This page intentionally left blank
## Contents

Preface ix
Acknowledgments xi

1 The Development of Statistics 1

- Evolution, Biometrics, and Eugenics 1
- The Definition of Statistics 6
- Probability 7
- The Normal Distribution 12
- Biometrics 14
- Statistical Criticism 18

2 Science, Psychology, and Statistics 21

- Determinism 21
- Probabilistic and Deterministic Models 26
- Science and Induction 27
- Inference 31
- Statistics in Psychology 33

3 Measurement 36

- In Respect of Measurement 36
- Some Fundamentals 39
- Error in Measurement 44
CONTENTS

4 The Organization of Data

- The Early Inventories 47
- Political Arithmetic 48
- Vital Statistics 50
- Graphical Methods 53

5 Probability

- The Early Beginnings 56
- The Beginnings 57
- The Meaning of Probability 60
- Formal Probability Theory 66

6 Distributions

- The Binomial Distribution 68
- The Poisson Distribution 70
- The Normal Distribution 72

7 Practical Inference

- Inverse Probability and the Foundations of Inference 77
- Fisherian Inference 81
- Bayes or \( p \leq .05 \)? 83

8 Sampling and Estimation

- Randomness and Random Numbers 85
- Combining Observations 88
- Sampling in Theory and in Practice 95
- The Theory of Estimation 98
- The Battle for Randomization 101

9 Sampling Distributions

- The Chi-Square Distribution 105
- The \( t \) Distribution 114
- The \( F \) Distribution 121
- The Central Limit Theorem 124
10 Comparisons, Correlations, and Predictions 127
Comparing Measurements 127
Galton’s Discovery of Regression 129
Galton’s Measure of Co-relation 138
The Coefficient of Correlation 141
Correlation - Controversies and Character 146

11 Factor Analysis 154
Factors 154
The Beginnings 156
Rewriting the Beginnings 162
The Practitioners 164

12 The Design of Experiments 171
The Problem of Control 171
Methods of Inquiry 173
The Concept of Statistical Control 176
The Linear Model 181
The Design of Experiments 182

13 Assessing Differences and Having Confidence 186
Fisherian Statistics 186
The Analysis of Variance 187
Multiple Comparison Procedures 195
Confidence Intervals and Significance Tests 199
A Note on ‘One-Tail’ and ‘Two-Tail’ Tests 203

14 Treatments and Effects: The Rise of ANOVA 205
The Beginnings 205
The Experimental Texts 205
The Journals and the Papers 207
The Statistical Texts 212
Expected Means Squares 213
15 The Statistical Hotpot

Times of Change 216
Neyman and Pearson 217
Statistics and Invective 224
Fisher versus Neyman and Pearson 228
Practical Statistics 233

References 236

Author Index 254

Subject Index 258
Preface

In this second edition I have made some corrections to errors in algebraic expressions that I missed in the first edition and I have briefly expanded on some sections of the original where I thought such expansion would make the narrative clearer or more useful. The main change is the inclusion of two new chapters; one on factor analysis and one on the rise of the use of ANOVA in psychological research. I am still of the opinion that factor analysis deserves its own historical account, but I am persuaded that the audience for such a work would be limited were the early mathematical contortions to be fully explored. I have tried to provide a brief non-mathematical background to its arrival on the statistical scene.

I realized that my account of ANOVA in the first edition did not do justice to the story of its adoption by psychology, and largely due to my re-reading of the work of Sandy Lovie (of the University of Liverpool, England) and Pat Lovie (of Keele University, England), who always write papers that I wish I had written, decided to try again. I hope that the Lovies will not be too disappointed by my attempt to summarize their sterling contributions to the history of both factor analysis and ANOVA.

As before, any errors and misinterpretations are my responsibility alone. I would welcome correspondence that points to alternative views.

I would like to give special thanks to the reviewers of the first edition for their kind comments and all those who have helped to bring about the revival of this work. In particular Professor Niels Waller of Vanderbilt University must be acknowledged for his insistent and encouraging remarks. I hope that I have
deserved them. My colleagues and many of my students at York University, Toronto, have been very supportive. Those students, both undergraduate and graduate, who have expressed their appreciation for my inclusion of some historical background in my classes on statistics and method have given me enormous satisfaction. This relatively short account is mainly for them, and I hope it will encourage some of them to explore some of these matters further.

