STATISTICS:
a series of TEXTBOOKS and MONOGRAPHS

Survey Sampling
Theory and Methods
Second Edition

Arijit Chaudhuri
Horst Stenger

Survey Sampling

Statistical Distributions
Professor N. Balakrishnan
McMaster University

Statistical Process Improvement
Professor G. Geoffrey Vining
Virginia Polytechnic Institute

Stochastic Processes
Professor V. Lakshmikantham
Florida Institute of Technology

Survey Sampling
Professor Lynne Stokes
Southern Methodist University

Time Series
Sastry G. Pantula
North Carolina State University

1. The Generalized Jackknife Statistic, *H. L. Gray and W. R. Schucany*
2. Multivariate Analysis, *Anant M. Kshirsagar*
3. Statistics and Society, *Walter T. Federer*
4. Multivariate Analysis: A Selected and Abstracted Bibliography, 1957–1972, *Kocherlakota Subrahmaniam and Kathleen Subrahmaniam*
5. Design of Experiments: A Realistic Approach, *Virgil L. Anderson and Robert A. McLean*
6. Statistical and Mathematical Aspects of Pollution Problems, *John W. Pratt*
7. Introduction to Probability and Statistics (in two parts), Part I: Probability; Part II: Statistics, *Narayan C. Giri*
8. Statistical Theory of the Analysis of Experimental Designs, *J. Ogawa*
9. Statistical Techniques in Simulation (in two parts), *Jack P. C. Kleijnen*
10. Data Quality Control and Editing, *Joseph I. Naus*
11. Cost of Living Index Numbers: Practice, Precision, and Theory, *Kali S. Banerjee*
12. Weighing Designs: For Chemistry, Medicine, Economics, Operations Research, Statistics, *Kali S. Banerjee*
13. The Search for Oil: Some Statistical Methods and Techniques, *edited by D. B. Owen*
14. Sample Size Choice: Charts for Experiments with Linear Models, *Robert E. Odeh and Martin Fox*
15. Statistical Methods for Engineers and Scientists, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*
16. Statistical Quality Control Methods, *Irving W. Burr*

17. On the History of Statistics and Probability, *edited by D. B. Owen*
18. Econometrics, *Peter Schmidt*
19. Sufficient Statistics: Selected Contributions, *Vasant S. Huzurbazar (edited by Anant M. Kshirsagar)*
20. Handbook of Statistical Distributions, *Jagdish K. Patel, C. H. Kapadia, and D. B. Owen*
21. Case Studies in Sample Design, *A. C. Rosander*
22. Pocket Book of Statistical Tables, *compiled by R. E. Odeh, D. B. Owen, Z. W. Birnbaum, and L. Fisher*
23. The Information in Contingency Tables, *D. V. Gokhale and Solomon Kullback*
24. Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods, *Lee J. Bain*
25. Elementary Statistical Quality Control, *Irving W. Burr*
26. An Introduction to Probability and Statistics Using BASIC, *Richard A. Groeneveld*
27. Basic Applied Statistics, *B. L. Raktoe and J. J. Hubert*
28. A Primer in Probability, *Kathleen Subrahmaniam*
29. Random Processes: A First Look, *R. Syski*
30. Regression Methods: A Tool for Data Analysis, *Rudolf J. Freund and Paul D. Minton*
31. Randomization Tests, *Eugene S. Edgington*
32. Tables for Normal Tolerance Limits, Sampling Plans and Screening, *Robert E. Odeh and D. B. Owen*
33. Statistical Computing, *William J. Kennedy, Jr., and James E. Gentle*
34. Regression Analysis and Its Application: A Data-Oriented Approach, *Richard F. Gunst and Robert L. Mason*
35. Scientific Strategies to Save Your Life, *I. D. J. Bross*
36. Statistics in the Pharmaceutical Industry, *edited by C. Ralph Buncher and Jia-Yeong Tsay*
37. Sampling from a Finite Population, *J. Hajek*
38. Statistical Modeling Techniques, *S. S. Shapiro and A. J. Gross*
39. Statistical Theory and Inference in Research, *T. A. Bancroft and C.-P. Han*
40. Handbook of the Normal Distribution, *Jagdish K. Patel and Campbell B. Read*
41. Recent Advances in Regression Methods, *Hrishikesh D. Vinod and Aman Ullah*
42. Acceptance Sampling in Quality Control, *Edward G. Schilling*
43. The Randomized Clinical Trial and Therapeutic Decisions, *edited by Niels Tygstrup, John M Lachin, and Erik Juhl*
44. Regression Analysis of Survival Data in Cancer Chemotherapy, *Walter H. Carter, Jr., Galen L. Wampler, and Donald M. Stablein*
45. A Course in Linear Models, *Anant M. Kshirsagar*
46. Clinical Trials: Issues and Approaches, *edited by Stanley H. Shapiro and Thomas H. Louis*
47. Statistical Analysis of DNA Sequence Data, *edited by B. S. Weir*
48. Nonlinear Regression Modeling: A Unified Practical Approach, *David A. Ratkowsky*
49. Attribute Sampling Plans, Tables of Tests and Confidence Limits for Proportions, *Robert E. Odeh and D. B. Owen*

