

FUNDAMENTALS *of* STATISTICS

BY
TRUMAN LEE KELLEY
HARVARD UNIVERSITY



HARVARD UNIVERSITY PRESS
CAMBRIDGE, MASSACHUSETTS
1947

FUNDAMENTALS OF STATISTICS

Copyright, 1947
By Truman Lee Kelley

Printed in the United States of America

LONDON : GEOFFREY CUMBERLEGE

OXFORD UNIVERSITY PRESS

To
EXPERIMENTALISTS

Though the mists of life
obscure the distant scene
they are undismayed for,
in markings near about,
they discern the contour
of the land and the portal
to the future.

Kelley: FUNDAMENTALS OF STATISTICS

ERRATA

- Page 60, Problem 4: for second degree read third degree
- Page 133, Fourth line from the bottom: for $i=12$ read number of intervals =12
- Page 182, Fourth line: for 6.629 read 5.629
Third line from bottom: for .54.2 read 54.2
Last line: for 6.630 read 5.630
- Page 183, First line: for 6.629 read 5.629
- Page 201, Ninth line: for n^2-2 read n^2-n
Tenth line: for N^2-2 read N^2-N
Last line: for N_2-N read N^2-N
- Page 203, First line: for $-2(X_1)^2$ read $-2(\Sigma X_1)^2$
- Page 209, First line: for \tilde{V} read \tilde{V}
Fifth line: for $\tilde{V} = [N/(N-1)]\bar{V}$ read $\tilde{V} = [N/(N-1)]\bar{V}$
- Page 211, Formula [6:21]: for μ_m read μ_k
- Page 212, First column: for $V = \frac{N}{N-1} \tilde{V}$ read $\tilde{V} = \frac{N}{N-1} V$
- Page 222, Twelfth line from bottom: for 89.6666667 read 80.6666667
- Page 223, Formula [6:55]: for $(N-1)\tilde{U}_4 - (N-3)\tilde{V}^2$ read $[(N-1)\tilde{\mu}_4 - (N-3)\tilde{V}^2]$
- Page 246, Formula [7:08]: for $P_{.25} - P_{.25}$ read $P_{.75} - P_{.25}$
- Page 260, Fourth line: for σ_{w_0} read $\sigma_{f_{w_0}}$
- Page 283, Lines five and thirteen from bottom: for dromedary read camel
- Page 284, Twenty-fourth line: for VII read VIII
- Page 291, Formula [8:15]: for 735 read 945
- Page 292, Third line: for μ_0, μ_1, μ_1 read μ_0, μ_1', μ_1
- Page 293, Table VIII C column 2q: for .001 read .01
- Page 295, Chart VIII IV: for d read d_{ij}
- Page 307, Tenth line from bottom: for $\sigma_1\sigma_2\ldots\sigma_1$ read $\sigma_1\sigma_2\ldots\sigma_1$
- Page 308, Tenth line: for x_3 read x_1
Formula [8:37]: for $\frac{x_1^2}{V_2}$ read $\frac{x_1^2}{V_1}$
- Page 309, Formula [8:39]: for $F_{1\infty}$ read $F_{1\infty}$
- Page 323, Formula [9:15]: for $\frac{f_s}{N} \times \frac{f_s}{N}$ read $\frac{f_s}{N} \times \frac{f_t}{N}$
- Page 423, Fourth line: for z read z_w
- Page 513, Formula [13:29]: after $+a_{n+1}$ read =0
- Page 517, Formula [13:44]:
for $-\frac{\alpha_3}{2\sigma}$ read $-\frac{\alpha_3\sigma}{2}$ and after this read α of [13:44] is not α of [13:30]
- Page 520, Fourth line: for not read now
- Page 528, Seventh line from bottom: for [13:134] read [13:142]
- Page 575, Fifth line: for k th read j th
Middle of page: for pre-multiplied read post-multiplied and for post-multiplied read pre-multiplied
Fourth and fifth lines from bottom: for $(ABC)' = C'B'A'$ read $(ABC)' = C'B'A'$
- Page 583, Last line: for $(qn-2)$ read $(qn-2)/$
- Page 590, First line: for y read y_4
- Page 679, Last line: for $m!/m! (n-m)!!$ read $n!/m! (n-m)!!$.

