

LESLIE KISH

*SURVEY
SAMPLING*

SURVEY SAMPLING

Leslie Kish

*Professor of Sociology
Program Director, Survey Research Center
Institute for Social Research
The University of Michigan*

JOHN WILEY & SONS

New York • Chichester • Brisbane • Toronto • Singapore

Copyright © 1965 by
John Wiley & Sons, Inc.

All Rights Reserved.

Reproduction or translation of any part of this work beyond that permitted by Sections 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

15

Library of Congress Catalog Card Number: 65-19479
Printed in the United States of America

ISBN 0 471 48900 X

SURVEY SAMPLING

Preface

This is designed to be a simple book on sampling methods, with emphasis on and illustrations from surveys of human populations. Although I have designed samples for a variety of nonhuman populations, my experience has been mostly in social applications; the book emphasizes these areas, and it draws most of its examples from them. However, essentially the same methods can be used for sampling animals, plants, minerals, physical products, accounts, and inventories. Sampling specialists find that their skills possess broad generality and transferability; thus this book can also be useful to medical men, biologists, chemists, engineers, and accountants.

Sampling plays a vital role in research design involving human populations; it commands increasing attention from social scientists and practitioners. I include not only the disciplines of economics, sociology, anthropology, psychology, and political science, but also the professions of public health, biostatistics, education, social work, public administration, and business administration. Sampling problems are equally material to practitioners engaged in marketing, commerce, and industry.

While writing this book, I tried to keep a definite audience in mind and to maintain a dialogue with them. My first consideration was for quantitatively oriented students in the social sciences and allied fields. I believe that sampling methods should become roughly the third course of statistics for students in the social sciences. It should follow a course in the fundamentals of statistical reasoning, and another devoted to major statistical tools. I hope that this book can serve as a text for professors of social statistics who are not specialists in sampling.

Second, I have also aimed to provide a reference book on sampling

methods for professors, researchers, and officials who want to understand survey sampling without necessarily becoming sampling specialists. Improving their methods is crucial, since most samples in survey research represent occasional small or medium-sized efforts by researchers who are not specialists.

I have not written this for my colleagues in sampling, but chiefly for my fellow social scientists, whose specialties differ from mine. Reading one book will not turn anyone into an expert sampling statistician; but the careful reader should become able to design and execute valid samples of moderate dimensions and difficulty, to avoid selection biases, and to achieve reasonable efficiency. He should also become more adept at evaluating the sample results he encounters, to judge their validity, their limits of inference, applicability, and precision.

The reader will be able to design satisfactory rather than optimum samples—to *satisfice* (as Herbert Simon says), rather than to optimize. I paid less attention than others to some technical efforts for getting the last 2 percent efficiency out of samples. These efforts may be worthwhile for large, complicated samples—where it may be best to consult an experienced sampling statistician. In that case, the researcher with sufficient knowledge to meet him half-way will be better able to utilize the scarce time of the sampling consultant.

Social research places special emphasis on the comparative and analytical uses of samples. Hence, I have stressed statistics of subclasses in the sample, rather than focusing all attention on the entire sample. Subclasses are treated for each of the major designs; and the comparison of subclass means receives frequent attention. Other issues relating to analytical statistics from complex samples are also discussed. These discussions frequently utilize research developed under a grant (G-7571) from the National Science Foundation.

The book is oriented toward providing a working knowledge of practical sampling methods, with an understanding of their theoretical background. The necessary working formulas are given. I have tried to state clearly the assumptions underlying the formulas, and to outline the potentialities and limitations inherent in the assumptions. I often use illustrations to explain in detail the meaning of formulas and definitions. I also provide a variety of examples, with computations laid out in painstaking detail.

The book contains many practical procedures, the “domestic arts” of sampling along with its science; the valuable “tricks” that are usually learned only in apprenticeship. Perhaps the principal task of any book or course on methods should be the drastic shortening of the period of apprenticeship in a discipline. Throughout the book the reader can

discover my attempts to synthesize complete sets of coherent rules for conduct ordinarily explained—if at all—with *ad hoc* rules-of-thumb.