50. Experimental Design, Statistical Models, and Genetic Statistics, *edited by Klaus Hinkelmann*
51. Statistical Methods for Cancer Studies, *edited by Richard G. Cornell*
52. Practical Statistical Sampling for Auditors, *Arthur J. Wilburn*
53. Statistical Methods for Cancer Studies, *edited by Edward J. Wegman and James G. Smith*
54. Self-Organizing Methods in Modeling: GMDH Type Algorithms, *edited by Stanley J. Farlow*
55. Applied Factorial and Fractional Designs, *Robert A. McLean and Virgil L. Anderson*
56. Design of Experiments: Ranking and Selection, *edited by Thomas J. Santner and Ajit C. Tamhane*
57. Statistical Methods for Engineers and Scientists: Second Edition, Revised and Expanded, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*
58. Ensemble Modeling: Inference from Small-Scale Properties to Large-Scale Systems, *Alan E. Gelfand and Crayton C. Walker*
59. Computer Modeling for Business and Industry, *Bruce L. Bowerman and Richard T. O'Connell*
60. Bayesian Analysis of Linear Models, *Lyle D. Broemeling*
61. Methodological Issues for Health Care Surveys, *Brenda Cox and Steven Cohen*
62. Applied Regression Analysis and Experimental Design, *Richard J. Brook and Gregory C. Arnold*
63. Statpal: A Statistical Package for Microcomputers—PC-DOS Version for the IBM PC and Compatibles, *Bruce J. Chalmer and David G. Whitmore*
64. Statpal: A Statistical Package for Microcomputers—Apple Version for the II, II+, and IIe, *David G. Whitmore and Bruce J. Chalmer*
65. Nonparametric Statistical Inference: Second Edition, Revised and Expanded, *Jean Dickinson Gibbons*
66. Design and Analysis of Experiments, *Roger G. Petersen*
67. Statistical Methods for Pharmaceutical Research Planning, *Sten W. Bergman and John C. Gittins*
68. Goodness-of-Fit Techniques, *edited by Ralph B. D'Agostino and Michael A. Stephens*
69. Statistical Methods in Discrimination Litigation, *edited by D. H. Kaye and Mikel Aickin*
70. Truncated and Censored Samples from Normal Populations, *Helmut Schneider*
71. Robust Inference, *M. L. Tiku, W. Y. Tan, and N. Balakrishnan*
72. Statistical Image Processing and Graphics, *edited by Edward J. Wegman and Douglas J. DePriest*
73. Assignment Methods in Combinatorial Data Analysis, *Lawrence J. Hubert*
74. Econometrics and Structural Change, *Lyle D. Broemeling and Hiroki Tsurumi*
75. Multivariate Interpretation of Clinical Laboratory Data, *Adelin Albert and Eugene K. Harris*
76. Statistical Tools for Simulation Practitioners, *Jack P. C. Kleijnen*
77. Randomization Tests: Second Edition, *Eugene S. Edgington*

78. A Folio of Distributions: A Collection of Theoretical Quantile-Quantile Plots, *Edward B. Fowlkes*
79. Applied Categorical Data Analysis, *Daniel H. Freeman, Jr.*
80. Seemingly Unrelated Regression Equations Models: Estimation and Inference, *Virendra K. Srivastava and David E. A. Giles*
81. Response Surfaces: Designs and Analyses, *Andre I. Khuri and John A. Cornell*
82. Nonlinear Parameter Estimation: An Integrated System in BASIC, *John C. Nash and Mary Walker-Smith*
83. Cancer Modeling, *edited by James R. Thompson and Barry W. Brown*
84. Mixture Models: Inference and Applications to Clustering, *Geoffrey J. McLachlan and Kaye E. Basford*
85. Randomized Response: Theory and Techniques, *Arijit Chaudhuri and Rahul Mukerjee*
86. Biopharmaceutical Statistics for Drug Development, *edited by Karl E. Peace*
87. Parts per Million Values for Estimating Quality Levels, *Robert E. Odeh and D. B. Owen*
88. Lognormal Distributions: Theory and Applications, *edited by Edwin L. Crow and Kunio Shimizu*
89. Properties of Estimators for the Gamma Distribution, *K. O. Bowman and L. R. Shenton*
90. Spline Smoothing and Nonparametric Regression, *Randall L. Eubank*
91. Linear Least Squares Computations, *R. W. Farebrother*
92. Exploring Statistics, *Damaraju Raghavarao*
93. Applied Time Series Analysis for Business and Economic Forecasting, *Sufi M. Nazem*
94. Bayesian Analysis of Time Series and Dynamic Models, *edited by James C. Spall*
95. The Inverse Gaussian Distribution: Theory, Methodology, and Applications, *Raj S. Chhikara and J. Leroy Folks*
96. Parameter Estimation in Reliability and Life Span Models, *A. Clifford Cohen and Betty Jones Whitten*
97. Pooled Cross-Sectional and Time Series Data Analysis, *Terry E. Dielman*
98. Random Processes: A First Look, Second Edition, Revised and Expanded, *R. Syski*
99. Generalized Poisson Distributions: Properties and Applications, *P. C. Consul*
100. Nonlinear L_p-Norm Estimation, *Rene Gonin and Arthur H. Money*
101. Model Discrimination for Nonlinear Regression Models, *Dale S. Borowiak*
102. Applied Regression Analysis in Econometrics, *Howard E. Doran*
103. Continued Fractions in Statistical Applications, *K. O. Bowman and L. R. Shenton*
104. Statistical Methodology in the Pharmaceutical Sciences, *Donald A. Berry*
105. Experimental Design in Biotechnology, *Perry D. Haaland*
106. Statistical Issues in Drug Research and Development, *edited by Karl E. Peace*
107. Handbook of Nonlinear Regression Models, *David A. Ratkowsky*

108. Robust Regression: Analysis and Applications, *edited by Kenneth D. Lawrence and Jeffrey L. Arthur*
109. Statistical Design and Analysis of Industrial Experiments, *edited by Subir Ghosh*
110. *U*-Statistics: Theory and Practice, *A. J. Lee*
111. A Primer in Probability: Second Edition, Revised and Expanded, *Kathleen Subrahmaniam*
112. Data Quality Control: Theory and Pragmatics, *edited by Gunar E. Liepins and V. R. R. Uppuluri*
113. Engineering Quality by Design: Interpreting the Taguchi Approach, *Thomas B. Barker*
114. Survivorship Analysis for Clinical Studies, *Eugene K. Harris and Adelin Albert*
115. Statistical Analysis of Reliability and Life-Testing Models: Second Edition, *Lee J. Bain and Max Engelhardt*
116. Stochastic Models of Carcinogenesis, *Wai-Yuan Tan*
117. Statistics and Society: Data Collection and Interpretation, Second Edition, Revised and Expanded, *Walter T. Federer*
118. Handbook of Sequential Analysis, *B. K. Ghosh and P. K. Sen*
119. Truncated and Censored Samples: Theory and Applications, *A. Clifford Cohen*
120. Survey Sampling Principles, *E. K. Foreman*
121. Applied Engineering Statistics, *Robert M. Bethea and R. Russell Rhinehart*
122. Sample Size Choice: Charts for Experiments with Linear Models: Second Edition, *Robert E. Odeh and Martin Fox*
123. Handbook of the Logistic Distribution, *edited by N. Balakrishnan*
124. Fundamentals of Biostatistical Inference, *Chap T. Le*
125. Correspondence Analysis Handbook, *J.-P. Benzécri*
126. Quadratic Forms in Random Variables: Theory and Applications, *A. M. Mathai and Serge B. Provost*
127. Confidence Intervals on Variance Components, *Richard K. Burdick and Franklin A. Graybill*
128. Biopharmaceutical Sequential Statistical Applications, *edited by Karl E. Peace*
129. Item Response Theory: Parameter Estimation Techniques, *Frank B. Baker*
130. Survey Sampling: Theory and Methods, *Arijit Chaudhuri and Horst Stenger*
131. Nonparametric Statistical Inference: Third Edition, Revised and Expanded, *Jean Dickinson Gibbons and Subhabrata Chakraborti*
132. Bivariate Discrete Distribution, *Subrahmaniam Kocherlakota and Kathleen Kocherlakota*
133. Design and Analysis of Bioavailability and Bioequivalence Studies, *Shein-Chung Chow and Jen-pei Liu*
134. Multiple Comparisons, Selection, and Applications in Biometry, *edited by Fred M. Hoppe*
135. Cross-Over Experiments: Design, Analysis, and Application, *David A. Ratkowsky, Marc A. Evans, and J. Richard Alldredge*
136. Introduction to Probability and Statistics: Second Edition, Revised and Expanded, *Narayan C. Giri*

