ELMER B. MODE

Elements of STATISTICS

Third Edition

Elements of STATISTICS

third edition

Elmer B. Mode

Professor of Mathematics / Boston University

PRENTICE-HALL, INC. Englewood Cliffs, New Jersey Current printing (last digit):
15 14 13 12 11 10 9 8

© 1961 by PRENTICE-HALL, INC. Englewood Cliffs, N. J.

All rights reserved. No part of this book may be reproduced in any form, by mimeograph or any other means, without permission in writing from the publishers.

Library of Congress Catalog Card Number 61–10792

Printed in the United States of America
27254C

Elements of STATISTICS

"The time has come," the Walrus said,
"To talk of many things:
Of shoes—and ships—and sealing-wax—
Of cabbages—and kings—"

此为试读,需要完整PDF请访问: www.ertongbook.com

PREFACE

It has been said that the peculiar characteristic which sets man apart from the other animals is his capacity for capitalizing the achievements of his ancestors. Surely statistics offers a striking example of man's attempt to employ the records of the past in solving the problems of the present and in plotting the course of the future.

Much of the field of elementary statistics has been pretty well organized, so that the general route of progress for the beginner can be mapped out with little difficulty. Yet this route, plainly marked as it is, may be traversed with such facile mathematical skill—and nothing else—as to cause the novice to lose some of the most valuable lessons in simple critical analysis which statistics can teach. There is a peculiar logic associated with the study of statistics. Some call it common sense; others prefer a more exact characterization. In any case, it is difficult to dissociate the purely mathematical aspects from those which are not exactly mathematical but which do require a good brand of clear thinking. It is hoped that this book, dealing as it does with the elementary phases of statistics, will develop on the part of the reader an appropriate attitude of critical reasonableness toward the subject. There is no reason why sound mathematical conclusions should not be accompanied by a healthy appreciation of what such conclusions do not assert.

A large amount of important work can be done without extensive preparation in mathematics. Most persons who study elementary statistics are motivated by the fact that it has a direct bearing on a related field of interest:

business, economics, sociology, biology, psychology, education, and others. They want to acquire "sufficient statistical terminology and technique" to enable them to read intelligently and somewhat critically the statistical content of the literature in these fields. If called upon to perform a simple statistical analysis, they wish to be able to handle the fundamental procedures.

The present book favors no particular field of interest. The subjects enumerated above are well represented. There is an unusual abundance and a wide variety of exercises, none of which has been borrowed from other textbooks.

The arrangement of material allows much flexibility in the choice of topics. Many sections and some chapters may be eliminated, if so desired, without affecting the continuity of the book. Those who desire a brief course on methodology can omit such proofs as are present, although these cast valuable light on the concepts and processes involved. In particular, the writer has aimed to describe the full purport of each distinctive statistical concept. A list of references appropriate to an introductory study of this kind will be found at the end of the book.

The present edition, the third, presents a considerable revision of the earlier texts. In a field whose rate of extension and clarification is very rapid, fairly frequent changes in concepts and method of presentation are inevitable. The chief features of this revision are (1) a separate chapter on probability—the cornerstone upon which statistics rests; (2) the exposition of the normal distribution following that of the binomial; (3) a more systematic treatment of the testing of hypotheses; (4) some diminution in the computational aspects of statistics; (5) greater emphasis on the interpretation of results; (6) more up-to-date tables and charts of critical values; and (7) a brief optional chapter on nonparametrics.

The newer expositions have proved to be teachable at the level of the earlier editions of this book, and it has been the author's experience that the more challenging chapters excite the student's interest most of all.

The author is indebted to Sir Ronald A. Fisher and to Messrs. Oliver and Boyd, Ltd., Edinburgh, for permission to reprint Tables III, IV, and Va, from their book, *Statistical Methods for Research Workers*. The author is also grateful to Professor George W. Snedecor and the Iowa State College Press, to Professor Egon S. Pearson and the *Biometrika* Trustees, to Professors Wilfrid J. Dixon and Alexander M. Mood, to Drs. F. Swed and Churchill Eisenhart, and to Professor E. G. Olds, for the use of Tables appearing in the back of the book.

