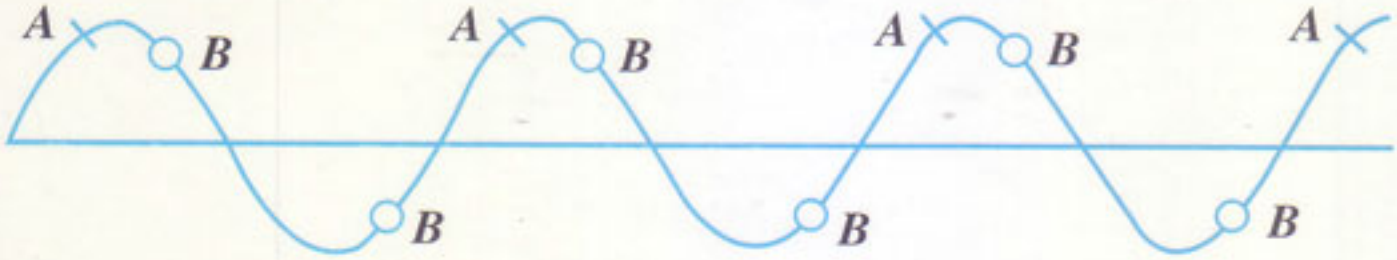
Third Edition

M. Nurul Islam

An Introduction to SAMPLING METHODS

Theory and Applications



Mullick & Brothers

ABOUT THE AUTHOR

Dr. M. Nurul Islam is former Selection Grade Professor of Statistics, Faculty of Science at the University of Dhaka, Bangladesh. Currently Dr. Islam is Pro-Vice Chancellor at the World University of Bangladesh. He joined the University of Dhaka as a Lecturer in 1971 and acted as Chairman of the Department of Statistics during 1985-1987. He held the coveted position of the Vice Chancellor of the Mawlana Bhashani Science and Technology University in Tangail during 2009-2013. He is an elected member of famed International Statistical Institute (ISI) of the Netherlands. Dr. Islam is now holding the position of President of the Bangladesh Statistical Association. He also held the same poison of the Association during 2010–2012. Dr. Islam was an elected member of the Dhaka University Senate for the period 2002-2004. He was also a part-time member of University Grants Commission during 2010-2012. During his 45 years' career as a teacher and researcher, he has written nearly one hundred scientific papers and survey reports related to his discipline, which have been published in both local and international journals. He has supervised a number of M.Sc and PhD works. He provided consultancy services to USAID, DANIDA, UNICEF, CIDA, CARE, SIDA, NETZ, Bangladesh Bureau of Statistics (BBS), Ministry of Health and Family Welfare and Ministry of Textile. He independently conducted a number of field surveys as a Principal Investigator. He extensively toured a number of countries including USA, UK, Canada, Egypt, Indonesia, Singapore, Nepal and India and presented papers there in conferences and seminars. He is author of four textbooks entitled (i) **An Introduction to Statistics and Probability, (ii) An Introduction to Research Methods, (iii) An Introduction to Sampling Methods and (IV) An Introduction to Demographic Techniques** published by Mullick & Brothers. The author has also published his entire research works in a single volume entitled **Demographic Research in Bangladesh: An Update,** with the financial assistance from Social Science Research Council, Ministry of Planning. He has been an Executive Editor of several national and international peer reviewed journals. Dr. Islam is also the recipient of the UGC Award and the Dhaka University Faculty Award for 2007 and 2004 respectively for two of his incomparable textbooks.

CONTENTS

		IX
	PREFACE	XI
1	**INTRODUCTION**	**1**
1.1	Concept of Sampling...	1
1.2	Importance of Sampling...	2
1.3	Sampling: Historical Perspectives...	3
1.4	Technical Terms...	4
1.5	Sampling Frame...	7
1.6	Census and Survey...	7
1.7	Steps in Planning and Executing a Sample Survey	8
1.8	Advantages of Sampling over Complete Count	12
1.9	Limitations of Sampling...	13
1.10	Sample Design and Survey Design...	13
1.11	Evaluation of a Sample Design...	14
1.12	Sampling with and without Replacement... ...	16
1.13	Confidence Interval...	18
1.14	Probability and Non-probability Sample...	19
1.15	Sampling Distribution...	20
2	**SIMPLE RANDOM SAMPLING**	**24**
2.1	Introduction...	24
2.2	Properties of Simple Random Sampling... ...	26
2.3	Drawing a Simple Random Sample...	29
2.4	Estimation of Population Mean and Total	38
2.5	Properties of the Estimators...	39
2.6	Relative Error of the Estimators...	48
2.7	Improving the Estimates...	58
2.8	Advantages and Disadvantages of SRS... ...	59
3	**MORE ON SIMPLE RANDOM SAMPLING**	**63**
3.1	Introduction...	63
3.2	Preliminaries...	63
3.3	Variance of the Estimators...	65
3.4	Estimating Sample Size in SRS...	78
3.5	Developing a Formula for Sample Size...	82
3.6	Sample Size for Estimating Population Proportion...	83
3.7	Sample Size for Continuous Data...	87
3.8	Problems in Estimating C_y^{2}	91
3.9	Probability Proportional to Size Sampling in SRS	92