137. Applied Analysis of Variance in Behavioral Science, *edited by Lynne K. Edwards*
138. Drug Safety Assessment in Clinical Trials, *edited by Gene S. Gilbert*
139. Design of Experiments: A No-Name Approach, *Thomas J. Lorenzen and Virgil L. Anderson*
140. Statistics in the Pharmaceutical Industry: Second Edition, Revised and Expanded, *edited by C. Ralph Buncher and Jia-Yeong Tsay*
141. Advanced Linear Models: Theory and Applications, *Song-Gui Wang and Shein-Chung Chow*
142. Multistage Selection and Ranking Procedures: Second-Order Asymptotics, *Nitis Mukhopadhyay and Tumulesh K. S. Solanky*
143. Statistical Design and Analysis in Pharmaceutical Science: Validation, Process Controls, and Stability, *Shein-Chung Chow and Jen-pei Liu*
144. Statistical Methods for Engineers and Scientists: Third Edition, Revised and Expanded, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*
145. Growth Curves, *Anant M. Kshirsagar and William Boyce Smith*
146. Statistical Bases of Reference Values in Laboratory Medicine, *Eugene K. Harris and James C. Boyd*
147. Randomization Tests: Third Edition, Revised and Expanded, *Eugene S. Edgington*
148. Practical Sampling Techniques: Second Edition, Revised and Expanded, *Ranjan K. Som*
149. Multivariate Statistical Analysis, *Narayan C. Giri*
150. Handbook of the Normal Distribution: Second Edition, Revised and Expanded, *Jagdish K. Patel and Campbell B. Read*
151. Bayesian Biostatistics, *edited by Donald A. Berry and Dalene K. Stangl*
152. Response Surfaces: Designs and Analyses, Second Edition, Revised and Expanded, *André I. Khuri and John A. Cornell*
153. Statistics of Quality, *edited by Subir Ghosh, William R. Schucany, and William B. Smith*
154. Linear and Nonlinear Models for the Analysis of Repeated Measurements, *Edward F. Vonesh and Vernon M. Chinchilli*
155. Handbook of Applied Economic Statistics, *Aman Ullah and David E. A. Giles*
156. Improving Efficiency by Shrinkage: The James-Stein and Ridge Regression Estimators, *Marvin H. J. Gruber*
157. Nonparametric Regression and Spline Smoothing: Second Edition, *Randall L. Eubank*
158. Asymptotics, Nonparametrics, and Time Series, *edited by Subir Ghosh*
159. Multivariate Analysis, Design of Experiments, and Survey Sampling, *edited by Subir Ghosh*
160. Statistical Process Monitoring and Control, *edited by Sung H. Park and G. Geoffrey Vining*
161. Statistics for the 21st Century: Methodologies for Applications of the Future, *edited by C. R. Rao and Gábor J. Székely*
162. Probability and Statistical Inference, *Nitis Mukhopadhyay*

163. Handbook of Stochastic Analysis and Applications, *edited by D. Kannan and V. Lakshmikantham*
164. Testing for Normality, *Henry C. Thode, Jr.*
165. Handbook of Applied Econometrics and Statistical Inference, *edited by Aman Ullah, Alan T. K. Wan, and Anoop Chaturvedi*
166. Visualizing Statistical Models and Concepts, *R. W. Farebrother and Michael Schyns*
167. Financial and Actuarial Statistics, *Dale Borowiak*
168. Nonparametric Statistical Inference, Fourth Edition, Revised and Expanded, *edited by Jean Dickinson Gibbons and Subhabrata Chakraborti*
169. Computer-Aided Econometrics, *edited by David EA. Giles*
170. The EM Algorithm and Related Statistical Models, *edited by Michiko Watanabe and Kazunori Yamaguchi*
171. Multivariate Statistical Analysis, Second Edition, Revised and Expanded, *Narayan C. Giri*
172. Computational Methods in Statistics and Econometrics, *Hisashi Tanizaki*
173. Applied Sequential Methodologies: Real-World Examples with Data Analysis, *edited by Nitis Mukhopadhyay, Sujay Datta, and Saibal Chattopadhyay*
174. Handbook of Beta Distribution and Its Applications, *edited by Richard Guarino and Saralees Nadarajah*
175. Item Response Theory: Parameter Estimation Techniques, Second Edition, *edited by Frank B. Baker and Seock-Ho Kim*
176. Statistical Methods in Computer Security, *William W. S. Chen*
177. Elementary Statistical Quality Control, Second Edition, *John T. Burr*
178. Data Analysis of Asymmetric Structures, *edited by Takayuki Saito and Hiroshi Yadohisa*
179. Mathematical Statistics with Applications, *Asha Seth Kapadia, Wenyaw Chan, and Lemuel Moyé*
180. Advances on Models Character and Applications, *N. Balakrishnan and I. G. Bayramov*
181. Survey Sampling: Theory and Methods, Second Edition, *Arijit Chaudhuri and Horst Stenger*

Survey Sampling
Theory and Methods
Second Edition

Arijit Chaudhuri
Indian Statistical Institute
Calcutta, India

Horst Stenger
University of Manheim
Manheim, Germany

Chapman & Hall/CRC
Taylor & Francis Group
Boca Raton London New York Singapore

Published in 2005 by
Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 0-82475-754-8 (Hardcover)
International Standard Book Number-13: 978-0-8247-5754-0 (Hardcover)
Library of Congress Card Number 2004058264

Library of Congress Cataloging-in-Publication Data

Chaudhuri, Arijit, 1940-
Survey sampling : theory and methods / Arijit Chaudhuri, Horst Stenger.—2nd ed.
p. cm. -- (Statistics, textbooks and monographs ; v. 181)
Includes bibliographical references and index.
ISBN 0-82475-754-8
1. Sampling (Statistics) I. Stenger, Horst, 1935- II. Title. III. Series.

