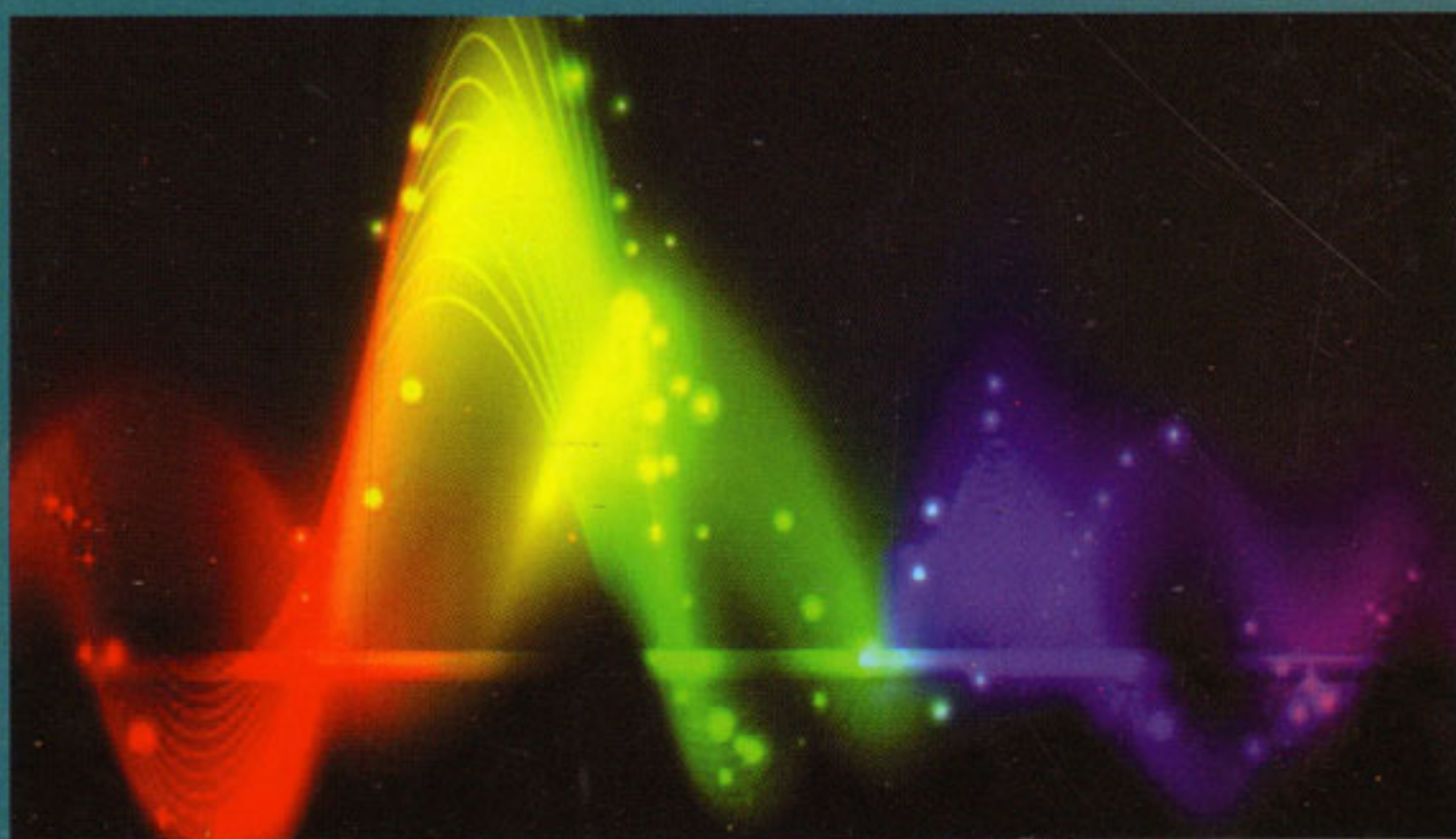


**FOURTH REVISED  
EDITION**

# **An Introduction to STATISTICS AND PROBABILITY**

**M. Nurul Islam**



**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 position 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

<b>1</b>	<b>STATISTICS AND ITS ORIGIN</b>	<b>1–20</b>
1.1	Statistics: Its origin and Development	1
1.2	General Field of Statistics	4
1.3	Defining Statistics	5
1.4	Characteristics of Statistics	10
1.5	Uses and Importance of Statistics	12
1.6	Population and Sample	15
1.7	Sources of Statistical Data	16
1.8	Distrust in Statistics	18
1.9	Computer and Statistics	18
<b>2</b>	<b>SUMMARIZING DATA</b>	<b>21–89</b>
2.1	Introduction	21
2.2	Meaning of Data	21
2.3	Level of Measurement	22
2.4	Variable and Attribute	26
2.5	Summarizing and Presenting Data	29
2.6	Frequency Distribution	31
2.7	Frequency Distribution for Continuous Data	33
2.8	Steps in Constructing Grouped Distribution	36
2.9	Formation of Discrete Frequency Distribution	39
2.10	Formation of Continuous Frequency Distribution	41
2.11	Other Forms of Frequency Distribution	47
2.12	Cumulative Frequency Distribution	48
2.13	Desirable Features of a Frequency Table	53
2.14	Presenting Data by Graphs and Diagrams	53
2.15	Presentation of Qualitative Data	54
2.16	Presentation of Quantitative Data	64
2.17	Use of Computers in Summarizing Data	83
<b>3</b>	<b>DESCRIPTIVE STATISTICS I: CENTRAL TENDENCY</b>	<b>90–167</b>
3.1	Introduction	90
3.2	Measures of Central Tendency	91
3.3	The Arithmetic Mean	92
3.4	The Median	112
3.5	Quartiles, Percentiles and Deciles	122
3.6	The Mode	129
3.7	The Geometric Mean	140



3.8	The Harmonic Mean	146
3.9	Other Measures of Average	150
3.10	Comparing the Averages	152

## **4 DESCRIPTIVE STATISTICS II: DISPERSION 168–252**

4.1	Meaning of Dispersion	168
4.2	Measures of Dispersion	170
4.3	Absolute Measures of Dispersion	171
4.4	Relative Measures of Dispersion	195
4.5	Empirical Relations among Measures of Dispersion	198