There seems to be a slow realization among statistical consumers in psychology that there is more to the enterprise than null hypothesis significance testing, and other controversies to exercise us. It is still my firm belief that just a little more mathematical sophistication and just a little more historical knowledge would do a great deal for the way we carry on our research business in psychology.

The editors and production people at Lawrence Erlbaum, ever sharp and efficient, get on with the job and bring their expertise and sensible advice to the project and I very much appreciate their efforts.

My wife has sacrificed a great deal of her time and given me considerable help with the final stages of this revision and she and my family, even yet, put up with it all. Mere thanks are not sufficient.

Michael Cowles
Acknowledgments

I wish to express my appreciation to a number of individuals, institutions, and publishers for granting permission to reproduce material that appears in this book:


Excerpts from a number of papers by R.A. Fisher. Reprinted by permission of Professor J.H. Bennett on behalf of the copyright holders, the University of Adelaide.

Excerpts, figures and tables by permission of Hafner Press, a division of Macmillan Publishing Company from *Statistical methods for research workers* by R.A. Fisher. Copyright (c) 1970 by University of Adelaide.

Excerpts from volumes of *Biometrika*. Reprinted by permission of the Biometrika Trustees and from Oxford University Press.


ACKNOWLEDGMENTS

An excerpt reprinted with permission of Macmillan Publishing Company from *An introduction to the study of experimental medicine* by C. Bernard (H. C. Greene, Trans.), (c) 1927 (Original work published in 1865) and from *Science and human behavior* by B. F. Skinner (c) 1953 by Macmillan Publishing Company, renewed 1981 by B. F. Skinner.


Excerpts from various volumes of the Journal of the Royal Statistical Society, reprinted by permission of the Royal Statistical Society.


Excerpts from various volumes of Nature reprinted by permission of Macmillan Magazines Ltd.


ACKNOWLEDGMENTS


The central concern of the life sciences is the study of variation. To what extent does this individual or group of individuals differ from another? What are the reasons for the variability? Can the variability be controlled or manipulated? Do the similarities that exist spring from a common root? What are the effects of the variation on the life of the organisms? These are questions asked by biologists and psychologists alike.

The life-science disciplines are defined by the different emphases placed on observed variation, by the nature of the particular variables of interest, and by the ways in which the different variables contribute to the life and behavior of the subject matter. Change and diversity in nature rest on an organizing principle, the formulation of which has been said to be the single most influential scientific achievement of the 19th century: the theory of evolution by means of natural selection. The explication of the theory is attributed, rightly, to Charles Darwin (1809–1882). His book *The Origin of Species* was published in 1859, but a number of other scientists had written on the principle, in whole or in part, and these men were acknowledged by Darwin in later editions of his work.

Natural selection is possible because there is variation in living matter. The struggle for survival within and across species then ruthlessly favors the individuals that possess a combination of traits and characters, behavioral and physical, that allows them to cope with the total environment, exist, survive, and reproduce.

Not all sources of variability are biological. Many organisms to a greater or
lesser extent reshape their environment, their experience, and therefore their behavior through learning. In human beings this reshaping of the environment has reached its most sophisticated form in what has come to be called cultural evolution. A fundamental feature of the human condition, of human nature, is our ability to process a very great deal of information. Human beings have originality and creative powers that continually expand the boundaries of knowledge. And, perhaps most important of all, our language skills, verbal and written, allow for the accumulation of knowledge and its transmission from generation to generation. The rich diversity of human civilization stems from cultural, as well as genetic, diversity.

Curiosity about diversity and variability leads to attempts to classify and to measure. The ordering of diversity and the assessment of variation have spurred the development of measurement in the biological and social sciences, and the application of statistics is one strategy for handling the numerical data obtained.

