

BOOK REVIEW

**Sampling: Design and Analysis. Sharon L. Lohr. 2nd Edition,
International Publication, CB, ©2010, 608 Pages ISBN-10:
0495105279; ISBN-13: 9780495105275**

Professor Sharon L. Lohr has already published a number of papers and books devoted to sampling surveys. I have had an opportunity to study some of them, and especially the first edition of her book on *Sampling: Design and Analysis*, published in 1999.

Now I have studied the 2nd edition of the book *Sampling: Design and Analysis*, published in 2010. The book covers many topics not found in other textbooks at this level. The book provides a modern introduction to the field of survey sampling intended for a wide audience of statistics students. The author concentrates on the statistical aspects of taking and analyzing a sample, incorporating a multitude of applications from a variety of disciplines. The text gives guidance on how to tell when a sample is valid or not, and how to design and analyze many different forms of sample surveys. Recent research on theoretical and applied aspects of sampling is included, as well as optional technology instructions for using statistical software with survey data.

Six main features distinguish this book from other texts about sampling methods.

- 1) The book is *accessible to students with a wide range of statistical backgrounds*, and is flexible for content and level. By approximate choice of sections, this book can be used for a first-year graduate course of statistics students or for class with students for business, sociology, psychology or biology who wants to learn about designing and analyzing data from sample surveys. It is also *useful for a person doing research who wants to learn more about statistical aspects of surveys and recent developments*.
- 2) The author used *real data as much as possible*. The examples and exercises come from *social sciences, engineering, agriculture, ecology, medicine* and a variety of other disciplines, and are selected *to illustrate the wide applicability of sampling methods*. A number of data sets have extra variables not specifically referred to the text; an instructor can use these for additional exercises or variations.

- 3) The author has incorporated *model-based* as well as *randomization theory* into the text, with the goal of placing sampling methods within the framework used in other areas of statistics. *Many of the important results in the last years of sampling research* have involved models, and *an understanding of both approaches is essential for the survey practitioners.*
- 4) As I stressed above, *the book covers many topics not found in other textbooks at this level.* Chapters 7 through 15 discuss how to *analyze complex surveys* such as those conducted by different countries¹, computer-intensive methods for estimating variances in complex surveys, *what to do if there is nonresponse*, and how to perform chi-squared tests and regression analyses using data from complex surveys.
- 5) This book emphasized *the importance of graphing the data.* Graphical analysis of survey data is challenging because of the large sizes and complexity of survey datasets but graphs can provide insight into the data structure.
- 6) Design of surveys is emphasized throughout, and is related to methods for analyzing the data from a survey. The book *presents the philosophy that the design* is by far the most important aspect of any survey: *no amount of statistical analysis can be compensated for a badly designed survey.* Models are used to motivate designs, and graphs are presented to check the sensitivity of the design to model assumptions.

Chapters 1 through 6 cover the building blocks of simple random, stratified, and cluster sampling, as well as ratio and regression estimation. To *read them requires* familiarity with *basic ideas of expectation, sampling distributions, confidence intervals, and linear regression* – material covered in most introductory statistics classes. Optional sections on the statistical theory for designs are marked with asterisks – these require to be familiar with calculus and mathematical statistics.

Chapter 7 is devoted to **complex surveys** and deals with assembling *design components, sampling weights, estimating a distribution function, plotting data from a complex survey, design effects*. Chapter 8 deals with **nonresponse** and begins with effects of ignoring nonresponse; *designing surveys to reduce nonsampling errors*; callbacks and two-phase sampling; *mechanisms for nonresponse*; *weighting methods for nonresponse*; *imputation*; parametric models for nonresponse; *what is an acceptable response rate.*

¹ Comp. e.g.: J. Kordos, A. Zięba: Development of Standard Error Estimation Methods in Complex Household Sample Surveys in Poland, *Statistics in Transition - new series*, Vol. 11, No 2, , 2010, pp. 231-252.

The material in *Chapter 9 through 15* can be covered in almost any order, with topics chosen to fit the needs of the students. Appendix A reviews probability concepts.

The second edition introduces some organizational changes to the chapters. The central concept of sampling weights is now introduced in Chapter 2. Stratified sampling has been moved earlier to Chapter 3, preceding ratio and regression estimation. This allows students to become more familiar with the use of weights to account for inclusion probabilities before they are exposed to adjusting weight for calibration. Chapter 9 – ***variance estimation in complex surveys*** – has expanded treatment of computer-intensive methods such as *linearization (Taylor series) methods, jackknife and bootstrap*; generalized variance functions, and confidence intervals.

Material in Chapter 12 of the first edition has been expanded in Chapters 12 to 14 of the second edition. Chapter 15 – ***Quality survey*** – on total survey design is *completely new*, and ties together much of the material in the earlier chapters. I will devote a special attention to this chapter later.

