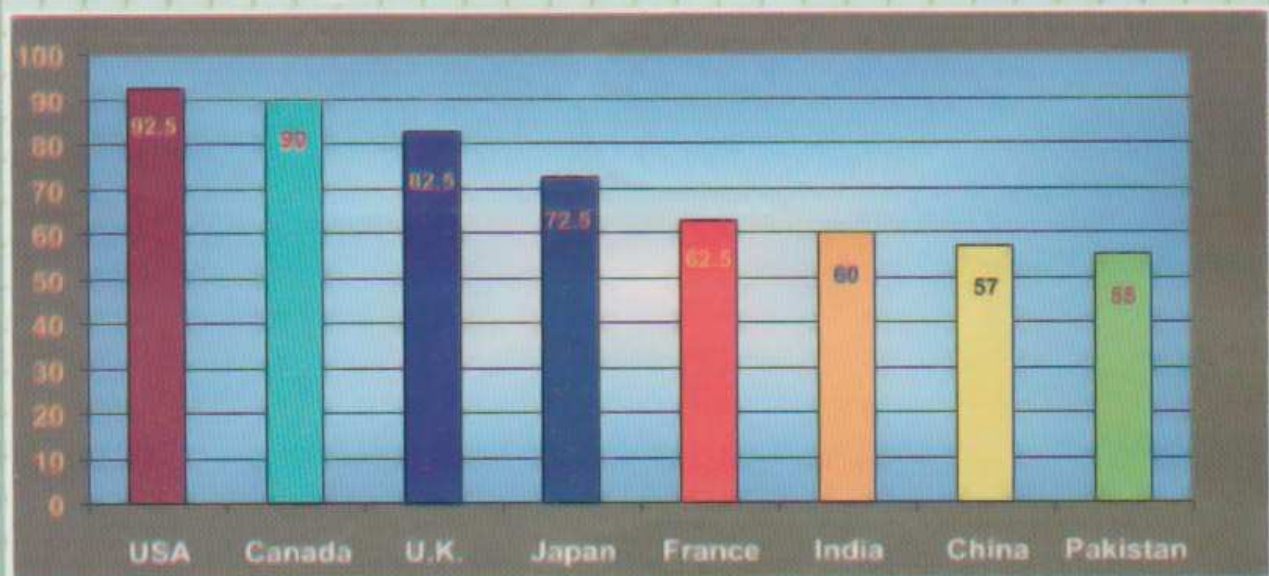


INTRODUCTION TO STATISTICAL THEORY

Part-1

Prof. Sher Muhammad Chaudhry
Prof. Dr. Shahid Kamal

Percentages of Adult Population using Wireless Phones
in Major Cities of 8 Countries in June 2008



ILMI KITAB KHANA

Preface to Sixth Edition

The sixth edition of An Introduction to Statistical Theory, revised jointly, has the same primary objectives as its earlier editions. However, it contains a considerable amount of new material and has a new format.

The great progress of the subject in previous decades has made it necessary to rewrite a number of sections and subsections in most of the chapters, to add new sections in a few places and to rearrange the subject matter as well as the order of some chapters to have a more logical sequence of topics. Numerous new examples and exercises drawn from past examination papers of various universities have been added.

We would like to express our appreciation to many professors, students and other readers who used the previous editions of this book. The changes, additions, deletions and corrections incorporated in the sixth edition, are mainly based on their invaluable comments and suggestions.

Students preparing for different examinations may omit the sections/chapters not meant for them.

Thanks are due to Messrs. Ilmi Kitab Khana, the publishers for their assistance in the production of this edition.

Suggestions for further improvement of the books will be welcome.

Lahore
September 1996

Sher Muhammad Chaudhry
Shahid Kamal

Preface to Eighth Edition

In response to a number of suggestions received both from teachers and students about the improvement of the book, some changes are made in the various chapters; some completely solved examples are either added or replaced to further clarify the techniques; at the end of each chapter Objective questions are given. In addition, all figures – diagrams and graphs – are redrawn. An effort is also made to remove the misprints.

It is confidently hoped that the book in the present form, will be found more useful.