QA276.6.C43 2005
519.5'2--dc22 2004058264

T&F informa
Taylor & Francis Group
is the Academic Division of T&F Informa plc.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Foreword

ARIJIT CHAUDHURI and HORST STENGER are well known in sampling theory. The present book further confirms their reputation. Here the authors have undertaken the large task of surveying the sampling literature of the past few decades to provide a reference book for researchers in the area. They have done an excellent job. Starting with the unified theory the authors very clearly explain subsequent developments. In fact, even the most modern innovations of survey sampling, both methodological and theoretical, have found a place in this concise volume. In this connection I may specially mention the authors' presentation of estimating functions. With its own distinctiveness, this book is indeed a very welcome addition to the already existing rich literature on survey sampling.

V. P. GODAMBE
University of Waterloo
Waterloo, Ontario, Canada

Preface to the Second Edition

It is gratifying that our Publishers engaged us to prepare this second edition. Since our first edition appeared in 1992, *Survey Sampling* acquired a remarkable growth to which we, too, have made a modest contribution. So, some addition seems due. Meanwhile, we have received feedback from our readers that prompts us to incorporate some modifications.

Several significant books of relevance have emerged after our write-up for the first edition went to press that we may now draw upon, by the following authors or editors: SÄRNDAL, SWENSSON and WRETMAN (1992), BOLFARINE and ZACKS (1992), S. K. THOMPSON (1992), GHOSH and MEEDEN (1986), THOMPSON and SEBER (1996), M. E. THOMPSON, (1997) GODAMBE (1991), COX (1991) and VALLIANT, DORFMAN and ROYALL (2000), among others.

Numerous path-breaking research articles have also appeared in journals keeping pace with this phenomenal progress. So, we are blessed with an opportunity to enlighten ourselves with plenty of new ideas. Yet we curb our impulse to cover the salient aspects of even a sizeable section of this current literature. This is because we are not inclined to reshape

the essential structure of our original volume and we are aware of the limitations that prevent us from such a venture.

As in our earlier presentation, herein we also avoid being dogmatic—more precisely, we eschew taking sides. *Survey Sampling* is at the periphery of mainstream statistics. The speciality here is that we have a tangible collection of objects with certain features, and there is an intention to pry into them by getting hold of some of these objects and attempting an inference about those left untouched. This inference is traditionally based on a theory of probability that is used to exploit a possible link of the observed with the unobserved. This probability is not conceived as in statistics, covering other fields, to characterize the interrelation of the individual values of the variables of our interest. But this is created by a survey sampling investigator through arbitrary specification of an artifice to select the samples from the populations of objects with preassigned probabilities. This is motivated by a desire to draw a representative sample, which is a concept yet to be precisely defined. Purposive selection (earlier purported to achieve representativeness) is discarded in favor of this sampling design-based approach, which is theoretically admitted as a means of yielding a legitimate inference about an aggregate from a sampled segment and also valued for its objectivity, being free of personal bias of a sampler. NEYMAN's (1934) pioneering masterpiece, followed by survey sampling texts by YATES (1953), HANSEN, HURWITZ and MADOW (1953), DEMING (1954) and SUKHATME (1954), backed up by exquisitely executed survey findings by MAHALANOBIS (1946) in India as well as by others in England and the U.S., ensured an unstinted support of probability sampling for about 35 years.

But ROYALL (1970) and BREWER (1963) installed a rival theory dislodging the role of the selection probability as an inferential tool in survey sampling. This theory takes off postulating a probability model characterizing the possible links among the observed and the unobserved variate values associated with the survey population units. The parameter of the surveyor's inferential concern is now a random variable rather than a constant. Hence it can be predicted, not estimated.

The basis of inference here is this probability structure as modeled.

Fortunately, the virtues of some of the sampling design-supported techniques like stratification, ratio method of estimation, etc., continue to be upheld by this model-based prediction theory as well. But procedures for assessing and measuring the errors in estimation and prediction and setting up confidence intervals do not match.

The design-based approach fails to yield a best estimator for a total free of design-bias. By contrast, a model-specific best predictor is readily produced if the model is simple, correct, and plausible. If the model is in doubt one has to strike a balance over bias versus accuracy. A procedure that works well even with a wrong model and is thus robust is in demand with this approach. That requires a sample that is adequately balanced in terms of sample and population values of one or more variables related to one of the primary inferential interest. For the design-based classical approach, currently recognized performers are the estimators motivated by appropriate prediction models that are design-biased, but the biases are negligible when the sample sizes are large. So, a modern compromise survey approach called model assisted survey sampling is now popular. Thanks to the pioneering efforts by SÄRNDAL (1982) and his colleagues the generalized regression (GREG) estimators of this category are found to be very effective in practice.

Regression modeling motivated their arrival. But an alternative calibration approach cultivated since the early nineties by ZIESCHANG (1990), DEVILLE and SÄRNDAL (1992), and others renders them purely design-based as well with an assured robustness or riddance from model-dependence altogether.

A predictor for a survey population total is a sum of the sampled values plus the sum of the predictors for the unsampled ones. A design-based estimator for a population total, by contrast, is a sum of the sampled values with multiplicative weights yielded by specific sampling designs. A calibration approach adjusts these initial sampling weights, the new weights keeping close to them but satisfying certain

consistency constraints or calibration equations determined by one or more auxiliary variables with known population totals.

This approach was not discussed in the first edition but is now treated at length. Adjustments here need further care to keep the new weights within certain plausible limits, for which there is considerable documentation in the literature. Here we also discuss a concern for outliers—a topic which also recommends adjustments of sampling weights. While calibration and restricted calibration estimators remain asymptotically design unbiased (ADU) and asymptotically design consistent (ADC), the other adjusted ones do not.

Earlier we discussed the QR predictors, which include (1) the best predictors, (2) projection estimators, (3) generalized regression estimators, and (4) the cosmetic predictors for which (1) and (3) match under certain conditions. Developments since 1992 modify QR predictors into restricted QR predictors (RQR) as we also recount.