Each chapter now concludes with a *chapter summary*, including key terms and references for further exploration.

The exercises in the second edition have been reordered into four categories in each chapter, with many new exercises added to the book's already extensive problem sets.

- *Introductory Exercises* give more routine problems intended to develop skills at the basic ideas in the book. Many of these would be suitable for hand calculations.
- *Working with Survey Data* exercises ask students to analyze data from real surveys.
- *Working with Theory* exercises are intended for a more mathematically oriented class, allowing students to work through proofs of results in a step-by-step manner and explore the theory of sampling in more depth.

As I have mentioned above, my attention was particularly drawn to chapter 15 on *Survey quality*. As the author has stressed in this chapter, “In many surveys, the margin of error reported is based entirely on the sampling error; nonsampling errors are sometimes acknowledged in the text, but generally are not included in the reported measures of uncertainty”. This is common practice in many countries and the author here concentrates on *survey quality*. First, the author quotes Dalenius (1977, p. 21) who referred to the practice of reporting only sampling error and ignoring other sources of error as “‘strain at a gnat and swallow a camel’; this characterization applies especially to the practice with respect to the accuracy: the sampling error plays the role of the gnat; sometimes

malformed, while the nonsampling error plays the role of the camel, often of unknown size and always of unwieldy shape”.

In this chapter, the author explores approaches to survey design and analysis taking into account **total survey error** which is the sum of five components: *coverage error, nonresponse error, measurement error, processing error and sampling error*. Such approach to *total survey error* does not exist in any other book on survey sampling. Sampling error produces variability in the estimates, and measures only precision of the estimates. It means that we may have very precise and not accurate estimates.

Additionally, I would like to cite Eurostat publication on data quality assessment: methods and tools². This publication underlines that *”production of high quality statistics depends on the assessment of data quality. Without a systematic assessment of data quality, the statistical office will risk to lose control of the various statistical processes such as data collection, editing or weighting. Doing without data quality assessment would result in assuming that the processes cannot be further improved and that problems will always be detected without systematic analysis. At the same time, data quality assessment is a precondition for informing the users about the possible uses of the data, or which results could be published with or without a warning. Certainly, without good approaches for data quality assessment statistical institutes are working in the blind and can make no justified claim of being professional and of delivering quality in the first place. Assessing data quality is therefore one of the core aspects of a statistical institute’s work”.*

For details see contents.

Contents

Chapter 1. Introduction.

- 1.1 A Sample Controversy.
- 1.2 Requirements of a Good Sample.
- 1.3 Selection Bias.
- 1.4 Measurement Error.
- 1.5 Questionnaire Design.
- 1.6 Sampling and Nonsampling Errors.
- 1.7 Exercises.

Chapter 2. Simple Probability Samples.

- 2.1 Types of Probability Samples.
- 2.2 Framework for Probability Sampling.
- 2.3 Simple Random Sampling.

² Eurostat (2007), *Handbook on Data Quality Assessment: Methods and Tools*.

- 2.4 Sampling Weights.
- 2.5 Confidence Intervals.
- 2.6 Sample Size Estimation.
- 2.7 Systematic Sampling.
- 2.8 Randomization Theory Results for Simple Random Sampling.
- 2.9 A Prediction Approach for Simple Random Sampling.
- 2.10 When Should a Simple Random Sample Be Used?
- 2.11 Chapter Summary.
- 2.12 Exercises.

Chapter 3. *Stratified Sampling.*

- 3.1 What Is Stratified Sampling?
- 3.2 Theory of Stratified Sampling.
- 3.3 Sampling Weights in Stratified Random Sampling.
- 3.4 Allocating Observations to Strata.
- 3.5 Defining Strata.
- 3.6 Model-Based Inference for Stratified Sampling.
- 3.7 Quota Sampling.
- 3.8 Chapter Summary.
- 3.9 Exercises.

Chapter 4. *Ratio and Regression Estimation.*

- 4.1 Ratio Estimation in a Simple Random Sample.
- 4.2 Estimation in Domains.
- 4.3 Regression Estimation in Simple Random Sampling.
- 4.4 Poststratification.
- 4.5 Ratio Estimation with Stratified Samples.
- 4.6 Model-Based Theory for Ratio and Regression Estimation.
- 4.7 Chapter Summary.
- 4.8 Exercises.

Chapter 5. *Cluster sampling with equal probabilities.*

- 5.1 Notation for Cluster Sampling.
- 5.2 One-Stage Cluster Sampling.
- 5.3 Two-Stage Cluster Sampling.
- 5.4 Designing a Cluster Sample.
- 5.5 Systematic Sampling.
- 5.6 Model-Based Inference in Cluster Sampling.
- 5.7 Chapter Summary.
- 5.8 Exercises.