My efforts are perhaps most evident in methods for dealing with frame problems. They receive a great deal of emphasis, especially in an early summary (2.7) of frame defects and techniques for treating them, and in a detailed treatment in Sections 11.1 to 11.4. Throughout the book the reader's attention is called to possible frame defects and their effects on sample design. Problems invite the reader to participate in finding and overcoming defects he is likely to encounter in research situations.

Sampling methods are not developed here for their own sake, but as means to ends originating in substantive research problems, especially in the social sciences and their applications. They are designed to provide valid, scientific, and economical tools for those research problems. Hence, the approach is directed less toward delivering a neat theoretical package than toward practical research projects in all their complexity, while preserving a fundamental simplicity of presentation.

To achieve simplicity I frequently resort to useful approximations. Briefly, this book attempts to preach what reasonable sampling statisticians actually practice, rather than an arid dogma. For example, the factor $(N - 1)/N$, if inconvenient, can be ignored when the population size N is large. The book emphasizes estimated variances that the researcher actually computes, rather than their expected values. And so on.

I did not try to make this book "self-contained"—an illusory aim. For the algebraic steps of derivations the interested reader is directed, by frequent references, to relevant passages in several fine textbooks. In these references I sought neither to discover original sources, nor the prestige of the most elegant mathematical treatment. Instead, I sought the most relevant, readable, and available sources, preferably textbooks; there the reader can find further references.

However, I am glad that, on the urging of friends, I abandoned my original intention to avoid derivations completely. My derivations are few and simple, yet they cover the essentials of sampling theory adequately for readers who understand certain fundamental concepts of statistical theory. Sampling theory was condensed by eliminating some repetition and detail. The basic derivations are in Sections 2.8, 4.5, 4.6, 5.6, 6.6, 8.5, and in scattered remarks. These and other technical portions carry a (*); continuity can be maintained without them.

My original intention was to formalize and disseminate the notes that my students and I have evolved from teaching a one-semester course. This restricted but vital core is presented in the simple, nontechnical, carefully organized and related sections of Chapters 1 to 8 in Part I: Fundamentals of Survey Sampling. To complete the course, I add

Sections 9.1 to 9.5 on area sampling, plus most of Chapter 13 on non-sampling errors. Brief selections may also be added from Chapters 11 and 12, which deal with special techniques.

These two chapters and the technical portions of others have served as the base for advanced courses in special problems. However, they are designed primarily to serve as sources of reference for individual, technical problems. These sections are mostly self-contained, rather than interconnected. I hope that researchers will often find in one of them ready answers to specific technical problems.

Those who can avoid the complexities of cluster sampling will find a self-contained treatment of element sampling in Chapters 1 to 4. To this can be added a modest treatment of the essentials of cluster sampling: 6.4B, 6.5A and B, 7.2, 7.3, and 9.1 to 9.3. These could comprise a third of a course, incorporated into a full year course in statistics.

For a briefer introduction to sampling, teachers have used the following sections of 81 pages: 1.0–1.7, 2.6, 2.7, 3.1, 3.4, 3.5, 5.1, 5.4, 6.4B, 6.5A, 7.2, 7.3, 9.1–9.3.

I hope that readers will especially like the problems. They should give much credit to my students, who eliminated the worst, stimulated the best, and sharpened their phrasing; also some credit to my many colleagues and clients, whose questions and research projects gave birth to most of these problems. I tried hard to make them life-like; hence, the student's crucial task is to discover what the principal issues are, then pose appropriate questions. When this is well done, the answers come rather smoothly. The focus is on principal issues, rather than on fine nuances. This was made possible by eliminating cumbersome details, while retaining the essence of actual problems. In addition to testing the student's ability to apply the book's methods, these problems perform a double task. They deepen understanding of those methods by extending them to new and varied areas. Furthermore, the extensions themselves represent valuable methodological tools. Work on these problems should yield half the value the student can extract from this book. Of a subject devoted to methods, we can say with Sophocles: "One must learn by doing the thing; for though you think you know it, you have no certainty, until you try."