3.10	Estimated Variance for PPS Sampling...	96
3.11	Design Effect...	102
4	**STRATIFIED SAMPLING**	**108**
4.1	Concept of Stratification...	108
4.2	Principles of Stratification...	110
4.3	Steps Involved in Stratified Sampling...	112
4.4	Estimators and Their Properties...	114
4.5	Confidence Interval...	131
4.6	Estimating Sample Size...	134
4.7	Allocating Sample Size to Strata...	139
4.8	Disproportionate Stratified Sampling...	144
4.9	Precision of Stratified Sampling...	162
4.10	Advantages and Disadvantages of Stratified Sampling...	164
5	**RATIO ESTIMATES**	**171**
5.1	Introduction...	171
5.2	Ratio Estimation under Simple Random Sampling...	175
5.3	Variance of the Ratio Estimator...	183
5.4	Sample Estimates of Variance...	188
5.5	Confidence Interval...	190
5.6	Measuring the Efficiency of Ratio Estimators... .	193
5.7	Bias in Ratio Estimation...	198
5.8	Ratio Estimation in Stratified Sampling...	204
5.9	Ratio Estimation under Modified Random Sampling...	210
5.10	Unbiased Ratio Type Estimator...	216
5.11	Product Estimator...	217
5.12	Multivariate Ratio Estimator...	219
5.13	Comparing the Difference of Two Estimators... .	221
6	**REGRESSION ESTIMATES**	**238**
6.1	Linear Regression Estimator...	238
6.2	Properties of Regression Estimators...	241
6.3	Sampling Variance of Regression Estimator ...	244
6.4	Comparing Regression Estimator with Ratio and Mean Per Unit Estimators	248
6.5	Bias of the Regression Estimator...	251
6.6	Regression Estimator in Stratified Sampling...	252
6.7	Limitations of Separate and Combined Estimators	261
6.8	Difference Estimator...	261

6.9 Choice of Estimators... 264

7 SYSTEMATIC SAMPLING 274

7.1 Introduction 274
7.2 How to Draw a Systematic Sample... 275
7.3 Systematic Sampling in Relation to Other Designs 282
7.4 Advantages and Disadvantages of Systematic Sampling... 283
7.5 Applications of Systematic Sampling.... 284
7.6 Sample Mean and Its Variance... 284
7.7 Implications of ρ_w on $V(y_{sy})$... 289
7.8 Comparison with Stratified Sampling 290
7.9 Relative Precision of Systematic Sample Mean... 292
7.10 Estimating Population Characteristics... 295
7.11 Estimated Variance 300
7.12 Systematic Sampling in Special populations... . 305
7.13 Repeated Systematic Sampling... 312
7.14 PPS Approach in Systematic Sampling 315

8 SINGLE-STAGE CLUSTER SAMPLING 319

8.1 Introduction... 319
8.2 Some Important Features of Cluster Sampling... 321
8.3 Reasons for Using Cluster Sampling... 322
8.4 Disadvantages of Cluster Sampling... 322
8.5 Cluster Sampling in Relation to Other Designs ... 323
8.6 Applications of Cluster Sampling... 324
8.7 Clustering Principles... 324
8.8 Simple One-stage Cluster Sampling... 326
8.9 *Variance in Terms of Intra-class Correlation*... .. 334
8.10 Design Effect and Intra-class Correlation... ... 337
8.11 Relationship between ρ, Deff. and M... 337
8.12 Efficiency of Cluster Sampling... 338
8.13 Optimum Cluster Size... 345
8.14 Variable-Size Cluster Sampling... 349
8.15 Cluster Sampling with PPS... 357
8.16 Horvitz-Thomson Estimator... 364
8.17 Estimation of Proportions in Cluster Sampling... 377
8.18 Estimating Sample Size for Cluster Sampling... 384
8.19 Relative Efficiency of Unequal Cluster... 387
8.20 Use of Stratification in Cluster Sampling... 388
8.21 Systematic Sampling Approach in Cluster Sampling 390

9	**TWO–STAGE SAMPLING**	**394**
9.1	Introduction...	394
9.2	Drawing Two-stage Cluster Sampling...	392
9.3	Advantages of Two-stage Sampling...	396
9.4	Estimation of Population Mean and Total	398
9.5	A Simple Interpretation of Two-stage Sample...	402
9.6	Estimating a Population Proportion...	406
9.7	Estimates for Equal-Sized Clusters...	412
9.8	Two-stage Sampling with PPS...	420
10	**FURTHER ASPECTS OF SAMPLING**	**428**
10.1	Introduction...	428
10.2	Non-probability Sampling...	428
10.3	Multi-stage Sampling...	434
10.4	Multi-phase Sampling...	436
10.5	Post-stratification...	441
10.6	Area Sampling...	444
10.7	Acceptance Sampling...	445
10.8	Pre-testing...	445
10.9	Pilot Survey....	444
10.10	Design Weight...	447
10.11	Errors in Sample Survey...	451
10.12	Estimation of Variance in Complex Designs	461
10.13	Questionnaire and its Construction...	466
10.14	Steps Involved in Designing a Questionnaire...	468
10.15	Important Surveys in Bangladesh: An Overview	470
10.16	Demographic Surveys...	471
10.17	Contraceptive Prevalence Surveys (CPS).... ...	474
10.18	Bangladesh Demographic and Health Survey...	478
1019	Bangladesh Health and Demographic Survey ...	482
10.20	Health and Nutrition Survey...	482
10.21	Bangladesh Child Nutrition Survey...	485
10.22	Bangladesh Maternal Mortality Surveys	485
10.23	Labour Force Survey...	487
10.23	Household Expenditure Survey...	488
10.24	Bootstrapping	491
	RANDOM NUMBER	498
	BIBLIOGRAPHY	499
	SUBJECT INDEX	501