Chapter 6. *Sampling with Unequal Probabilities.*

- 6.1 Sampling One Primary Sampling Unit.
- 6.2 One-Stage Sampling with Replacement.
- 6.3 Two-Stage Sampling with Replacement.
- 6.4 Unequal Probability Sampling Without Replacement.
- 6.5 Examples of Unequal Probability Samples. Randomization

- 6.6 Theory Results and Proofs.
- 6.7 Models and Unequal Probability Sampling.
- 6.8 Chapter Summary.
- 6.9 Exercises

Chapter 7. *Complex Surveys.*

- 7.1 Assembling Design Components.
- 7.2 Sampling Weights.
- 7.3 Estimating a Distribution Function.
- 7.4 Plotting Data from a Complex Survey.
- 7.5 Univariate Plots. Design Effects.
- 7.6 The National Crime Victimization Survey.
- 7.7 Sampling and Experiment Design.
- 7.8 Chapter Summary.
- 7.9 Exercises.

Chapter 8. *Nonresponse.*

- 8.1 Effects of Ignoring Nonresponse.
- 8.2 Designing Surveys to Reduce Nonsampling Errors.
- 8.3 Callbacks and Two-Phase Sampling.
- 8.4 Mechanisms for Nonresponse.
- 8.5 Weighting Methods for Nonresponse.
- 8.6 Imputation.
- 8.7 Parametric Models for Nonresponse.
- 8.8 What Is an Acceptable Response Rate?
- 8.9 Chapter Summary.
- 8.10 Exercises.

Chapter 9. *Variance Estimation in Complex Surveys.*

- 9.1 Linearization (Taylor Series) Methods.
- 9.2 Random Group Methods.
- 9.3 Resampling and Replication Methods.
- 9.4 Generalized Variance Functions
- 9.5 Confidence Intervals.
- 9.6 Chapter Summary.
- 9.7 Exercises.

Chapter 10. *Categorical Data Analysis in Complex Surveys.*

- 10.1 Chi-Square Tests with Multinomial Sampling.
- 10.2 Effects of Survey Design on Chi-Square Tests.
- 10.3 Corrections to χ^2 Tests.
- 10.4 Loglinear Models.
- 10.5 Chapter Summary.
- 10.7 Exercises.

Chapter 11. *Regression with Complex Survey Data.*

- 11.1 Model-Based Regression in Simple Random Samples.
- 11.2 Regression in Complex Surveys.

- 11.3 Using Regression to Compare Domain Means.
- 11.4 Should Weights Be Used in Regression?
- 11.5 Mixed Models for Cluster Samples.
- 11.6 Logistic Regression.
- 11.7 Generalized Regression Estimation for Population Totals.
- 11.8 Chapter Summary.
- 11.9 Exercises.

Chapter 12. Two-Phase Sampling.

- 12.1 Theory for Two-Phase Sampling.
- 12.2 Two-Phase Sampling with Stratification.
- 12.3 Two-Phase Sampling with Ratio Estimation.
- 12.4 Subsampling Nonrespondents.
- 12.5 Designing a Two-Phase Sample.
- 12.6 Chapter Summary.
- 12.7 Exercises.

Chapter 13. Estimating Population Size.

- 13.1 Capture-Recapture Estimates.
- 13.2 Multiple Recapture estimation.
- 13.3 Chapter Summary.
- 13.4 Exercises.

Chapter 14. Rare Populations and Small Area Estimation.

- 14.1 Sampling Rare Population.
- 14.2 Small Area Estimation.
- 14.3 Chapter Summary.
- 14.4 Exercises.

Chapter 15. Survey Quality.

- 15.1 Coverage Error.
 - 15.1.1 Measuring Coverage and Coverage Bias.
 - 15.1.2 Coverage and Survey Mode.
 - 15.1.3 Improving Coverage.
- 15.2 Nonresponse Error.
- 15.3 Measurement Error.
 - 15.3.1 Measuring And Modeling Measurement Error
 - 15.3.2 Reducing Measurement Error.
- 15.4 Sensitive Questions.
 - 15.4.1 Nonresponse and Measurement Error.
 - 15.4.2 Randomized Response.
- 15.5 Processing Error.
- 15.6 Total Survey Quality.
- 15.7 Chapter Summary.
- 15.8 Exercises.

Appendices: Probability Concepts Used in Sampling.

Probability.

Random Variables and Expected Value.

Conditional Probability.

Conditional Expectation.

REFERENCES

Summing up, I would like to stress that the book is very comprehensive, covering both fundamental sampling theory and analysis, and real survey examples. The book also covers many of the current survey sampling topics that are not usually covered in standard sampling books, such as sampling weights, nonresponse, complex data analysis, and survey quality.

Information for the Polish statisticians: this book is available at the Central Statistical Library of the Central Statistical Office of Poland.

Prepared by: Jan Kordos, Warsaw School of Economics, e-mail: jan1kor2@aster.pl