As science has progressed, it has become increasingly concerned with quantification as a means of describing events. It is felt that precise and economical descriptions of events and the relationships among them are best achieved by measurement. Measurement is the link between mathematics and science, and the apparent (at any rate to mathematicians!) clarity and order of mathematics foster the scientist's urge to measure. The central importance of measurement was vigorously expounded by Francis Galton (1822–1911): "Until the phenomena of any branch of knowledge have been submitted to measurement and number it cannot assume the status and dignity of a Science."

These words formed part of the letterhead of the Department of Applied Statistics of University College, London, an institution that received much intellectual and financial support from Galton. And it is with Galton, who first formulated the method of correlation, that the common statistical procedures of modern social science began.

The nature of variation and the nature of inheritance in organisms were much-discussed and much-confused topics in the second half of the 19th century. Galton was concerned to make the study of heredity mathematical and to bring order into the chaos.

Francis Galton was Charles Darwin's cousin. Galton's mother was the daughter of Erasmus Darwin (1731–1802) by his second wife, and Darwin's father was Erasmus's son by his first. Darwin, who was 13 years Galton's senior, had returned home from a 5-year voyage as the naturalist on board H.M.S. Beagle (an Admiralty expeditionary ship) in 1836 and by 1838 had conceived of the principle of natural selection to account for some of the observations he had made on the expedition. The careers and personalities of Galton and Darwin were quite different. Darwin painstakingly marshaled evidence and single-mindedly buttressed his theory, but remained diffident about it, apparently
uncertain of its acceptance. In fact, it was only the inevitability of the announce­
ment of the independent discovery of the principle by Alfred Russell
Wallace (1823–1913) that forced Darwin to publish, some 20 years after he had
formed the idea. Galton, on the other hand, though a staid and formal Victorian,
was not without vanity, enjoying the fame and recognition brought to him by
his many publications on a bewildering variety of topics. The steady stream of
lectures, papers and books continued unabated from 1850 until shortly before
his death.

The notion of correlated variation was discussed by the new biologists. Darwin observes in *The Origin of Species*:

Many laws regulate variation, some few of which can be dimly seen, . . . I will
here only allude to what may be called correlated variation. Important changes
in the embryo or larva will probably entail changes in the mature animal . . .
Breeders believe that long limbs are almost always accompanied by an elongated
head . . . cats which are entirely white and have blue eyes are generally deaf . . .
it appears that white sheep and pigs are injured by certain plants whilst
dark-coloured individuals escape . . . (Darwin, 1859/1958, p. 34)

Of course, at this time, the hereditary mechanism was unknown, and, partly
in an attempt to elucidate it, Galton began, in the mid-1870s, to breed sweet
peas.1 The results of his study of the size of sweet pea seeds over two generations
were published in 1877. When a fixed size of parent seed was compared with
the mean size of the offspring seeds, Galton observed the tendency that he called
then *reversion* and later *regression* to the mean. The mean offspring size is not
as extreme as the parental size. Large parent seeds of a particular size produce
seeds that have a mean size that is larger than average, but not as large as the
parent size. The offspring of small parent seeds of a fixed size have a mean size
that is smaller than average but now this mean size is not as small as that of the
fixed parent size. This phenomenon is discussed later in more detail. For the
moment, suffice it to say that it is an arithmetical artifact arising from the fact
that offspring sizes do not match parental sizes absolutely uniformly. In other
words, the *correlation* is imperfect.

Galton misinterpreted this statistical phenomenon as a real trend toward a
reduction in population variability. Paradoxically, however, it led to the forma­
tion of the *Biometric School* of heredity and thus encouraged the development
of a great many statistical methods.

---

1 Mendel had already carried out his work with edible peas and thus begun the science of
genetics. The results of his work were published in a rather obscure journal in 1866 and the wider
scientific world remained oblivious of them until 1900.
Over the next several years Galton collected data on inherited human characteristics by the simple expedient of offering cash prizes for family records. From these data he arrived at the regression lines for hereditary stature. Figures showing these lines are shown in Chapter 10.