The author's thanks are due his kind colleague Professor Albert Morris, for helpful comments made after reading the manuscript of the first edition, and especially to Professor Albert A. Bennett, whose constructive criticisms

PREFACE

and invaluable suggestions added greatly to the interest of the first two editions. In particular the author is deeply in debt to Professor Elizabeth A. Shuhany for her many recommendations for the improvement of the presentation of the material in this, the third edition.

E.B.M.

CONTENTS

1

INTRODUCTION

page 1

1.1: The origin of statistics, 1.
1.2: The meaning of statistics, 2.
1.3: Sample and population, 3.
1.4: Random sampling numbers, 5.
1.5: The nature of statistics, 5.
1.6: Some nonmathematical aspects, 7.
1.7: Misuses of statistics, 8.
1.8: Exact and approximate numbers, 9.
1.9: Significant digits, 10.
1.10: The addition and subtraction of approximate numbers, 11.
1.11: The multiplication and division of approximate numbers, 12.
1.12: Remarks concerning approximate computation, 13.
1.13: The slide rule, 13.
1.14: Computation tables, 14.
1.15: Bibliography, 14.

2

BASIC MEASURES OF CENTRAL TENDENCY AND VARIABILITY

page 18

2.1: Introduction, 18. 2.2: A system of notation, 19. 2.3: The arithmetic mean, 21. 2.4: A mathematical property of the arithmetic mean, 22. 2.5: Uses of the arithmetic mean, 22. 2.6: The weighted arithmetic mean, 23. 2.7: The median, 24. 2.8: The range, 25. 2.9: The mean deviation, 25. 2.10: The variance, 26. 2.11: Properties of sums of squares, 27. 2.12: The standard deviation, 28. 2.13: The relation between the standard deviation and the range, 28. 2.14: The appropriateness of an average, 30.

3

FREQUENCY DISTRIBUTIONS

page 40

3.1: Tabulation, 40.
3.2: The frequency distribution, 42.
3.3: Class intervals,
44.
3.4: Graphical representation, 45.
3.5: Frequency polygons, 48.
3.6: Applications, 49.
3.7: Cumulative frequency, 51.
3.8: The cumulative frequency diagram, 53.

THE STATISTICS OF FREQUENCY DISTRIBUTIONS

page 65

4.1: The arithmetic mean, 65. 4.2: Remarks concerning the computation, 68. 4.3: The variance and standard deviation, 69. 4.4: Remarks concerning the computation of s^2 and s, 70. 4.5: Uses of the standard deviation, 71. 4.6: The median, 71. 4.7: Remarks concerning the computation of the median, 72. 4.8: Properties of the median, 73. 4.9: The usefulness of the median, 73. 4.10: The percentiles, 74. 4.11: The quartile deviation, 75. 4.12: Relations among the mean deviation, quartile deviation, and standard deviation, 76. 4.13: The coefficient of variation, 76. 4.14: The mode, 77. 4.15: Further remarks on the appropriateness of an average, 78. 4.16: Standard deviates, 79. 4.17: Moments, 79. 4.18: The third moment: skewness, 80. 4.19: The fourth moment, 81.

5

PROBABILITY

page 87

5.1: Introduction, 87. 5.2: Experiments, events, and sample space, 87. 5.3: Discrete and continuous space, 91. 5.4: Probability, 92. 5.5: The addition theorem, 94. 5.6: The multiplication theorem, 95. 5.7: Expected value, 97.