Chapter 1

INTRODUCTION

1.1 CONCEPT OF SAMPLING

During the past 30 years or so, the methods and techniques of sampling have reached a high level of scientific development. As a result, the uses of sampling have been extended into a wide variety of fields. From the standpoint of statistical data collection, sampling is a means for selecting a relatively small number of households, persons or other units for inclusion in a survey of some kind and inferring conclusions on the basis of these limited number of instances. This selection is done because enumeration of all units in the **target population** (population for which information is needed) is a large and complex undertaking that is almost always affected by limitations of time, budget and availability of experienced personnel. Not only that, it is unnecessary as well from the standpoint of precision and statistical reliability. Many countries have found, moreover, that sampling can play an important role in an overall census program (UN, 1997).

Let us now introduce the concept of sampling by an example.

EXAMPLE 1.1: Very frequently we talk about banning or restricting students' politics in the university campus. This is a very sensitive issue. We sometime wonder whether our views on this issue are shared by the student community, who are directly or indirectly involved in this important issue. We may want to know the actual percentage of students of Dhaka University who do not approve of banning students' politics in the campus. This percentage could be obtained by asking every student in the campus

inadequate. The works of Gosset appeared in Biometrika in 1908 under the name of 'Student', Gosset's pseudonym. (A famous story has it that Gosset was afraid to publish under his own name for fear that his employers, the Guinness Brewery, would be unhappy to discover that one of its chemists was doing research in statistics.) Today, student's t is a basic tool of statisticians and experimenters.

R. A. Fisher (1890–1962), who is known as the father of statistics, made numerous and significant contributions to statistics. He was influenced by Karl Pearson and Gosset, the two world known statisticians. His pioneering works made valuable contributions to evolve statistical procedures in many fields particularly agriculture, biology and genetics. He is known for his contribution in the field of Analysis of Variance and Experimental Design. Meanwhile, Francis Galton (1822–1921) gave the concept of **regression line** while working on heredity of men and laws governing the transmission of physical and mental characteristics from one generation to another'. His work was confirmed by his friend Karl Pearson.

The development of the theory of statistics was also effected by the pioneering works of a great number of authors, among whom Pareto, Adam, Edgeworth, Bowley, Yule, Stuart and Kendall deserve special mention.

J. Neyman (1894–1981) and E. S. Pearson (1895–1980) will remain ever known for their original contributions to the theory of testing statistical hypothesis in 1936 and 1938. This theory promoted considerable research works of practical use. Indian statistician Mahalanobis (1893–1972) contributed significantly to the field of sample survey. He is well known for his D^2 statistic. He established Indian Statistical Institute (ISI) in 1931 and founded **Sankhya**, an international journal in Statistics. Qazi Motahar Hussain (1897–1981), an authority in statistics, made significant contributions towards the development of statistics in Bangladesh. He is known worldwide for his works **'Hussain Chain'**. He was the founder of the Department of Statistics and the Institute of Statistical Research and Training at the University of Dhaka.

Deming's (1900–1993) philosophical thoughts towards improving the quality of manufactured products through statistical quality control techniques is noteworthy. Genichi Tagguci (1924–) promoted the use of experimental designs for product improvement.

CHAPTER

1

STATISTICS AND ITS ORIGIN

1.1 STATISTICS: ITS ORIGIN AND DEVELOPMENT

It is not precisely known how the word **statistics** was originated. However, most people believe that the term statistics, derived from word **state,** was used to refer to a collection of facts of interest to the state. Some believe that the word **statistic** has been originated from the Italian word **statista**, the French word **statistique** and the German word **statistik**. This background tends to suggest that the term statistics has its origin from the ancient time. At that time, the word statistics had been believed to be in use as an indicator or yardstick of a country's economic, political and social conditions. In this sense, the word statistics serves as an index of a country's overall condition. In modern days too, relevant statistical data are of immense importance to understand the level of development of a country. Presumably, all cultures that recorded history internationally also recorded statistics.

The term **statistics** is an old one. As people abandoned their nomadic way of life, and started to live in groups, they urgently felt the need to know each other's wealth, manpower, strength and position for their survival and safety. The group chiefs or rulers used to collect information on the above aspects in order to impose and levy tax to strengthen their economic condition. Caesar Augustus decreed that the entire world should be taxed.