Lahore
May 2009

Sher Muhammad Chaudhry
Shahid Kamal

CONTENTS

	<u>Page</u>
Preface	
1. INTRODUCTION	
1.1 Meaning of Statistics	1
1.1.1 Use of Statistical Information	1
1.1.2 Characteristics of Statistics	2
1.1.3 Descriptive and Inferential Statistics	3
1.1.4 Populations and Samples	4
1.1.5 Importance of Statistics	4
1.2 Observations and Variables	5
1.2.1 Variables	5
1.2.2 Discrete and Continuous Variables	5
1.2.3 Measurement Scales	6
1.2.4 Errors of Measurement	7
1.2.5 Significant Digits	8
1.2.6 Rounding off a Number	8
1.3 Collection of Data	9
1.3.1 Collection of Primary Data	9
1.3.2 Collection of Secondary Data	10
1.3.3 Editing of Data	11
1.3.4 Uses and Misuses of Statistics	11
Exercises	11
2. PRESENTATION OF DATA	
2.1 Introduction	15
2.2 Classification	15
2.2.1 Aims of Classification	15
2.2.2 Basic Principles of Classification	15
2.3 Tabulation	15
2.3.1 Types of Tables	16
2.3.2 Main Parts of a Table and its Construction	16
2.4 Frequency Distribution	20
2.4.1 Class-limits	20
2.4.2 Class-boundaries	20
2.4.3 Class-Mark	20
2.4.4 Class Width or Interval	20
2.4.5 Constructing a Grouped Frequency Distribution	21
2.4.6 Cumulative Frequency Distribution	26
2.5 Stem-and-Leaf Display	27
2.6 Graphical Representation	28
2.7 Diagrams	29
2.7.1 Simple Bar Chart	30
2.7.2 Multiple Bar Chart	30
2.7.3 Component Bar Chart	31

2.7.4	Rectangles and Sub-divided Rectangles	32
2.7.5	Pictograms	33
2.7.6	Pie Diagrams	34
2.7.7	Profit and Loss Chart	35
2.8	Graphs	35
2.8.1	Graph of Time Series – Historigram	36
2.8.2	Histogram	37
2.8.3	Frequency Polygon	39
2.8.4	Frequency Curve	40
2.8.5	Cumulative Frequency Polygon or Ogive	40
2.8.6	Ogive for a Discrete Variable	41
2.8.7	Types of Frequency Curves	41
2.8.8	Ratio Charts or Semi-logarithmic Graphs	43
	Exercises	44
3.	MEASURES OF CENTRAL TENDENCY OR AVERAGES	
3.1	Introduction	55
3.2	Criteria of a Satisfactory Average	55
3.3	Types of Averages	55
3.4	The Arithmetic Mean	56
3.4.1	The Weighted Arithmetic Mean	57
3.4.2	Properties of the Arithmetic Mean	57
3.4.3	Mean from Grouped Data	59
3.4.4	Change of Origin and Scale	61
3.5	The Geometric Mean	62
3.6	The Harmonic Mean	64
3.7	The Median	67
3.7.1	Quantiles	68
3.8	The Mode	72
3.9	Empirical Relation between Mean, Median and Mode	74
3.10	The Box Plots	74
3.11	Relative Merits and Demerits of Various Averages	75
3.11.1	The Arithmetic Mean	75
3.11.2	The Geometric Mean	75
3.11.3	The Harmonic Mean	76
3.11.4	The Median	76
3.11.5	The Mode	76
	Exercises	77
4.	MEASURES OF DISPERSION, MOMENTS AND SKEWNESS	
4.1	Introduction	87
4.2	Range	87
4.3	The Semi-Interquartile Range or the Quartile Deviation	88
4.4	The Mean (or Average) Deviation	89

4.5	The Variance and Standard Deviation	91
4.5.1	Change of Origin and Scale	96
4.5.2	Interpretation of the Standard Deviation	97
4.5.3	Co-efficient of Variation	98
4.5.4	Properties of Variance and Standard Deviation	100
4.5.5	Standardized Variables	103
4.6	Trimmed and Winsorized Measures	104
4.7	Moments	105
4.7.1	Moments about the Mean in terms of Moments about an arbitrary origin, say a , and conversely	106
4.7.2	Sheppard's Corrections	108
4.7.3	Moment-Ratios	108
4.7.4	Change of Origin and Scale	110
4.7.5	Charlier Check	111
4.8	Skewness	114
4.9	Kurtosis	115
4.10	Describing a Frequency Distribution Exercises	116

5. INDEX NUMBERS

5.1	Introduction	131
5.1.1	Simple and Composite Index Numbers	131
5.1.2	Problems Involved in Index Number Construction	131
5.2	Main Steps in the Construction of Index Numbers of Wholesale Prices	132
5.2.1	Selection of Commodities for Inclusion	132
5.2.2	Selection of the Base Period	132
5.2.3	Selection of Average	134
5.2.4	Selection of Appropriate Weights	135
5.3	Unweighted Index Numbers	135
5.3.1	Simple Aggregative Index	135
5.3.2	Simple Average of Relatives	136
5.4	Weighted Index Numbers	140
5.4.1	Weighted Aggregative Price Index Numbers	140
5.4.2	Weighted Average of Relatives Price Index Number	145
5.5	Quantity Index Numbers	147
5.6	Tests for Index Number Formulae	151
5.6.1	Time Reversal Test	151
5.6.2	Factor Reversal Test	153
5.6.3	Circular Test	155
5.7	Consumer Price Index Number	156
5.7.1	Meaning	156
5.7.2	Construction of Consumer Price Index Numbers	156

