

MSP-E5.1**45 Hours**

ADVANCED STATISTICAL PACKAGES AND APPLICATION IN LARGE SCALE DATA

Learning objectives:

1. To comprehend the need for big data in monitoring and evaluation of health and population policies
2. To introduce national and international big data and guide students in managing these data
3. To familiarize students to survey softwares

Unit I: Scope of large scale surveys and big data

Concept of big data, need for big data for planning and monitoring of public health programmes, introduction to large scale demographic and health surveys (DHS): NFHS, DLHS, WHO-SAGE, LASI-objectives, designs, instruments, sample size. Cleaning of big data- range and consistency checks, missing data, long and wide format conversion, merging files. Hands on exercise, Ethical considerations in large-scale sample surveys

Unit II: Software and Quality assurance procedures

Introduction to STATA for survey data analysis-SVYSET, SVYTAB, SVYMEAN, SVYPROP, SVYTOTAL, SVYLC. Summarization of big DHS data, Conversion of ASCII and SPSS data into STATA format. Hands on exercise. Revisit of sub-samples, field check tables, non-response pattern, and quality lot assurance, roles of supervisors, editors, field and nodal agencies. Third party audit.

Introduction to R: reading ASCII file, data summarization: frequency and graphical representation, survey data summarization using R. Installation of libraries: sampling, survey, samplingbook, pps. Use of svydesign, svytotal, svymean.

Unit III: Use of STATA and R for sampling and estimates

Sampling and estimation by simple random sampling, stratified, cluster, systematic and multi-stage sampling, PPS sampling using STATA and R.

Reading List

Suggested readings:

1. Lumley, T. Complex Surveys: A Guide to Analysis Using R
2. Damico, A. Step-by-step instructions to analyze major public-use survey data sets with the R language
3. Ladusingh, L. Survey Sampling Methods
4. Fares Qeadan. Sampling Methods Using STATA