It is with genuine pleasure that I acknowledge my debt to many friends and colleagues. Since Yates' book in 1948, six fine books on sampling have appeared in English, and I have learned and taught from each in turn. Deming, Cochran, Hansen, and Hurwitz have been personal teachers, friends, and consultants. I was fortunate to have, at different times, Roe Goodman, Benjamin Tepping, and Irene Hess as colleagues in the Sampling Section of the Survey Research Center. Angus Campbell and Rensis Likert, in creating and directing that Center and the Institute

for Social Research, have provided valuable support and opportunity. Questions, problems, and issues posed daily by the social scientists at the Center provide a uniquely stimulating atmosphere. I also benefited from daily contact with colleagues in the Sociology and Mathematics Departments. Of the many statisticians who have given me personal stimulation, support, and criticism, I must single out L. J. Savage, Bruce Hill, and Howard Raiffa; William Ericson gave me valuable criticism on much of the final draft, tightening many a loose phrase or formulation. They stimulated and helped me to make this book more suitable as a text in departments of statistics than I first planned. V. K. Sethi and C. T. Tharakan eliminated many technical errors. Above all, I was fortunate to find in my wife, Rhea, an expert and devoted editor, who has greatly improved the book's style.

Primary and essential stimulation came from the many students in about forty courses in sampling I have given in fifteen years. They represent a good cross section of researchers, teachers, practitioners, and students from the social sciences, mathematics, and allied fields. The students in the seventeen yearly Summer Institutes on Survey Techniques have been mostly officials, scientists, and professors; they came from all over the globe, representing yearly an average of ten countries. Their experience, searching questions, suggestions, and arguments have contributed substantially to the origin, form, and content of this book.

Leslie Kish

Ann Arbor, Mich.

June 1965

SURVEY SAMPLING

Contents

Part I FUNDAMENTALS OF SURVEY SAMPLING	1
1. Introduction	3
1.0 Guide to the Introduction	3
1.1 Survey Design and Sample Design	4
1.2 A Taxonomy of Survey Units and Concepts	6
1.3 Population Values and Statistics	9
1.4 Statistical Inference in Surveys	14
1.5 A Taxonomy of Selection Procedures	17
1.6 Criteria of Sample Design	23
1.7 Mechanical Selection, Randomization, and Frames	26
2. Basic Concepts of Sampling	35
2.0 Some Basic Symbols	35
2.1 Procedures for Simple Random Sampling (SRS)	36
2.2 Mean and Variance of SRS	40
2.3 The "fpc"; Sampling without Replacement	43
2.4 Proportions; the Means of Binomials	45
2.5 Relative Error	47
2.6 The Design of Economic Samples	49
2.7 Frame Problems	53
*2.8 Basic Formulas and Derivations	59
2.8A Some Fundamentals	59
2.8B Variance in Simple Random Sampling Without Replacement	62
2.8C The Expectations of Some Important Types of Samples	64
2.8D Variances for Linear Combinations	69
3. Stratified Sampling	75
3.1 Definition and Purposes	75
3.2 The Weighted Mean and Its Variance	77
3.3 Mean and Variance for Stratified Element Sampling	80
3.4 Proportionate Sampling of Elements	82
3.4A The Mean and Variance of Proportionate Samples	82
3.4B The Design of Proportionate Samples	86
3.4C Post-Stratification or Stratification after Selection	90