PREFACE

An alternative title for this book which would emphasize the point of view is *Statistics, Its Philosophy and Method*. This work is more than a revision of the writer's earlier *Statistical Method* (1923). That work covered a field which in the last twenty years has grown to such magnitude that nothing short of an encyclopedia could cover it today. The endeavor herein has been to place a great emphasis upon the logic and principles underlying the statistical study of phenomena, to provide, in the early chapters, such basic issues as will integrate thoughtful and investigative moods with statistical processes, and, in the later chapters, to give such treatment of modern processes as is required in handling many experimental situations and as will open to the reader the wealth of thought in current statistical literature.

The early chapters constitute an elementary text, and call for no mathematical background other than that of arithmetic and elementary algebra. They seek, first, to show the place of statistics in the social process of solving problems and, second, to provide the most basic tools. The approach to statistics of most beginning students is that it holds some useful

tricks which, with a certain drudgery, can be learned. The content of statistics is not a bag of tricks, but rational thought processes which serve in problem situations. That there is a nicety in these processes gives zest to their discovery and pleasure in their use. This viewpoint can scarcely be overemphasized, and it is hoped that teachers of this text will further emphasize and elaborate upon the utility of statistical concepts in the everyday problems of the physical and biological scientist, the schoolman, the doctor, the farmer, the economist, and the businessman.

The parallel work of the author, *Statistical Tables*, is designed specifically for laboratory statisticians and is not a requisite to students of this text in view of the abridged set of tables given herein.

In addition to the numerous acknowledgments cited throughout this text, the writer is deeply indebted to Eric F. Gardner for statistical research in connection with sequential analysis and otherwise and he is also greatly indebted to him and to Katherine Ytredal for expert assistance in the difficult task of preparation of tables and of the manuscript for the printer.

Truman Lee Kelley

Cambridge, Massachusetts

CONTENTS

I. THE DIGNITY OF DATA AND THE BACKGROUND OF STATISTICS

1. Statistics, Psychology, and Logic	3
2. Historical Antecedents of Modern Statistics	10
3. Occasions for Resort to Statistics	12
4. The Mathematical Background of Students of Statistics	24
5. Plan of Study	56
Problems	58

II. STATISTICAL SERIES

1. Types of Data	62
2. Types of Issues	69
3. Types of Statistical Processes	74
4. The Service Rendered by Elementary and by Advanced Statistics	79
Problems	82

III. STATISTICAL TABLES

1. Characteristics of a Good Table in General	84
2. Special Characteristics of the Three Types of Tables	89
Problems	95

IV. GRAPHIC METHODS

1. General Fields in Which Serviceable . . .	98
2. Temporal Series	101
3. Quantitative Series	120
4. Qualitative Series	150
5. Geographic Series	153
6. Graphic Portrayal in Three Dimensions . .	166
Problems	173

V. THE STABLE FEATURES OF PHENOMENA

1. The Quest for Certainty	185
2. Biological and Social Stability	187
3. Temporal Data	193
4. Geographical Data	193
5. Quantitative and Qualitative Data	194
6. Complex Data	197
7. Employable Instruments in the Quest For Certainty	198

VI. MEASURES OF VARIABILITY

1. The General Concept of Variability . . .	199
2. Degrees of Freedom	205
3. The Variance and the Standard Deviation of the Mean	207
4. Sample and Population Moments	210
5. Cumulants and k -Statistics	215
6. The Computation of the Mean and of Moments	216
7. Decimal Places to be Kept in Published Results	222
8. Variance Errors and Standard Errors of Moments	223

VII. MEASURES OF CENTRAL TENDENCY

1. The Effect of Form of Distribution Upon Different Averages	234
2. The Median	240
3. The Mode	248
4. The Geometric Mean	265