A common theme in Galton's work, and later that of Karl Pearson (1857–1936), was a particular social philosophy. Ronald Fisher (1890–1962) also subscribed to it, although, it must be admitted, it was not, as such, a direct influence on his work. These three men are the founders of what are now called classical statistics and all were eugenists. They believed that the most relevant and important variables in human affairs are inherited. One's ancestors rather than one's environmental experiences are the overriding determinants of intellectual capacity and personality as well as physical attributes. Human well-being, human personality, indeed human society, could therefore, they argued, be improved by encouraging the most able to have more children than the least able. MacKenzie (1981) and Cowan (1972, 1977) have argued that much of the early work in statistics and the controversies that arose among biologists and statisticians reflect the commitment of the founders of biometry, Pearson being the leader, to the eugenics movement.

In 1884, Galton financed and operated an anthropometric laboratory at the International Health Exhibition. For a charge of threepence, members of the public were measured. Visual and auditory acuity, weight, height, limb span, strength, and a number of other variables were recorded. Over 9,000 data sets were obtained, and, at the close of the exhibition, the equipment was transferred to the South Kensington Museum where data collection continued. Francis Galton was an avid measurer.

Karl Pearson (1930) relates that Galton's first forays into the problem of correlation involved ranking techniques, although he was aware that ranking methods could be cumbersome. How could one compare different measures of anthropometric variables? In a flash of illumination, Galton realized that characteristics measured on scales based on their own variability (we would now say standard score units) could be directly compared. This inspiration is certainly one of the most important in the early years of statistics. He recalls the occasion in Memories of my Life, published in 1908:

As these lines are being written, the circumstances under which I first clearly grasped the important generalisation that the laws of heredity were solely concerned with deviations expressed in statistical units are vividly recalled to my memory. It was in the grounds of Naworth Castle, where an invitation had been given to ramble freely. A temporary shower drove me to seek refuge in a reddish recess in the rock by the side of the pathway. There the idea flashed
across me and I forgot everything else for a moment in my great delight. (Galton, 1908, p. 300) 

This incident apparently took place in 1888, and before the year was out, *Co-relations and Their Measurement Chiefly From Anthropometric Data* had been presented to the Royal Society. In this paper Galton defines co-relation: "Two variable organs are said to be co-related when the variation of one is accompanied on the average by more or less variation of the other, and in the same direction" (Galton, 1888, p. 135).

The last five words of the quotation indicate that the notion of negative correlation had not then been conceived, but this brief but important paper shows that Galton fully understood the importance of his statistical approach. Shortly thereafter, mathematicians entered the picture with encouragement from some, but by no means all, biologists.

Much of the basic mathematics of correlation had, in fact, already been developed by the time of Galton's paper, but the utility of the procedure itself in this context had apparently eluded everyone. It was Karl Pearson, Galton's disciple and biographer, who, in 1896, set the concept on a sound mathematical foundation and presented statistics with the solution to the problem of representing covariation by means of a numerical index, the *coefficient of correlation*.

From these beginnings spring the whole corpus of present-day statistical techniques. George Udny Yule (1871–1951), an influential statistician who was not a eugenist, and Pearson himself elaborated the concepts of *multiple* and *partial* correlation. The general psychology of individual differences and research into the structure of human abilities and intelligence relied heavily on correlational techniques. The first third of the 20th century saw the introduction of *factor analysis* through the work of Charles Spearman (1863–1945), Sir Godfrey Thomson (1881–1955), Sir Cyril Burt (1883–1971), and Louis L. Thurstone (1887–1955).

A further prolific and fundamentally important stream of development arises from the work of Sir Ronald Fisher. The technique of *analysis of variance* was developed directly from the method of *intra-class correlation* – an index of the extent to which measurements in the same category or family are related, relative to other categories or families.