SÄRNDAL (1996), DEVILLE (1999), BREWER (1999a, 1999b), and BREWER and GREGOIRE (2000) are prescribing a line of research to justify omission of the cross-product terms in the quadratic forms, giving the variance and mean square error (MSE) estimators of linear estimators of population totals, by suitable approximations. In this context SÄRNDAL (1996) makes a strong plea for the use of generalized regression estimators based either on stratified (1) simple random sampling (SRS) or (2) Bernoulli sampling (BS), which is a special case of Poisson sampling devoid of cross-product terms. This encourages us to present an appraisal of Poisson sampling and its valuable ramifications employing permanent random numbers (PRN), useful in coordination and exercise of control in rotational sampling, a topic we omitted earlier.

Among other novelties of this edition we mention the following. We give essential complements to our earlier discussion of the minimax principle. In the first edition, exact results were presented for completely symmetric situations and approximate results for large populations and samples. Now, following STENGER and GABLER (1996) an exact minimax

property of the expansion estimator in connection with the LAHIRI-MIDZUNO-SEN design is presented for arbitrary sample sizes.

An exact minimax property of a Hansen-Hurwitz estimator proved by GABLER and STENGER (2000) is reviewed; in this case a rather complicated design has to be applied, as sample sizes are arbitrary.

A corrective term is added to SEN (1953) and YATES and GRUNDY's (1953) variance estimator to make it unbiased even for non-fixed-sample-size designs with an easy check for its uniform non-negativity, as introduced by CHAUDHURI and PAL (2002). Its extension to cover the generalized regression estimator analogously to HORVITZ and THOMPSON's (1952) estimator is but a simple step forward.

In multistage sampling DURBIN (1953), RAJ (1968) and J. N. K. RAO's (1975a) formulae for variance estimation need expression in general for single-stage variance formulae as quadratic forms to start with, a condition violated in RAJ (1956), MURTHY (1957) and RAO, HARTLEY and COCHRAN (1962) estimators, among others. Utilizing commutativity of expectation operators in the first and later stages of sampling, new simple formulae are derived bypassing the above constraint following CHAUDHURI, ADHIKARI and DIHIDAR (2000a, 2000b).

The concepts of borrowing strength, synthetic, and empirical Bayes estimation in the context of developing small domain statistics were introduced in the first edition. Now we clarify how in two-stage sampling an estimator for the population total may be strengthened by employing empirical Bayes estimators initiated through synthetic versions of GREG estimators for the totals of the sampling clusters, which are themselves chosen with suitable unequal probabilities. A new version of cluster sampling developed by CHAUDHURI and PAL (2003) is also recounted.

S. K. THOMPSON (1992) and THOMPSON and SEBER's (1996) adaptive and network sampling techniques have been shown by CHAUDHURI (2000a) to be generally applicable for any sampling scheme in one stage or multistages with or without stratification. It is now illustrated how adaptive sampling

may help the capture of rare units with appropriate network formations; vide CHAUDHURI, BOSE and GHOSH (2003).

In the first edition as well as in the text by CHAUDHURI and MUKERJEE (1988), randomized response technique to cover qualitative features was restricted to simple random sampling with replacement (SRSWR) alone. Newly emerging extension procedures to general sampling designs are now covered.

In the first edition we failed to cover SITTER's (1992a, 1992b) mirror-match and extended BWO bootstrap procedures and discussed RAO and WU's (1985, 1988) rescaled bootstrap only cursorily; we have extended coverage on them now.

Circular systematic sampling (CSS) with probability proportional to size (PPS) is known to yield zero inclusion probabilities for paired units. But this defect may now be removed on allowing a random, rather than a predetermined, sampling interval—a recent development, which we now cover. Barring these innovations and a few stylistic repairs the second edition mimics the first.

Of course, the supplementary references are added alphabetically. We continue to remain grateful to the same persons and institutions mentioned in the first edition for their sustained support.

In addition, we wish to thank Mrs. Y. CHEN for typing and organizing typesetting of the manuscript.

ARIJIT CHAUDHURI
HORST STENGER

Preface to the First Edition

Our subject of attention is a finite population with a known number of identifiable individuals, bearing values of a characteristic under study. The main problem is to estimate the population total or mean of these values by surveying a suitably chosen sample of individuals. An elaborate literature has grown over the years around various criteria for appropriate sampling designs and estimators based on selected samples so designed. We cover this literature selectively to communicate to the reader our appreciation of the current state of development of essential aspects of theory and methods of survey sampling.

Our aim is to reach graduate and advanced level students of sampling and, at the same time, researchers in the area looking for a reference book. Practitioners will be interested in many techniques of sampling that, we believe, are not adequately covered in most textbooks. We have avoided details of foundational aspects of inference in survey sampling treated in the texts by CASSEL, SÄRNDAL and WRETMAN (1977) and CHAUDHURI and VOS (1988).

In the first four chapters we state fundamental results and provide proofs of many propositions, although often

leaving some of them incomplete purposely in order to save space and invite our readers to fill in the gaps themselves. We have taken care to keep the level of discussion within reach of the average graduate-level student.

The first four chapters constitute the core of the book. Although not a prerequisite, they are nevertheless helpful in giving motivations for numerous theoretical and practical problems of survey sampling dealt with in subsequent chapters, which are rather specialized and indicate several lines of approach. We have collected widely scattered materials in order to aid researchers in pursuing further studies in areas of specific interest. The coverage is mostly review in nature, leaving wide gaps to be bridged with further reading from sources cited in the References.

In chapter 1 we first formulate the problem of getting a good point estimator for a finite population total. We suppose the number of individuals is known and each unit can be assigned an identifying label. Consequently, one may choose an appropriate sample of these labels. It is assumed that unknown values can be ascertained for the individuals sampled. First we discuss the classical design-based approach of inference and present GODAMBE (1955) and GODAMBE and JOSHI's (1965) celebrated theorems on nonexistence of the best estimator of a population total. The concepts of likelihood and sufficiency and the criteria of admissibility, minimaxity, and completeness of estimators and strategies are introduced and briefly reviewed. Uses and limitations of well-known superpopulation modeling in finding serviceable sampling strategies are also discussed. But an innovation worth mentioning is the introduction of certain preliminaries on GODAMBE's (1960b) theory of estimating equations. We illustrate its application to survey sampling, bestowing optimality properties on certain sampling strategies traditionally employed ad hoc.

The second chapter gives RAO and VIJAYAN's (1977) procedure of mean square error estimation for homogeneous linear estimators and mentions several specific strategies to which it applies.