5.7.3	Shortcomings or Drawbacks of Consumer Price Index Numbers	159
5.8	Uses of Index Numbers	160
5.9	Limitations of Index Numbers	160
	Exercises	161
6.	PROBABILITY	
6.1	Introduction	173
6.2	An Aside – Sets	173
6.2.1	Subsets	174
6.2.2	Venn Diagram	175
6.2.3	Operations on Sets	175
6.2.4	The Algebra of Sets	177
6.2.5	Partition of Sets	177
6.2.6	Class of Sets	177
6.2.7	Cartesian Product Sets	177
6.2.8	Relation and Function	178
6.3	Random Experiment	179
6.3.1	Sample Space	179
6.3.2	Events	180
6.3.3	Events and Symbolic Representations	181
6.3.4	Counting Sample Points	181
6.4	Definitions of Probability	184
6.4.1	Subjective or Personalistic Probability	187
6.5	Laws of Probability	189
6.6	Conditional Probability	196
6.7	Independent and Dependent Events	205
	Exercises	214
7.	RANDOM VARIABLES	
7.1	Introduction	227
7.2	Distribution Function	227
7.3	Discrete Random Variables and its Probability Distribution	228
7.4	Continuous Random Variable and its Probability Density Function	233
7.5	Joint Distributions	237
7.5.1	Bivariate Distribution Function	237
7.5.2	Bivariate Probability Functions	238
7.5.3	Marginal Probability Functions	238
7.5.4	Conditional Probability Functions	239
7.5.5	Independence	239
7.5.6	Continuous Bivariate Distributions	243
7.6	Mathematical Expectation of a Random Variable	248
7.6.1	Expectation of a Function of a Random Variable	250

7.6.2	Properties of Expected Values	256
7.6.3	Covariance of Two Random Variables	263
7.6.4	Variance of the Sum or Difference of Two Random Variables	263
7.6.5	Correlation Co-efficient of Random Variables	263
7.7	Medians and Modes of Continuous Random Variables	266
7.8	Chebyshev's Inequality	268
7.9	Moment Generating Function	269
7.9.1	Cumulant Generating Function	271
7.9.2	Relation between Cumulants and Moments	272
7.9.3	Characteristic Function	273
	Exercises	273
8.	DISCRETE PROBABILITY DISTRIBUTIONS	
8.1	Introduction	285
8.2	Binomial Probability Distribution	285
8.2.1	Derivation of Binomial Probability Distribution	285
8.2.2	Binomial Frequency Distribution	291
8.2.3	Properties of the Binomial Probability Distribution	292
8.2.4	The Recurrence Formula for Binomial Distribution	297
8.2.5	Fitting a Binomial Distribution to Observed Data	298
8.2.6	Moment Generating and Cumulant Generating Functions of the Binomial Distribution	299
8.3	Hypergeometric Probability Distribution	302
8.3.1	Derivation of Hypergeometric Distribution	302
8.3.2	Properties of Hypergeometric Distribution	305
8.4	Poisson Distribution	308
8.4.1	Derivation of Poisson Approximation to the Binomial	309
8.4.2	Poisson Frequency Distribution	312
8.4.3	Properties of the Poisson Distribution	313
8.4.4	The Recurrence Formula for Poisson Distribution	316
8.4.5	Fitting a Poisson Distribution to Observed Data	317
8.4.6	Poisson Process	318
8.4.7	Moment Generating and Cumulant Generating Functions of the Poisson Distribution	320
8.5	Negative Binomial Distribution	321
8.5.1	Derivation of the Negative Binomial Distribution	322
8.5.2	Properties of the Negative Binomial Distribution	324
8.6	Geometric Distribution	325
8.6.1	Derivation of the Geometric Distribution	325
8.6.2	Properties of the Geometric Distribution	326
8.6.3	Moment Generating Function of the Geometric Distribution	327
8.7	Multinomial Distribution	328
8.7.1	Derivation of the Multinomial Distribution	328