6

THE BINOMIAL DISTRIBUTION

page 101

6.1: Introduction, 101. 6.2: A fundamental assumption, 101. 6.3: Factorials, 102. 6.4: Permutations and combinations, 102. 6.5: Fundamental theorems concerning permutations, 102. 6.6: A fundamental theorem concerning combinations, 104. 6.7: Binomial probability, 104. 6.8: A decision problem, 105. 6.9: Rejection and acceptance regions, 106. 6.10: The type I error, 108. 6.11: The type II error, 109. 6.12: Remarks, 110. 6.13: The binomial distribution, 111. 6.14: The mean of a binomial distribution, 112. 6.15: The standard deviation of a binomial distribution, 112. 6.16: Other properties of the binomial distribution, 113. 6.17: Binomial probability calculations, 114. 6.18: Binomial probability paper, 115.

THE NORMAL FREQUENCY FUNCTION

page 119

7.1: Introduction, 119. 7.2: The normal frequency function, 120. 7.3: Area and probability, 124. 7.4: The standard form, 125. 7.5: Ordinates of the normal curve, 127. 7.6: Areas under the normal curve, 128. 7.7: Properties of the standard curve, 130. 7.8: $\emptyset(z)$ and f(x), 132. 7.9: Some applications, 133. 7.10: Applications to binomial distributions, 136. 7.11: Normal curve fitting, 139. 7.12: Graduation of normal data, 140. 7.13: Tests of normality, 141. 7.14: The cumulative normal curve, 142. 7.15: Probability paper, 143.

8

INFERENCES FROM SAMPLE MEANS

page 152

8.1: Introduction, 152. 8.2: The distribution of the sample mean, 152. 8.3: Applications, 154. 8.4: Quality control charts for means, 158. 8.5: A general theorem, 159. 8.6: Remarks, 161. 8.7: The Student-Fisher *t*-distribution, 161. 8.8: Degrees of freedom, 163. 8.9: Applications, 164. 8.10: The difference between two sample means when the population variances are unknown, 164. 8.11: Paired observations, 165. 8.12: The power of a test, 166. 8.13: Nonparametric tests, 169.

9

DISTRIBUTIONS RELATED TO BINOMIAL AND NORMAL DISTRIBUTIONS

page 176

9.1: The distribution of a proportion, 176.9.2: Control charts for p, 1799.3: The difference between two proportions, 180.9.4: The Poisson distribution, 181.9.5: The multinomial distribution, 184.

xiii

ESTIMATION

page 190

10.1: Introduction, 190. 10.2: The point estimate, 191. 10.3: The unbiased estimate of p, 191. 10.4: Estimates of σ^2 , 192. 10.5: Confidence limits for μ , 193. 10.6: Confidence limits for $\mu_x - \mu_y$, 195. 10.7: Confidence limits for μ_d , 196. 10.8: Confidence intervals and tests of hypotheses, 196. 10.9: Confidence limits for p, 196. 10.10: Confidence limits for σ^2 , 198. 10.11: Properties of point estimates, 198.

11

THE CHI-SQUARE DISTRIBUTION

page 202

11.1: An example, 202. 11.2: The Chi-square function, 203. 11.3: Chi-square and variance, 204. 11.4: Testing a multinomial distribution, 205. 11.5: Remarks on the Chi-square test, 206. 11.6: Goodness of fit to a normal curve, 207. 11.7: 2×2 tables, 208. 11.8: The Yates correction, 209. 11.9: An alternative formula, 210. 11.10: Mathematical models for 2×2 tables, 211. 11.11: $p \times q$ tables, 211. 11.12: The *F*-distribution: a test for $\sigma_1^2 = \sigma_2^2$, 212.

12

CURVE FITTING

page 219

12.1: Introduction, 219. 12.2: The straight line, 222. 12.3: Fitting a straight line by the method of least squares, 223. 12.4: An example, 225. 12.5: The derivation of the least squares formulas, 226. 12.6: Time series, 227. 12.7: Polynomial curve fitting, 228. 12.8: Regression, 228. 12.9: Explained and unexplained variance, 229. 12.10: A test for linearity, 233. 12.11: Testing hypotheses concerning slopes, 234.