The third chapter introduces ROYALL's (1970) linear prediction approach in sampling. Here one does not speculate

about what may happen if another sample is drawn with a preassigned probability. On the contrary, the inference is based on speculation on the possible nature of the finite population vector of variate values for which one may postulate plausible models. It is also shown how and why one needs to revise appropriate predictors and optimal purposive sampling designs to guard against possible mis-specifications in models and, at the same time, seek to employ robust but nonoptimal procedures that work well even when a model is inaccurately hypothesized. This illustrates how these sampling designs may be recommended when a model is correctly but simplistically postulated. Later in the chapter, Bayes estimators for finite population totals based on simplistic priors are mentioned and requirements for their replacements by empirical Bayes methods are indicated with examples. Uses of the JAMES–STEIN technique on borrowing strength from allied sources are also emphasized, especially when one has inadequate sample data specific to a given situation.

In chapter 4 we first note that if a model is correctly postulated, a design-unbiased strategy under the model may be optimal yet poorer than a comparable optimal predictive strategy. On the other hand, the optimal predictive strategy is devoid of design-based properties and modeling is difficult. Hence the importance of relaxing design-unbiasedness for the design-based strategy and replacing the optimal predictive strategy by a nonoptimal robust alternative enriched with good design properties. The two considerations lead to inevitable asymptotics. We present, therefore, contemporary activities in exploring competitive strategies that do well under correct modeling but continue to have desirable asymptotic design-based features in case of model failures. Although achieving robustness is a guiding motive in this presentation, we do not repeat here alternative robustness preserving techniques, for example, due to GODAMBE (1982). However, the asymptotic approaches for minimax sampling strategies are duly reported to cover recently emerging developments.

In chapter 5 we address the problem of mean square error estimation covering estimators and predictors and we follow procedures that originate from twin considerations of designs

and models. In judging comparative efficacies of competing procedures one needs to appeal to asymptotics and extensive empirical investigations demanding Monte Carlo simulations; we have illustrated some of the relevant findings of established experts in this regard.

Chapter 6 is intended to supplement a few recent developments of topics concerning multistage, multiphase, and repetitive sampling. The time series methods applicable for a fuller treatment are not discussed.

Chapter 7 recounts a few techniques for variance estimation involving nonlinear estimators and complex survey designs including stratification, clustering, and selection in stages.

The next chapter deals with specialized techniques needed for domain estimation, poststratification, and estimation from samples taken using inadequate frames. The chapter emphasizes the necessity for conditional inference involving speculation over only those samples having some recognizable features common with the sample at hand.

Chapter 9 introduces the topic of analytic rather than descriptive studies where the center of attention is not the survey population at hand but something that lies beyond and typifies it in some discernible respect. Aspects of various methodologies needed for regression and categorical data analyses in connection with complex sampling designs are discussed as briefly as possible.

Chapter 10 includes some accounts of methods of generating randomized data and their analyses when there is a need for protected privacy relating to sensitive issues under investigation.

Chapter 11 presents several methods of analyzing survey data when there is an appreciable discrepancy between those gathered and those desired. The material presented is culled intensively from the three-volume text on incomplete data by MADOW et al. (1983) and from KALTON's (1983a,b) texts and other sources mentioned in the references.

The concluding chapter sums up our ideas about inference problems in survey sampling.

We would like to end with the following brief remarks. In employing a good sampling strategy it is important to acquire knowledge about the background of the material under investigation. In light of the background information at one's command one may postulate models characterizing some of the essential features of the population on which an inference is to be made. While employing the model one should guard against its possible incorrectness and hence be ready to take advantage of the classical design-based approach in adjusting the inference procedures. While deriving in full the virtue of design-based arguments one should also examine if appropriate conditional inference is applicable in case some cognizable features common to the given sample are discernible. This would allow averaging over them instead of over the entire set of samples.

ARIJIT CHAUDHURI gratefully acknowledges the facilities for work provided at the Virginia Polytechnic Institute and University of Mannheim as a visiting professor and the generosity of the Indian Statistical Institute in granting him the necessary leave and opportunities for joint research with his coauthor. He is also grateful to his wife, Mrs. BINATA CHAUDHURI, for her nonacademic but silent help.

HORST STENGER gratefully acknowledges the support of the Deutsche Forschungsgemeinschaft offering the opportunity of an intensive cooperation with the coauthor. His thanks also go to the Indian Statistical Institute, where joint research could be continued. In addition, he wishes to thank Mrs. R. BENT, Mrs. H. HARYANTO, and, especially, Mrs. P. URBAN, who typed the manuscript through many versions.

Comments on inaccuracies and flaws in our presentation will be appreciated and necessary corrective measures are promised for any future editions.

ARIJIT CHAUDHURI
HORST STENGER

The Authors

ARIJIT CHAUDHURI is a CSIR (Council of Scientific and Industrial Research) Emeritus Scientist and a visiting professor at the Applied Statistics Unit, Indian Statistical Institute in Kolkata, India, where he served as a professor from 1982 to 2002. He has served as a visiting professor at the Virginia Polytechnic Institute and State University, the University of Nebraska — Lincoln, the University of Mannheim, Germany and other institutes abroad. He is the chairman of the Advanced Survey Research Centre in Kolkata and a life member of the Indian Statistical Institute, the Calcutta Statistical Association, and the Indian Society of Agricultural Statistics.

Dr. CHAUDHURI holds a Ph.D. in statistics from Calcutta University, and undertook a postdoctoral fellowship for two years at the University of Sydney. He has published more than 100 research papers in numerous journals and is the coauthor of three research monographs: the first edition of the current

volume (1992), *Randomized Response: Theory and Techniques* (with Rahul Mukerjee) (1988), *Unified Theory and Strategies of Survey Sampling* (with J.W.E. Vos) (1988).

HORST STENGER, professor of statistics at the University of Mannheim, Germany, has written several journal articles and two books on survey sampling, *Stichprobentheorie* (1971) and *Stichproben* (1986). He is also the coauthor of three books on general statistics, *Grundlagen der Statistik* (1978, 1979) with A. Anderson, W. Popp, M. Schaffranek, K. Szameitat; *Bevölkerungs- und Wirtschaftsstatistik* (1983) with A. Anderson, M. Schaffranek, K. Szameitat; and *Schätzen und Testen* (1976, 1997), with A. Anderson, W. Popp, M. Schaffranek, D. Steinmetz.