5. The Harmonic Mean	271
Problems	274

VIII. THE NORMAL DISTRIBUTION

1. The Normal Distribution as Descriptive of Chance Distributions and Distributions of Errors	275
2. The Normal Distribution as Descriptive of Biological and Social Phenomena	277
3. The Normal Distribution as Related to Advanced Statistical Theory	284
4. The Normal Distribution as Related to the Design of Experiments	286
5. Derivation and Tables of the Normal Curve	287
6. Certain Properties of the Normal Distribution	290
7. Statistical Constants Descriptive of Portions of a Normal Distribution	295
8. The Significance of Differences Between the Means of Tail Portions of a Normal Distribution	300
9. Fitting a Normal Curve to Data	301
10. The Distribution of the Sum of Normal Variables	305
11. The Relationship of Chi-Square, T-Square, and F to Normal Distribution	306

IX. THE STATISTICS OF ATTRIBUTES

1. Situations Wherein Qualitative Series Arise	311
2. The Frequency in a Class	311
3. Chi-Square	318
4. χ^2 From the General Contingency Table	321
5. Transformations Normalizing the Variance Ratio Distribution	325

X. ESTIMATION, REGRESSION, AND CORRELATION

1. A Brief Perspective of the Field . . .	332
2. The Problem of Concomitant Variation in The Sciences	337
3. Findings Resulting from Galton's Graphic Treatment	339
4. Algebraic Statement of Galton's Graphic Findings and Derivation of Correlation Formulas	344
5. Derivation of Computational Formulas for b and r	348
6. The Regression Equation and the Analysis of Variance	354
7. The Trustworthiness of a Correlation Coefficient	358
8. Averaging Correlation Coefficients . .	363
9. The Trustworthiness of a Point on the Regression Line	363
10. Correlation Between Ranks	365
11. Biserial Correlation	370
12. Two-By-Twofold Correlation	379
13. Adjustments for Coarseness of Grouping	388
14. Correlations and Other Statistics of Sums and Differences	395

XI. FURTHER CORRELATION ISSUES

1. The Various Consequences of Employing Semi-Reliable Initial Measures . . .	399
Reliability of Measures and Issues Involving Two Variables	411
Reliability and Other Issues Involving Three or More Unequally Reliable Measures of the Same Thing	419
2. The Effect of Variability in Range Upon Correlation	425
3. Three-Variable Multiple Correlation . .	433
4. Nonlinear Regression	442
5. The Correlation Ratio	449

XII. THE GENERAL MULTIPLE LINEAR REGRESSION PROBLEM

1. A Statement of Relationships in Connection with Standard Score Variables	454
2. A Modified Doolittle Solution	458
3. The Determinantal Expression of the Solution of Normal Equations	471
4. The Use of Matrices in Expressing Multiple Correlation Relationships	475

XIII. SUNDRY STATISTICAL ISSUES AND PROCEDURES

1. Evidence of Periodicity in Short Time Series	482
2. Lead and Lag in Time Series	492
3. Imposed Conditions	497
4. Comparable Measures	499
5. Quotients	501
6. The Most Reliable Weighted Average of Independent Measures	505
7. Fitting of Curves to Observations	506
8. The Variance Error of a Statistic Derived from One Having a Known Variance Error	523
9. The Variance Error of a Coefficient Corrected for Attenuation	526
10. The Equi-Probable and the Mean Ranges for Samples of Different Size Drawn from a Normal Population	529
11. The Optimal Size of Interval for Histogram or Frequency Polygon	531
12. Direct and Inverse Interpolation	538
13. The Machine Extraction of Square and Cube Roots	543
14. Occasional Formulas	545
15. Sequential Analysis	555

XIV. MATHEMATICS, THE MENTOR OF STATISTICAL INGENUITY

1. Ingenuity in Research	571
2. Matrices and Determinants	573
Determinantal Solution of Simultaneous Equations	580
3. The Point Binomial	581
4. The Poisson Distribution	582
5. The Hypergeometric Series	583
6. Factorials and the Gamma Function	585
7. The Numerical Solution of Higher Degree Parabolic and of Exponential Equations	589
The Numerical Solution of Complicated Simultaneous Equations	591
8. Transforming Rank and Percentage Positions into Quantitative Scores	592
Normalizing a Distribution	592
Transforming Rank, or Percentage, Position into Equally Reliable Deviation Scores	593
9. The Square Root Transformation	598
10. The Cube Root Transformation	599
11. Certain Properties of Differences in Tabled Entries	599
12. Expanding A Table By Interpolating Values	601
13. Trigonometric Functions of Sums and Differences	603
14. Space of Two Dimensions	603
15. Space of Three or More Dimensions	605
16. LaGrange Multipliers	607
17. Growth Curves	608
18. The Binomial Theorem	609
19. Stirling's Approximation to the Factorial	609
20. The Sum of the Positive Powers of the First k Numbers	609
21. Fourier Series	610
22. Taylor's Theorem	610
23. The Euler-Maclaurin Formula for Evaluating a Definite Integral	610