4.6	Comparing the Measures of Dispersions	200
4.7	A Few More Theorems on Dispersion	203
4.8	The Moments	213
4.9	Central Mom in Terms of Raw Moments	216
4.10	Effects of Changes in Origin and Scale on Moments	221
4.11	Sheppard's Correction for Moments	228
4.12	Shape Characteristics of a Distribution	232
4.13	Box and Whisker Plots	240
4.14	More Theorems and Examples	246

## **5 SIMPLE REGRESSION AND CORRELATION 253–339**

5.1	Regression Analysis: An Introduction	253
5.2	Simple Linear Regression Model	257
5.3	Types of Regression Analysis	260
5.4	Scatter Diagram	262
5.5	The Least Squares Method	263
5.6	Properties of Regression Coefficient	272
5.7	Partitioning the Total variation in Regression	273
5.8	Goodness of Fit in Regression	274
5.9	Correlation Analysis	282
5.10	Measuring the Correlation	283
5.11	Rank Correlation	290
5.12	Some Useful Results	303
5.14	More Problems on Correlation and Regression	314
5.15	Pearson's $r$ from Bivariate frequency Table	318
5.16	Correlation Ratio	322
5.17	Further Aspects of Correlation	326

## **6 MULTIPLE REGRESSION ANALYSIS 340–374**

6.1	Multiple Regression Model	340
-----	---------------------------	-----



6.2	Estimating the Parameters in the Model	343
6.3	Some Properties of the Estimators	345
6.4	Partitioning the Total Sum of Squares	348
6.5	Problems in Interpreting the Constants	353
6.6	Multiple Correlation	355
6.7	Partial Correlation	359
6.8	Polynomial Regression	368

## **7 PROBABILITY: A MEASURE OF UNCERTAINTY 375–473**

7.1	Understanding Probability	375
7.2	Probability: Historical Perspectives	376
7.3	A Review of Set and Set Notations	377
7.4	Operations with Sets	381
7.5	Law of Sets	386
7.6	Random Phenomenon and Related Concepts	388
7.7	Counting Rules: Permutation and Combination	396
7.8	Assigning Probabilities to Experimental outcomes	400
7.9	The Odds	409
7.10	Joint Probability	411
7.11	Conditional Probability	414
7.12	Independence of Events	420
7.13	Selected Theorems on Probability	424
7.14	Bayes' Theorem	437
7.15	Miscellaneous Examples	445