Dr. STENGER is a member of the International Statistical Institute, the American Statistical Association and the Deutsche Statistische Gesellschaft. He received the Dr. rer. nat. degree (1965) in mathematical statistics and the habilitation qualification (1967) in statistics from the University of Munich, Germany. From 1967 to 1971 he was professor of statistics and econometrics at the University of Göttingen, Germany. He has been a visiting professor at the Indian Statistical Institute, Calcutta.

Contents

Chapter 1

Estimation in Finite Populations: A Unified Theory

1.1 INTRODUCTION

Suppose it is considered important to gather ideas about, for example, (1) the total quantity of food grains stocked in all the godowns managed by a state government, (2) the total number of patients admitted in all the hospitals of a country classified by varieties of their complaints, (3) the amount of income tax evaded on an average by the income earners of a city. Now, to inspect all godowns, examine all admission documents of all hospitals of a country, and make inquiries about all income earners of a city will be too expensive and time consuming. So it seems natural to select a few godowns, hospitals, and income earners, to get all relevant data for them and to be able to draw conclusions on those quantities that could be ascertained exactly only by a survey of all godowns, hospitals, and income earners. We feel it is useful to formulate mathematically as follows the essentials of the issues at hand common to the above and similar circumstances.

1.2 ELEMENTARY DEFINITIONS

Let N be a known number of units, e.g., godowns, hospitals, or income earners, each assignable identifying labels $1, 2, \ldots, N$ and bearing values, respectively, $Y_1, Y_2, \ldots, Y_N$ of a real-valued variable y, which are initially unknown to an investigator who intends to estimate the total

$$Y = \sum_{1}^{N} Y_i$$

or the mean $\overline{Y} = Y/N$.

We call the sequence $U = (1, \ldots, N)$ of labels a **population**. Selecting units leads to a sequence $s = (i_1, \ldots, i_n)$, which is called a **sample**. Here $i_1, \ldots, i_n$ are elements of U, not necessarily distinct from one another but the **order of its appearance** is maintained. We refer to $n = n(s)$ as the **size** of s, while the **effective sample size** $\nu(s) = |s|$ is the cardinality of s, i.e., the number of distinct units in s. Once a specific sample s is chosen we suppose it is possible to ascertain the values $Y_{i_1}, \ldots, Y_{i_n}$ of y associated with the respective units of s. Then

$$d = [(i_1, Y_{i_1}), \ldots, (i_n, Y_{i_n})] \quad \text{or briefly}$$
$$d = [(i, Y_i) | i \in s]$$

constitutes the **survey data**.

An **estimator** t is a real-valued function $t(d)$, which is free of Y_i for $i \notin s$ but may involve Y_i for $i \in s$. Sometimes we will express $t(d)$ alternatively by $t(s, \underline{Y})$, where $\underline{Y} = (Y_1, \ldots, Y_N)'$.

An estimator of special importance for $\overline{Y}$ is the **sample mean**

$$t(s, \underline{Y}) = \frac{1}{n(s)} \sum_{i=1}^{N} f_{si} Y_i = \overline{y}, \text{ say}$$

where f_{si} denotes the frequency of i in s such that

$$\sum_{i=1}^{N} f_{si} = n(s).$$

$N\overline{y}$ is called the **expansion estimator** for Y.

More generally, an estimator t of the form

$$t(s, \underline{Y}) = b_s + \sum_{i=1}^{N} b_{si} Y_i$$

with $b_{si} = 0$ for $i \notin s$ is called **linear** (L). Here b_s and b_{si} are free of $\underline{Y}$. Keeping $b_s = 0$ we obtain a **homogeneous linear** (HL) estimator.

We must emphasize that here $t(\underline{s}, \underline{Y})$ is linear (or homogeneous linear) in $Y_i, i \in s$. It may be a nonlinear function of two random variables, e.g., when $b_s = 0$ and $b_{si} = X/\Sigma_1^N f_{si} X_i$ so that

$$t(s, \underline{Y}) = \frac{\sum_1^N f_{si} Y_i}{\sum_1^N f_{si} X_i} X.$$

Here, X_i is the value of a variable x on $i \in \mathcal{U}$ and $X = \Sigma_1^N X_i$ (see section 2.2.)

In what follows we will assume that a sample is drawn at **random**, i.e., with each sample s is associated a selection probability $p(s)$. A **design** p may depend on related variables x, z, etc. But we assume, unless explicitly mentioned otherwise, that p is free of $\underline{Y}$. To emphasize this freedom, p is often referred to in the literature as a **noninformative design**.

If p involves any component of $\underline{Y}$ it is an **informative design**.

A design p is **without replacement** (WOR) if no repetitions occur in any s with $p(s) > 0$; otherwise, p is called **with replacement** (WR). A design p is of **fixed size** n (**fixed effective size** n) if $p(s) > 0$ implies that s is of size n (of effective size n). With respect to WOR designs there is, of course, no difference between fixed size and fixed effective size.

A design p is called **simple random sampling without replacement** (SRSWOR) if

$$p(s) = \frac{1}{\binom{N}{n} n!}$$

for s of size n without repetitions, while it is called **simple random sampling with replacement** (SRSWR) if

$$p(s) = \frac{1}{N^n}$$

for every s of size n, n fixed in advance.

The combination (p, t) denoting an estimator t based on s chosen according to a design p is called a **strategy**. Sometimes a redundant epithet **sampling** is used before design and strategy but we will avoid this usage.

Whatever $\underline{Y}$ may be, let

$$E_p(t) = \sum_s t(s, \underline{Y}) p(s)$$

denote the **expectation** of t and

$$M_p(t) = E_p(t - Y)^2 = \sum_s p(s)(t(s, \underline{Y}) - Y)^2$$

the **mean square error** (MSE) of t. If $E_p(t) = Y$ for every $\underline{Y}$, then t is called a **p-unbiased estimator** (UE) of Y. In this case $M_p(t)$ becomes the **variance** of t and is written

$$V_p(t) = E_p(t - E_p(t))^2.$$

For an arbitrary design p, consider the **inclusion probabilities**

$$\pi_i = \sum_{s \ni i} p(s)\,;\ i = 1, 2, \ldots, N$$

$$\pi_{ij} = \sum_{s \ni i,j} p(s)\,;\ i \neq j = 1, 2, \ldots, N$$

and, provided $\pi_1, \pi_2, \ldots, \pi_N > 0$, the **Horvitz–Thompson** (HT) **estimator** (HTE)

$$\bar{t} = \sum_{i \in s} \frac{Y_i}{\pi_i}$$

(see HORVITZ and THOMPSON, 1952) where the sum is over $|s|$ terms while s is of length $n(s)$. It is easily seen that $\bar{t}$ is HL and p-unbiased (HLU) for Y.