---

2 Karl Pearson (1914–1930) in the volume published in 1924, suggested that this spot deserves a commemorative plaque. Unfortunately, it looks as though the inspiration can never be so marked, for Kenna (1973), investigating the episode, reports that: "In the grounds of Naworth Castle there are not any rocks, reddish or otherwise, which could provide a recess, . . ." (p. 229), and he suggests that the location of the incident might have been Corby Castle.
Fisher studied mathematics at Cambridge but also pursued interests in biology and genetics. In 1913 he spent the summer working on a farm in Canada. He worked for a while with a City investment company and then found himself declared unfit for military service because of his extremely poor eyesight. He turned to school-teaching for which he had no talent and which he hated. In 1919 he had the opportunity of a post at University College with Karl Pearson, then head of the Department of Applied Statistics, but chose instead to develop a statistical laboratory at the Rothamsted Experimental Station near Harpenden in England, where he developed experimental methods for agricultural research. Over the next several years, relations between Pearson and Fisher became increasingly strained. They clashed on a variety of issues. Some of their disagreements helped, and some hindered, the development of statistics. Had they been collaborators and friends, rather than adversaries and enemies, statistics might have had a quite different history. In 1933 Fisher became Galton Professor of Eugenics at University College and in 1943 moved to Cambridge, where he was Professor of Genetics. Analysis of variance, which has had such far-reaching effects on experimentation in the behavioral sciences, was developed through attempts to tackle problems posed at Rothamsted.

It may be fairly said that the majority of texts on methodology and statistics in the social sciences are the offspring (diversity and selection notwithstanding!) of Fisher's books, Statistical Methods for Research Workers first published in 1925(a), and The Design of Experiments first published in 1935(a).

In succeeding chapters these statistical concepts are examined in more detail and their development elaborated, but first the use of the term statistics is explored a little further.

**The Definition of Statistics**

In an everyday sense when we think of statistics we think of facts and figures, of numerical descriptions of political and economic states (from which the word is derived), and of inventories of the various aspects of our social organization. The history of statistical procedures in this sense goes back to the beginnings of human civilization. When trade and commerce began, when governments imposed taxes, numerical records were kept. The counting of people, goods, and chattels was regularly carried out in the Roman Empire, the Domesday Book attempted to describe the state of England for the Norman conquerors, and government agencies the world over expend a great deal of money and

---

3 Maurice Kendall (1963) says of this work, "It is not an easy book. Somebody once said that no student should attempt to read it unless he had read it before" (p. 2).
energy in collecting and tabulating such information in the present day. Statistics are used to describe and summarize, in numerical terms, a wide variety of situations.

But there is another more recently-developed activity subsumed under the term statistics: the practice of not only collecting and collating numerical facts, but also the process of reasoning about them. Going beyond the data, making inferences and drawing conclusions with greater or lesser degrees of certainty in an orderly and consistent fashion is the aim of modern applied statistics. In this sense statistical reasoning did not begin until fairly late in the 17th century and then only in a quite limited way. The sophisticated models now employed, backed by theoretical formulations that are often complex, are all less than 100 years old. Westergaard (1932) points to the confusions that sometimes arise because the word statistics is used to signify both collections of measurements and reasoning about them, and that in former times it referred merely to descriptions of states in both numerical and non-numerical terms.

In adopting the statistical inferential strategy the experimentalist in the life sciences is accepting the intrinsic variability of the subject matter. In recognizing a range of possibilities, the scientist comes four-square against the problem of deciding whether or not the particular set of observations he or she has collected can reasonably be expected to reflect the characteristics of the total range. This is the problem of parameter estimation, the task of estimating population values (parameters) from a consideration of the measurements made on a particular population subset – the sample statistics. A second task for inferential statistics is hypothesis testing, the process of judging whether or not a particular statistical outcome is likely or unlikely to be due to chance. The statistical inferential strategy depends on a knowledge of probabilities.

This aspect of statistics has grown out of three activities that, at first glance, appear to be quite different but in fact have some close links. They are actuarial prediction, gambling, and error assessment. Each addresses the problems of making decisions, evaluating outcomes, and testing predictions in the face of uncertainty, and each has contributed to the development of probability theory.