3.5	Disproportionate Sampling or Optimum Allocation	92
3.6	Forming the Strata	98
3.6A	Classifying Sampling Units	98
3.6B	Stratification with Random "Quotas"	99
3.6C	Errors in Sorting	99
3.6D	Overlapping Lists	99
3.6E	Objectivity and Regularity Unnecessary	100
3.6F	Homogeneity within Strata	100
3.6G	Utilizing Available Variables	100
3.6H	Elaborate Controls Unnecessary	101
3.6I	Number of Strata	102
*3.6J	Effect of Increasing Numbers of Strata	102
3.6K	Equal Allocation	102
*3.6L	Optimum Stratification	104
4.	Systematic Sampling; Stratification Techniques	113
4.1	Procedures and Uses of Systematic Selection	113
4.1A	Uses of Systematic Samples	113
4.1B	Problems with Intervals	115
4.1C	Variances for Systematic Samples	117
4.2	Problems of Systematic Selection	120
4.3	Paired Selections	123
4.4	Replicated Sampling	127
4.5	Subclass Means, Totals, and Comparisons in Stratified Samples	132
4.5A	Subclass Means, Totals, and Their Variances	132
*4.5B	Derivations of Variances for Stratified Subclasses	136
*4.5C	Optimum Allocation for Subclasses	138
*4.6	Precision of Proportionate and Disproportionate Samples	139
4.6A	Precision of Proportionate Samples	139
4.6B	Precision of Optimum Allocation	142
5.	Cluster Sampling and Subsampling	148
5.1	Nature of Clusters	148
5.2	Random Choice of Clusters	151
5.3	Subsampling in Multistage Sampling	154
5.3A	The Aims of Subsampling	154
5.3B	Features of Simple Replicated Subsampling	156
5.4	Effects of Clustering; Intraclass Correlation, Roh	161
5.5	Stratified Cluster Sampling	164
*5.6	Components of the Variance	166
5.6A	Components of Two-Stage Random Sampling	166
5.6B	The Intraclass Correlation	170
5.6C	Tables for Analysis of Variance	173
6.	Unequal Clusters	182
6.1	Problems of Unequal Clusters	182
6.2	Random Selection of Unequal Clusters; Epsem Subsampling	184
6.3	Variances for Random Unequal Clusters	187
6.4	Stratified Sampling of Unequal Clusters	190
6.4A	Any Number (a_h) of Clusters	190
6.4B	Paired Selection of Clusters	193
6.5	Statistics for Stratified Unequal Clusters	195
6.5A	The Difference of Two Ratio Means	195

6.5B	Simple Variances for Complex Samples	197
6.5C	Variances for Systematic Primary Selections	202
6.5D	Expansions with Ratio Estimates	203
6.5E	Separate Ratio Estimates	206
*6.6	The Ratio Mean	206
6.6A	Variance of the Ratio Mean	206
6.6B	Bias of the Ratio Mean	208
6.6C	The Difference of Two Ratio Means	209
6.6D	The Product $y_{\bar{x}}$ and Relvariances	211
7.	Selection with Probabilities Proportional to Size Measures (PPS)	217
7.1	Control of Sample Size	217
7.2	Control of Subsample Size with PPS	220
7.3	Paired Selection of Primaries	223
7.4	Selecting Primaries without Replacement	227
7.4A	Single Selection from Random Half-Strata	227
*7.4B	Selecting Two Primaries without Replacement	229
7.4C	Systematic Selection of Primaries with PPS	230
7.4D	Single Selection from Each Stratum	231
*7.4E	Three or More Primaries per Stratum	232
7.5	Probabilities Proportional to Size; Problems and Procedures	234
7.5A	Selection with PPS Measures	234
7.5B	Subselection within Primary Units	238
7.5C	Variations in Actual Subsample Size	239
7.5D	Sampling Fractions, Strata and Measure Adjustments	241
7.5E	Undersized and Oversized Clusters	243
7.5F	Fractional Systematic Selection Intervals	246
8.	The Economic Design of Surveys	254
8.1	Planned Precision Based on Unit Variance	254
8.2	Estimates of Unit Variance; Design Effect	257
8.3	Models of Cost Functions	263
8.3A	A General Model of Cost Factors	263
8.3B	Specific Cost Function for Cluster Samples	268
8.4	Practicality	272
8.4A	Simple Designs	272
8.4B	Practical Field Instructions	274
8.4C	Random Numbers for Field Work	276
8.4D	Sequential Control of Sample Size	277
*8.5	Optimum Designs	279
*8.6	Techniques for Computing Variances	282
8.6A	Replicated Subsampling	282
8.6B	Collapsed Strata	283
8.6C	Special Techniques for Variance Computations	286
8.6D	Precision of Variance Estimates	289
8.6E	Components of the Variance	291
Part II	SPECIAL PROBLEMS AND TECHNIQUES	299
9.	Area Sampling	301
9.1	Area Frames for Dwellings	301
9.2	Preparing Maps	304