CORRELATION

page 238

13.1: Introduction, 238. 13.2: Two lines of regression, 239. 13.3: The coefficient of correlation, 239. 13.4: Properties related to r, 241. 13.5: The computation of r, 243. 13.6: The computation of r for ungrouped data, 243. 13.7: Testing the hypothesis that $\rho=0$, 246. 13.8: The use of the Student-Fisher t-distribution, 246. 13.9: The Fisher z-transformation, 246. 13.10: Testing the hypothesis, $\rho_1=\rho_2$, 248. 13.11: The correlation table, 249. 13.12: The computation of r from a correlation table, 249. 13.13: Remarks on the computation of r, 252 13.14: The correlation surface, 253. 13.15: Non-linear regression, 257. 13.16: The correlation ratio, 259. 13.17: Tetrachoric correlation, 260. 13.18: Multiple correlation, 261. 13.19: Partial correlation, 262. 13.20: The uses and misuses of correlation, 263. 13.21: Multiple linear regression, 263. 13.22: Rank correlation, 264.

14

THE ANALYSIS OF VARIANCE

page 269

14.1: Simple analysis of variance, 269. 14.2: Remarks, 273.

15

NONPARAMETRIC STATISTICS

page 276

15.1: Introduction, 276. 15.2: The Tchebycheff inequality, 277. 15.3: The law of large numbers, 278. 15.4: The sign test, 279. 15.5: The run test for randomness, 280. 15.6: The use of ranks, 280. 15.7: The Wilcoxon ranksum test, 282. 15.8: Correlation of ranked data, 283.

REFERENCES

page 288

MATHEMATICAL TABLES

page 291

INDEX

page 313

ΧV

INTRODUCTION

1

"In 1786, I found that in Germany they were engaged in a species of political inquiry, to which they had given the name of Statistics... an inquiry for the purpose of ascertaining the political strength of a country, or questions respecting matters of state."

SIR JOHN SINCLAIR

The Statistical Account of Scotland, Vol. 21 (1791-1799)

1.1. The Origin of Statistics

It is curious that the modern science of statistics traces its origin to two quite dissimilar human interests, political states and games of chance. In the mid-sixteenth century Girolamo Cardano, the Italian mathematician, physician, and gambler wrote his Liber de Ludo Aleae (The Book on Games of Chance) in which appeared the first known study of the principles of probability. About a hundred years later, the gambler Chevalier de Méré proposed to Blaise Pascal the famous "Problem of the Points," which may be described as follows: Two men are playing a game of chance. The one first gaining a certain number of points wins the stake. They are forced to quit before the game is completed. Given the number of points each has won, how should the stake be divided? This problem offered a real challenge to the wits of the two astute French mathematicians, Pascal and Fermat. A lengthy correspondence between the two men led to solutions, not only of the problems proposed, but of more general ones. The methods employed by Cardano and Pascal may be said to represent the beginnings of the mathematics of probability, about which modern statistical theory centers today. The publication by Laplace in 1812 of the epoch-making Théorie Analytique des Probabilités laid a firm foundation for this theory.

In the mid-eighteenth century, statistics itself was born as a word describing the study of "the political arrangement of the modern states of

the known world." The description of states was at first verbal, but the increasing proportion of numerical data in the descriptions gradually gave the new word the quantitative connotation that is associated with it now. From the rather restricted study of data pertaining to a state, statistics

From the rather restricted study of data pertaining to a state, statistics branched out into other fields of investigation.

Between 1835 and 1870, the Belgian astronomer Quetelet was applying the theory of probability to anthropological measurements. His conclusions may be summarized and extended by stating that the same general laws of variation governing gambler's luck may be discovered in the statures of soldiers, the intelligence quotients of children, the blood pressures of adults, the speeds of molecules of a gas, and innumerable other aggregates of observations.

