12:30 – 13:30

13:30 – 17:00

Exercise Session

Calibration (part I)

- Specifying calibration models symbolically
 - ✓ R-model formulae: syntax and semantics in the calibration context
 - ✓ Factorizable calibration models: global and partitioned calibration

DAY 3

9:00 – 12:30

Unfreezing

- Questions and comments on topics covered by previous lessons: clarification

Calibration (part II)

- Managing auxiliary information: population totals *templates*
 - ✓ Building the template
 - ✓ Describing the template
 - ✓ Checking the template
 - ✓ Filling the template with actual population aggregates
- Setting up and executing a calibration task
 - ✓ Distance functions: linear, raking and logit
 - ✓ Range restrictions on the g-weights
 - ✓ Unit-level and cluster-level weights adjustment
 - ✓ Homoskedastic and heteroskedastic calibration models
- Diagnostics on the calibration task
- Hints to advanced techniques
 - ✓ Simultaneous calibration to auxiliary information on *individuals* and *households*

Computing estimates and sampling errors (part I)

- Calibration and Horvitz-Thompson estimators
 - ✓ Totals and Means
 - ✓ Absolute and relative frequency distributions (marginal, conditional and joint)

(11:15 – 11:30 coffee break)

Lunch Break, 12:30 – 13:30

13:30 – 17:00

Exercise Session

DAY 4

9:00 – 12:30

Unfreezing

- Questions and comments on topics covered by previous lessons: clarification

Computing estimates and sampling errors (part II)

- Calibration and Horvitz-Thompson estimators
 - ✓ Ratios between totals
 - ✓ Shares and ratios between shares
 - ✓ Multiple regression coefficients
 - ✓ Quantiles

(11:15 – 11:30 coffee break)

Lunch Break, 12:30 – 13:30

13:30 – 17:00

Exercise Session

Case study

- A calibration task on real-world survey data

Conclusions

- Feedback from trainees and discussion