9.3	Selection Rates for Blocks and Dwellings	308
9.4	Some Practical Problems	313
9.4A	Compact Segments versus Listed Dwellings	313
9.4B	Problems of Block Size	316
9.4C	Supplements for the New, the Missed, and the Unusual	316
9.4D	Repeated Selection from a Listing	318
9.4E	Modification to Three or More Stages	319
9.4F	Maps, Photos, and Similar Materials	320
9.5	Compact Segments	322
9.5A	Procedures for Compact Segments	322
9.5B	Instructions for Segmenting	325
9.5C	Instructions for Interviewing in Segments	329
9.5D	Segmented Listing	331
9.6	Listing Dwellings	334
9.6A	Instructions for Listing Dwellings	334
9.6B	Selecting Listed Dwellings	338
9.6C	Identifying Dwellings at Sample Addresses	339
9.7	Creating Segments from Listed Buildings	343
9.7A	Special Features	343
9.7B	Instructions for Selecting Blocks	344
9.7C	Instructions for Listing Buildings	344
9.7D	Segmenting Blocks and Selecting Segments	348
9.7E	Instructions for Sampling the Segments	349
9.8	Sampling City Directories	352
9.8A	City Directories as Frames for Dwellings	352
9.8B	Selecting from Directories	354
9.8C	Instructions for Sample Dwellings at Selected Addresses	355
9.8D	Instructions for Block Supplement for Missed Dwellings	356
10.	<i>Multistage Sampling</i>	359
10.1	Three-Stage Area Samples	359
10.2	A National Sample of Primary Areas (Counties)	363
10.3	Sampling a County	370
10.3A	Measures for Major Strata	370
10.3B	Uniform Design for All Strata	372
10.3C	Separate Procedures for Strata	374
10.3D	The County as Part of a National Sample	374
10.4	The Current Population Survey	377
11.	<i>Sampling from Imperfect Frames</i>	384
11.1	Empty Listings and Foreign Elements; Variable Frame Densities	384
11.2	Duplicate Listings; Overlapping Frames	388
11.2A	How to Avoid the Problem	388
11.2B	Creating Unique Listings	389
11.2C	Weighting the Selection with $1/p_i$	390
11.2D	Matching from Overlapping Lists	394
11.3	Selecting Persons from Dwellings; Small Unequal Clusters	396
11.3A	Dwellings as Clusters of Persons	396
11.3B	Selecting One Adult from the Household	398
11.3C	Epsm Selection of Persons	401
11.4	Rare Elements; High Skewness	404