XV. STATISTICAL TABLES

1. Selected References	612
1814. BARLOW'S TABLES	612
1910. Peirce. A SHORT TABLE OF INTEGRALS. . .	613
1914. Pearson, Ed. TABLES FOR STATISTICIANS AND BIOMETRICIANS, Part I.	613
1919-1939. CAMBRIDGE UNIVERSITY TRACTS FOR COM- PUTERS. Pearson, Ed.	614
1923. Glover. TABLES OF APPLIED MATHEMATICS IN FINANCE, INSURANCE, STATISTICS	615
1930. Salvosa. TABLES OF PEARSON TYPE III FUNC- TIONS.	615
1931. Pearson, Ed. TABLES FOR STATISTICIANS AND BIOMETRICIANS, Part II	616
1932. Dunlap and Kurtz. HANDBOOK OF STATISTICAL NOMOGRAPHS, TABLES, AND FORMULAS	616
1933. Chesire, Saffir, and Thurstone. COMPUTING DIAGRAMS FOR THE TETRACHORIC CORRELATION CO- EFFICIENT	617
1933-1935. Davis. TABLES OF THE HIGHER MATHEMA- TICAL FUNCTIONS.	617
1934-1938. BIOMETRIKA PUBLICATIONS. Pearson, Ed.	618
1934. Dwight. TABLES OF INTEGRALS AND OTHER MATHEMATICAL DATA	619
1935. Kelley. ESSENTIAL TRAITS OF MENTAL LIFE	619
1938. Fisher and Yates. STATISTICAL TABLES FOR BIOLOGICAL, AGRICULTURAL AND MEDICAL RESEARCH	620
1939-1944. NATIONAL BUREAU OF STANDARDS TABLES Lowan, Technical Director.	621
1939. Sheppard. THE PROBABILITY INTEGRAL. . .	623
1941. Comrie and Hartley. TABLE OF LAGRANGIAN COEFFICIENTS...	623
1941. Thompson. TABLE OF PERCENTAGE OF THE χ^2 DISTRIBUTION.	623
1941. Thompson. TABLES OF PERCENTAGE POINTS OF THE INCOMPLETE BETA-FUNCTION	623
1942. Molina. POISSON'S EXPONENTIAL BINOMIAL LIMIT.	624

1942. SMITHSONIAN MATHEMATICAL TABLES, HYPERBOLIC FUNCTIONS. Becker and Van Orstand. . .	624
1942-1943. Pearson and Hartley. THE PROBABILITY INTEGRAL OF THE RANGE IN SAMPLES OF N OBSERVATIONS FROM A NORMAL POPULATION.	625
1943. Merrington and Thompson. TABLES OF PERCENTAGE POINTS OF THE INVERTED BETA (F) DISTRIBUTION	625
1944. Bateman and Archibald. A GUIDE TO TABLES OF BESSEL FUNCTIONS	626
In press, 1947. THE KELLEY STATISTICAL TABLES	626
2. Lagrangian Interpolation Coefficients. .	627
3. Normal Probability Functions	639
4. Square and Cube Roots.	652
5. Shorter Mathematical Tables.	657

APPENDIX A. MATHEMATICAL BACKGROUND TEST

1. The Background Test.	659
2. Scoring Keys	672
3. Percentile Norms and Reliability	675

APPENDIX B. REFERENCE LISTS

1. List of Common Statistical Symbols . . .	677
Greek Alphabet.	677
Literal Symbols	678
2. Non-Literal Symbols.	693
3. Mathematical Terms	695
4. Key to Formulas.	698
5. References	729

APPENDIX C. CORRELATION CHART 742

Index	743
-----------------	-----

FUNDAMENTALS OF STATISTICS