In more recent times, an English school of statisticians under the leader-ship of Karl Pearson (1857–1936) and Ronald A. Fisher, have made notable contributions to both theoretical and applied statistics. The power of general methods based on probability concepts became more clearly perceived, and as a result, applications have been made to many diverse fields of inquiry. An appreciation of the importance of the statistical method in man's attempt to come to grips with a marvelously complex physical and social world is a serious and legitimate aim of any educated person.

1.2. The Meaning of Statistics

The layman frequently conceives of statistics as a mass of figures or a collection of data such as we might find in the publications of the United States Census Bureau, among the records of a school principal, or in the files of a large hospital. The often repeated phrase "Statistics show ..." is likely to imply that a given mass of figures contains salient and unalterable characteristics that can easily be discerned among the mass by any person of normal intelligence. That the word *statistics* may apply to certain aggregates of figures is not to be denied, but that important facts contained therein are easily detected is by no means always true.

A second meaning of *statistics* is simply the plural of *statistic*, where a *statistic* is a certain kind of measure used to evaluate a selected property of statistic is a certain kind of measure used to evaluate a selected property of the collection of items under investigation. The average weight of a football squad, for example, may be found by adding the weights of the individual players and dividing by their number. The average thus obtained is a statistic.

A third meaning of statistics is of prime concern to us in this book. It is the science of assembling, analyzing, characterizing, and interpreting collections of data. In this sense, statistics is a field of study, a doctrine concerned with mathematical characterizations of aggregates of items.

Statistics, as a science, is fundamentally a branch of applied mathematics, just as mechanics is mathematics applied to problems connected with bodies subjected to forces. In statistics, the applications may be made to almost any aggregate of observations or measurements. For this reason it is useful in business, economics, sociology, biology, psychology, education, physics, chemistry, agriculture, and related fields.

1.3. Sample and Population

A population is a totality of all actual or conceivable objects of a certain class under consideration. More precisely, a population consists of numerical values connected with these objects. Head lengths of criminals, test scores of pupils, thicknesses of washers, lengths of life of electric light bulbs, or numbers of negative replies on a questionnaire may constitute populations of measurements or observations. Such aggregates may be finite or infinite, real or fictitious, but in this book we shall assume all populations to be essentially infinite. Problems connected with finite populations will not be considered here.

A sample is a finite number of objects selected from the population. If these are chosen in such a manner that one object has as good a chance of being selected as another, we say that we have a random sample. In this connection we should note that some objects may have identical measurements, so that one measurement is not necessarily as likely to occur as another.

A group of 10 washers taken at random from a barrel of them constitutes a sample from a larger aggregate or population, from, say 1000 washers, or from the even larger potential product of the machine manufacturing the washers.

If we know the average thickness of the 10 washers, what conclusions can we safely draw concerning the average thickness of all the washers in the barrel? From the statistical characteristics of the sample, what deductions may we make about the number of washers smaller than a desired dimension? This type of problem is basic in statistics and is one to which notable contributions have been made in recent years.

The greatest care must be exercised in selecting a sample that is truly random. The 10 washers should be selected from different parts of the barrel, not from a particular part such as the top. If a sample of student records is to be selected from an alphabetically arranged card file, it would be risky to select all the records from the same portion of the file. Student names might reflect racial factors. A sample of soil should be synthesized from several areas of the garden plot.