9. CONTINUOUS PROBABILITY DISTRIBUTIONS

9.1	Introduction	341
9.2	Uniform Distribution	341
9.2.1	Properties of the uniform distribution	341
9.2.2	Moment-Generating Function of Uniform Distribution	342
9.3	Exponential Distribution	342
9.3.1	Properties	343
9.3.2	Moment-Generating Function of Exponential Distribution	344
9.4	Gamma and Beta Distributions	344
9.4.1	Gamma Function	344
9.4.2	Beta Function	346
9.4.3	Gamma Distribution	347
9.4.4	Properties of Gamma Distribution	347
9.4.5	Moment-Generating Function of Gamma Distribution	348
9.4.6	Beta Distribution of the First Kind	349
9.4.7	Properties of $\beta_1(m, n)$	349
9.4.8	Beta Distribution of Second Kind	350
9.4.9	Properties of $\beta_2(m, n)$	351
9.5	Normal Distribution	351
9.5.1	Standardized Normal Distribution	353
9.5.2	Properties of Normal Distribution	355
9.5.3	Moment Generating and Cumulant Generating Functions of the Normal Distribution	362
9.5.4	Tabulated Areas of Normal Distribution	364
9.5.5	Inverse Use of Table of Areas under Normal Curve	370
9.5.6	Normal Approximation to the Binomial Distribution	374
9.5.7	Normal approximation to the Poisson Distribution	378
9.5.8	Fitting a Normal Distribution	379
9.5.9	Ordinates of the Normal Distribution	382
	Exercises	384

10. SIMPLE REGRESSION AND CORRELATION

10.1	Introduction	395
10.2	Deterministic and Probabilistic Relations or Models	395
10.3	Scatter Diagram	396
10.4	Simple Linear Regression Model	396
10.4.1	An Aside – The Principle of Least Squares	398
10.4.2	Least-Squares Estimates in Simple Linear Regression	398
10.4.3	Properties of the Least-Squares Regression Line	402
10.4.4	Standard Deviation of Regression or Standard Error of Estimate	402
10.4.5	Co-efficient of Determination	404

10.5	Correlation	406
10.5.1	Pearson Product Moment Correlation Co-efficient	406
10.5.2	Correlation and Causation	408
10.5.3	Properties of r	408
10.5.4	Correlation Co-efficient for Grouped Data	411
10.6	Rank Correlation	413
10.6.1	Derivation of Rank Correlation	414
10.6.2	Rank Correlation for Tied Ranks	416
10.6.3	Co-efficient of Concordance	417
	Exercises	419
11.	MULTIPLE REGRESSION AND CORRELATION	
11.1	Introduction	429
11.2	Multiple Linear Regression with two Regressors	429
11.2.1	Expression of Multiple Linear Regression in Deviation Form	431
11.2.2	Standard Error of Estimate	433
11.2.3	Co-efficient of Multiple Determination and Multiple Correlation	434
11.2.4	Subscript Notation	435
11.2.5	Properties of Residuals	436
11.2.6	Multiple Regression in terms of Linear Correlation Coefficients	437
11.3	Multiple Correlation Co-efficient	438
11.4	Partial Correlation	441
11.4.1	Relationship between Multiple and Partial Correlation Co-efficients	446
11.5	Curvilinear Regression	447
	Exercises	447
12.	CURVE FITTING BY LEAST SQUARES	
12.1	Introduction	455
12.2	Approximating Curves and the Principle of Least-Squares	455
12.2.1	Fitting a Straight Line	455
12.2.2	Fitting of a second degree Parabola	457
12.2.3	Fitting of Higher Degree Parabolic Curves	460
12.2.4	Change of Origin and Unit	461
12.3	Exponential Curves	462
12.4	Other Types of Curves	465
12.4.1	Modified Exponential Curve	465
12.4.2	The Gompertz Curve	467
12.4.3	The Logistic Curve	467
12.4.4	The Makeham Curve	467
12.5	Criteria for a Suitable Curve	468
12.6	Finding Plausible Values by LS Method	468

Exercises	470
13. TIME SERIES ANALYSIS	
13.1 Introduction	477
13.2 Components of a Time Series	478
13.2.1 Secular Trend	478
13.2.2 Seasonal Variations	479
13.2.3 Cyclical Fluctuations	479
13.2.4 Irregular or Random Variations	480
13.3 Time Series Decomposition	480
13.4 Analysing the Secular Trend	481
13.4.1 The Method of Freehand Curve	481
13.4.2 The Method of Semi-Averages	481
13.4.3 The Method of Moving Averages	483
13.4.4 The Method of Least-Squares	487
13.5 Detrending	491
13.6 Analysing the Seasonal Variations	492
13.6.1 The Percentage of Annual Average method	492
13.6.2 The Ratio-to-Moving-Average Method	493
13.6.3 The Ratio-to-Trend Method	495
13.6.4 The Link-Relative Method	497
13.7 Deseasonalization of Data	498
13.8 Analysing the Cyclical Variations	499
13.9 Analysing the Irregular Variations	500
13.10 Forecasting	501
13.10.1 Forecasting by Exponential Smoothing	501
13.11 Serial Correlation	501
Exercises	503
Answers to Exercises	515
Tables	539
Index	545