## **8 RANDOM VARIABLES AND ITS PROBABILITY DISTRIBUTION 474–532**

8.1	Random Variable	474
8.2	Probability Distribution	477
8.3	Discrete Probability Distribution	478
8.4	Discrete Distribution Function	486
8.5	Continuous Probability Distribution	490
8.6	Continuous Distribution Function	493
8.7	Joint Probability Distribution	499
8.8	Marginal Distribution	503
8.9	Conditional Distribution	506
8.10	Independence of Random Variable	510
8.11	A Few More Examples	514

## **9 MATHEMATICAL EXPECTATION 533–595**

9.1	Meaning of Mathematical Expectation	533
-----	-------------------------------------	-----



9.2	Expected Value of a Function of Two Random Variables	545
9.3	Conditional Expectation	556
9.4	Conditional Variance	558
9.5	Moment Generating Function	561
9.6	Cumulants and Cumulant Generating Function	569
9.7	Relations Between Moments and Cumulants	570
9.8	Characteristic Function	572
9.9	Probability Generating Function	576
9.10	Chebychev's Theorem	580
9.11	Special Mathematical Expectation	582

## **10 PROBABILITY DISTRIBUTIONS** **596–721**

10.1	Introduction	596
10.2	Bernoulli Distribution	595
10.3	Binomial Distribution	599
10.4	Hypergeometric Distribution	624
10.5	Poisson Distribution	631
10.6	Negative Binomial Distribution	654
10.7	Geometric Distribution	657
10.8	Multinomial Distribution	663
10.9	Discrete Uniform Distribution	666
10.10	Continuous Uniform Distribution	668
10.11	Normal Distribution	672
10.12	Exponential Distribution	704
10.13	Gamma Distribution	708
10.14	Beta Distribution	711

## **11 HYPOTHESIS TESTING** **722–833**

11.1	Hypothesis Testing: Concepts	722
11.2	Rationale of Hypothesis Testing	724
11.3	Statistical Significance	725
11.4	Types of Hypothesis	726
11.5	Level of Significance	731
11.6	Type I and Type II Errors	732
11.7	Critical Region	737
11.8	Test Statistic	740
11.9	Testing Procedure	740
11.10	The $p$ -Value	741



Consequently, he required that all persons report to the nearest tax collector. In the ancient Egypt also, such attempt was made to count people and land areas to redistribute the lands among the inhabitants of Egypt. This probably happened at the time of **Ramses II** in about 1400 BC. Further development of statistical records was noted in ancient Chinese, Greek and Roman territories. This feature has tempted many people to label statistics as the **science of king**.

The science of statistics developed gradually and its field of applications widened day by day. Its use and importance continued to be considered indispensable in all spheres of life as time elapsed. At government level, the statistics collected are of the same type as was done at the ancient time and serve almost the same purpose as before viz. the administrative, social, political and military policies of the government. The modern science of statistics really began to develop toward the end of the nineteenth century. It has to day grown into a vast and versatile subject in terms of its uses and applications to all sciences.

In about mid-seventeenth century, the theoretical development in modern statistics came with the introduction of the Theory of Probability. Gambling, in the form of games of chance, led to this theory of probability being originated by Pascal and Fermat because of their interest in the gambling experiences of the Chevalier de Mere.

The role of normal curve of error in the development of statistics can hardly be over emphasized. De Moivre first studied the equation of this curve in 1773. De Moivre's works remained unknown until 1924 when Karl Pearson (1857–1936) found it in a library. However, the same result was later developed independently by two famous mathematical astronomers Laplace (1749–1827) and Gauss (1777–1855).

The works of Charles Darwin (1809–1882), a famous biologist, were largely biomedical or statistical in nature. Mendel, too, extensively used statistical concepts in his study of plant hybrids.

The study of statistics gained impetus when Karl Pearson spent nearly half a century in statistical research. In addition, Karl Pearson founded the **Biometrika**, a famous journal in Statistics. He also established a school of statistics.

W. S. Gosset (1876–1937), a student of Karl Pearson, succeeded in working with samples, when large-sample theory was proving somewhat



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.