REMARK 1.1 *To mention another way to write* $\bar{t}$ *define*

$$I_{si} = \begin{cases} 1 & \text{if} \quad i \in s \\ 0 & \text{if} \quad i \notin s \end{cases}$$

for $i = 1, 2, \ldots, N$. *Then*

$$\bar{t} = \bar{t}(s, \underline{Y}) = \sum_{i=1}^{N} I_{si} \frac{Y_i}{\pi_i}.$$

where the sum is over $i = 1, 2, \ldots, N$

REMARK 1.2 *Assume* $i_0 \in U$ *exists with* $\pi_{i_0} = 0$ *for a design* p. *Then, for an estimator* t

$$E_p t = \sum_{s \ni i_0} p(s) t(s, \underline{Y}) + \sum_{s \not\ni i_0} p(s) t(s, \underline{Y}).$$

The second term on the right of this equation is obviously free of Y_{i_0}. *Since* $p(s) = 0$ *for all* s *with* $i_0 \in s$, *the first term is* 0. *Hence,* $E_p t$ *is free of* Y_{i_0} *and, especially, not equal to* $Y = \Sigma_1^N Y_i$. *Consequently, no* p*-unbiased estimator exists.*

1.3 DESIGN-BASED INFERENCE

Let Σ_1 be the sum over samples for which $|t(s, \underline{Y}) - Y| \geq k > 0$ and let Σ_2 be the sum over samples for which $|t(s, \underline{Y}) - Y| < k$ for a fixed $\underline{Y}$. Then from

$$\begin{aligned} M_p(t) &= \Sigma_1 p(s)(t - Y)^2 + \Sigma_2 p(s)(t - Y)^2 \\ &\geq k^2 \text{Prob}\big[|t(s, \underline{Y}) - Y| \geq k\big] \end{aligned}$$

one derives the **Chebyshev inequality**:

$$\text{Prob}[|t(s, \underline{Y}) - Y| \geq k] \leq \frac{M_p(t)}{k^2}.$$

Hence

$$\text{Prob}[t - k \leq Y \leq t + k] \geq 1 - \frac{M_p(t)}{k^2} = 1 - \frac{1}{k^2}\big[V_p(t) + B_p^2(t)\big]$$

where $B_p(t) = E_p(t) - Y$ is the **bias** of t. Writing $\sigma_p(t) = \sqrt{V_p(t)}$ for the standard error of t and taking $k = 3\sigma_p(t)$, it follows that, whatever $\underline{Y}$ may be, the random interval $t \pm 3\sigma_p(t)$

covers the unknown Y with a probability not less than

$$\frac{8}{9} - \frac{1}{9}\frac{B_p^2(t)}{V_p(t)}.$$

So, to keep this probability high and the length of this covering interval small it is desirable that both $|B_p(t)|$ and $\sigma_p(t)$ be small, leading to a small $M_p(t)$ as well.

EXAMPLE 1.1 *Let y be a variable with values 0 and 1 only. Then, as a consequence of $Y_i^2 = Y_i$,*

$$\begin{aligned}\sigma_{yy} &= \frac{1}{N}\sum(Y_i - \overline{Y})^2 \\ &= \overline{Y}(1-\overline{Y}) \le \frac{1}{4}.\end{aligned}$$

Therefore, with p SRSWR of size n,

$$\begin{aligned}V_p(N\overline{y}) &= N^2\frac{\sigma_{yy}}{n} \\ &\le \frac{N^2}{4n}.\end{aligned}$$

From

$$E_p\,\overline{y} = \overline{Y}$$

we derive that the random interval

$$N\,\overline{y} \pm 3\sqrt{N^2\frac{1}{4n}} = N\left[\overline{y} \pm \frac{3}{2\sqrt{n}}\right]$$

covers the unknown $N\overline{Y}$ with a probability of at least 8/9.

It may be noted that $\underline{Y}$ is regarded as fixed (nonstochastic) and s is a random variable with a probability distribution $p(s)$ that the investigator adopts at pleasure. It is through p alone that for a fixed $\underline{Y}$ the interval $t \pm 3\sigma_p(t)$ is a random interval. In practice an upper bound of $\sigma_p(t)$ may be available, as in the above example, or $\sigma_p(t)$ is estimated from survey data d plus auxiliary information by, for example, $\hat{\sigma}_p(t)$ inducing necessary changes in the above confidence statements.

If $|B_t(t)|$ is small, then we may argue that the average value of t over repeated sampling according to p is numerically close to Y and, if $M_p(t)$ is small, then we may say that

the average square error $E_p(t - Y)^2$ calculated over repeated sampling according to p is small.

Let us stress this point more fully. The parameter to be estimated may be written as $Y = \Sigma_s Y_i + \Sigma_r Y_i$, the sums being over the distinct units sampled and the remaining units of U, respectively. Its estimator is

$$t = \sum_s Y_i + \left(t - \sum_s Y_i \right).$$

Now, t is close to Y for a sample s at hand and the realized survey data $d = (i, Y_i \mid i \in s)$ if and only if $(t - \Sigma_s Y_i)$ is close to $\Sigma_r Y_i$, the first expression depending on Y_i for $i \in s$ and the second determined by Y_j for $j \notin s$. Now, so far we permit $\underline{Y}$ to be any vector of real numbers without any restrictions on the structural relationships among its coordinates. In this **fixed population setup** we have no way to claim or disclaim the required closeness of $(t - \Sigma_s Y_i)$ and $\Sigma_r Y_i$ for a given sample s. But we need a link between Y_i for $i \in s$ and Y_j for $j \notin s$ in order to provide a base on which our inference about Y from realized data d may stand. Such a link is established by the hypothesis of **repeated sampling**. The resulting **design-based** (briefly: p-based) theory following NEYMAN (1934) is developed around the faith that it is desirable and satisfactory to assess the performance of the strategy (p, t) over repeated sampling, even if in practice a sample will really be drawn once, yielding a single value for t.

This theory is unified in the sense that the performance of a strategy (p, t) is evaluated in terms of the characteristics $E_p(t)$ and $M_p(t)$, such that there is no need to refer to specific selection procedures.

1.4 SAMPLING SCHEMES

A unified theory is developed by noting that it is enough to establish results concerning (p, t) without heeding how one may actually succeed in choosing samples with preassigned probabilities. A method of choosing a sample draw by draw, assigning selection probabilities with each draw, is called a