In more elaborate analyses, samples must be scientifically constructed so as to include appropriate numbers of different representative groups. Thus, the sample taken in an election poll must contain different economic, geographic, social, racial, or other groups if it is to reflect the voting tendencies of the country as a whole. The problem of obtaining an adequate sample is

Table 1-1 A PAGE OF RANDOM DIGITS

Table 1-1 A Fage of Random Digits									
70079	99064	97423	68793	91763	14940	55550	19900	36879	27718
74372	99540	00119	55063	97512	73665	45331	93614	49512	08359
43658	71456	63894	28132	98307	83300	08001	11186	21446	35864
72448	27714	10704	36331	68905	18477	42727	72133	25167	41601
43269	47963	88026	79532	82919	03920	10924	02018	13708	05281
66360	47852	32769	59586	00133	72584	26480	00245	48371	37526
22043	77224	26075	68778	87332	83287	54373	96391	82132	89338
78519	43251	18412	30777	14380	13550	37902	46169	27785	10488
58454	13026	26618	18537	44015	73261	42001	06096	21918	94440
00666	78245	32662	03375	54485	89848	90606	55556	49481	35329
80043	26080	72508	53576	49390	35273	86769	07108	66688	24636
53787	10007	66163	88811	21977	92078	95503	43655	57975	25768
88907	42653	05541	13459	89731	89459	98306	55222	32363	68675
76654	24020	67332	62362	65014	18061	92185	08657	92167	47793
11675	96819	10965	31214	39215	29883	34235	27113	22919	31278
90066	91253	59174	58312	84990		64054	34864	00483	17913
29480	78114	48305		84990 85176	52539 50048	62792		52055	93273
93992	71132	91042	67868 96303	11372	13817	15490	82816 19452	08265	57612
79938	37498	27019	18573	88617	31245	60208	53962	52981	04301
20506	31384	51173	33453	93156	43166	33599	98112	09422	48744
43006	16020	49784	09917	50236	59837	18739	85767	49111	51512
45186	04205	76923	06181	81538	68226	73500	60779	65584	24305
49966	94867	62902	43090	37205	72584	78048	98669	83267	13303
62224	77713	14540	24003	20499	32752	42271	75891	45681	44445
73217	21643	46106	73942	02936	45948	74850	17297	44957	31068
11219	20296	59367	31426	31166	66247	54764	91861	83130	37507
02164	54666	21868	65824	97370	23627	39822	29285	31387	17045
73171	27920	41254	60089	00693	58712	88187	56810	92728	07894
48435	58944	61989	84538	67060	69031	28814	31405	82384	77694
45687	46494	61920	26751	54241	09903	71831	98113	33094	99925
64573	28270	63695	16900	25980	61906	38832	44327	01141	37889
36345	24793	88754	95921	99442	30336	07705	41314	53028	07381
37402	15236	64920	25909	25085	85456	00198	32419	54583	83635
27358	35142	91012	35570	50420	30509	44150	99868	77894	05250
17222	24172	26021	79527	44721	19041	04399	74266	15134	17952
48436	19800	03441	60218	83099	10869	27264	06777	70388	34992
08752	26430	45080	80472	35599	34343	90581	46482	13441	74151
79075	92335	12474	33423	72174	02953	37198	97172	98019	92623
73073	26360	19111	65852	87760	41988	77620	83328	24394	23932
48418	80642	09023	48310	25218	79006	12709	39456	02883	83600
01362	30222	93728	16044	23187	40562	71067	13330	11022	17378
38148	24320	87981	57518	37136	04182	67913	88235	61865	24638
27411	82008	23860	45246	03403	97639	28686	67623	00542	63666
48322	46340	31022	55657	58297	36244	25091	75297	14695	75932
38823	78043	75095	58043	95125	74783	24693	06360	66853	66663
87891	01449	19122	70232	38118	30249	76453	20802	76374	83474
11627	55036	51014	95142	41014	28968	77021	79801	95957	87132
43277	09284	89837	17654	84726	49893	29601	02749	77246	21271
18946	64377	60317	28724	82044	03820	25767	53052	43304	70629
04996	65987	16738	51367	54872	93628	69984	29220	58652	06087

Reproduced from "A Million Random Digits," Rand Corporation, with the permission of The Free Press of Glencoe, Ill., 119 Lake St., Chicago, Ill.