11.4A	Selection Techniques for Rare Traits	404
11.4B	Highly Skewed Populations	410
11.4C	Screening with Field Ratings	412
11.5	Supplements for Incomplete Frames	415
11.6	Observational Units of Variable Sizes	418
11.7	Weighting Problems	424
11.7A	Aims and Procedures of Weighting	424
11.7B	Losses from Random Duplication and Elimination	427
11.7C	Losses from Oversampling Strata	429
11.8	Estimating Totals	433
11.8A	The Advantages of $N\bar{y}$ over Fy	433
11.8B	Totals from Subclasses of SRS	434
11.8C	The Product Estimate $\bar{N}\bar{y}$ of Aggregates	436
12.	Some Selection Techniques	440
*12.1	Two-Phase Sampling in General and in Stratification	440
12.1A	The Economics of Two-Phase Sampling	440
12.1B	Stratification in Two-Phase Sampling	443
12.1C	Complex Two-Phase Selections	444
*12.2	Two-Phase Ratio and Regression Means	446
12.2A	Two-Phase Ratio Means	446
12.2B	Two-Phase Regression Estimates	448
12.2C	Comparisons of Regression, Ratio, and Stratified Means in Two-Phase Sampling	450
*12.3	Regression Means; Comparisons to Ratio and Difference Means	451
12.3A	Regression Means	451
12.3B	Comparisons with Ratio and Difference Means	454
12.3C	Stratified Regression Means	455
12.4	Correlations from Overlaps in Repeated Surveys	457
12.4A	Variances of Differences between Two Overlapping Samples	457
12.4B	Gains from Overlaps in Measuring Change	462
12.4C	Sums and Current Estimates from Overlapping Samples	466
12.5	Panel Studies and Designs for Measuring Changes	469
12.5A	Panels versus New Samples	469
12.5B	Partial Overlaps in Rotation Designs	471
12.5C	Choice of Overlapping Units	472
12.5D	Splitting a Large Survey into Repeated Samples; Time Sampling	474
12.6	Continuing Sampling Operations	477
12.6A	Master Frames	478
12.6B	Techniques for Cumulating Information	480
12.6C	A Stratum for "Surprises"	481
12.6D	Inertias of Continuing Operations	482
12.7	Changing Selection Probabilities	483
12.7A	A Procedure for Minimum Changes of Selections	483
12.7B	Modifications of the Procedure	485
12.7C	A Simple Procedure for Burgeoning Units	486
12.7D	Shifting Units between Strata	487
*12.8	Multiple Stratification, Lattice Sampling, Controlled Selection	488
12.9	Standard Errors for Medians and Quantiles	495
12.10	Trinomials and Matched Binomials	497

*12.11 Standard Errors for Combinations of Ratio Means	501
12.11A Linear Combinations of Ratio Means	501
12.11B Double Ratios; Comparisons; Indexes	503
Part III RELATED CONCEPTS	507
13. Biases and Nonsampling Errors	509
13.1 Relation of Bias to Variable Error	509
13.2 Sources of Survey Errors	514
13.2A Biases and Variable Errors	514
13.2B Descriptions of Diverse Errors	519
13.2C Some Effects of Errors	524
13.3 Noncoverage; Incomplete Frames, Missing Units	527
13.4 Nonresponse	532
13.4A Sources of Nonresponse	532
13.4B Effects of Nonresponse	535
13.4C Some Nonresponse Results and Call-Back Data	536
13.5 Control of Nonresponse	548
13.5A Methods and Aims of Control	548
13.5B Call-Backs	550
13.5C Subsampling Nonresponses	556
13.6 Four Proposed Remedies for Nonresponse	557
13.6A Estimation of Effects	557
13.6B Substitutions for Nonresponses	558
13.6C The Politz Scheme	559
13.6D A Replacement Procedure	560
13.7 Quota Sampling	562
*13.8 Effects of Bias on Probability Statements	566
14. Some Issues of Inference from Survey Data	574
14.1 Computation and Presentation of Sampling Errors	574
14.2 Analytical Statistics for Complex Samples	582
14.3 Some Remarks about Statistical Inference	587
14.4 Experiments and Surveys	593
14.5 Multiple Objectives; Multipurpose Surveys	597
APPENDICES	603
A Summary of Symbols and Terms	605
B Unit Analysis of Statistical Formulas	614
C Remarks on Computations	617
D Table of Random Numbers	622
E List of a Population of 270 Blocks	624
References	627
Answers to Selected Problems	637
Index	639

PART I
FUNDAMENTALS OF
SURVEY SAMPLING