**PROBABILITY**

Statistical operations are often thought of as practical applications of previously developed probability theory. The fact is, however, that almost all our present-day statistical techniques have arisen from attempts to answer real-life problems of prediction and error assessment, and theoretical developments have not always paralleled technical accomplishments. Box (1984) has reviewed the scientific context of a range of statistical advances and shown that the fundamental methods evolved from the work of practising scientists.
John Graunt, a London haberdasher, born in 1620, is credited with the first attempt to predict and explain a number of social phenomena from a consideration of actuarial tables. He compiled his tables from *Bills of Mortality*, the parish accounts of deaths that were regularly, if somewhat crudely, recorded from the beginning of the 17th century.

Graunt recognizes that the question might be asked: "To what purpose tends all this laborious buzzing, and groping? To know, 1. the number of the People? 2. How many Males, and Females? 3. How many Married, and single?" (Graunt, 1662/1975, p. 77), and says: "To this I might answer in general by saying, that those, who cannot apprehend the reason of these Enquiries, are unfit to trouble themselves to ask them." (p. 77).

Graunt reassured readers of this quite remarkable work:

The *Lunaticks* are also but few, viz. 158 in 229250 though I fear many more than are set down in our *Bills* . . .

So that, this *Casualty* being so uncertain, I shall not force my self to make any inference from the numbers, and proportions we finde in our Bills concerning it: onely I dare ensure any man at this present, well in his Wits, for one in the thousand, that he shall not die a *Lunatick* in *Bedlam*, within these seven years, because I finde not above one in about one thousand five hundred have done so. (pp. 35–36)

Here is an inference based on numerical data and couched in terms not so very far removed from those in reports in the modern literature. Graunt's work was immediately recognized as being of great importance, and the King himself (Charles II) supported his election to the recently incorporated Royal Society.

A few years earlier the seeds of modern probability theory were being sown in France. At this time gambling was a popular habit in fashionable society and a range of games of chance was being played. For experienced players the odds applicable to various situations must have been appreciated, but no formal methods for calculating the chances of various outcomes were available. Antoine Gombauld, the Chevalier de Méré, a "man-about-town" and gambler with a scientific and mathematical turn of mind, consulted his friend, Blaise Pascal (1623–1662), a philosopher, scientist, and mathematician, hoping that

---

4 But note that there are hints of probability concepts in mathematics going back at least as far as the 12th century and that Girolamo Cardano wrote *Liber de Ludo Aleae*, (The Book on Games of Chance) a century before it was published in 1663 (see Ore, 1953). There is also no doubt that quite early in human civilization, there was an appreciation of long-run relative frequencies, randomness, and degrees of likelihood in gaming, and some quite formal concepts are to be found in Greek and Roman writings.
he would be able to resolve questions on calculation of expected (probable) frequency of gains and losses, as well as on the fair division of the stakes in games that were interrupted. Consideration of these questions led to correspondence between Pascal and his fellow mathematician Pierre Fermat (1601–1665). No doubt their advice aided de Méré's game. More significantly, it was from this exchange that some of the foundations of probability theory and combinatorial algebra were laid.


Pascal had connected the study of probability with the arithmetic triangle (Fig. 1.1), for which he discovered new properties, although the triangle was known in China at least five hundred years earlier. Proofs of the triangle's properties were obtained by mathematical induction or reasoning by recurrence.

\[
\begin{array}{cccccccc}
& & & & & & & 1 \\
& & & & & 1 & & \\
& & & & 1 & 2 & 1 & \\
& & & 1 & 3 & 3 & 1 & \\
& & 1 & 4 & 6 & 4 & 1 & \\
& 1 & 5 & 10 & 10 & 5 & 1 & \\
1 & 6 & 15 & 20 & 15 & 6 & 1 & \\
1 & 7 & 21 & 35 & 35 & 21 & 7 & 1
\end{array}
\]

\[
\left(\frac{1}{2} + \frac{1}{2}\right)^4 = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16}
\]

**FIG. 1.1** Pascal's Triangle

---

5 Poisson (1781–1840), writing of this episode in 1837 says, "A problem concerning games of chance, proposed by a man of the world to an austere Jansenist, was the origin of the calculus of probabilities" (quoted by Struik, 1954, p. 145). De Méré was certainly "a man of the world" and Pascal did become austere and religious, but at the time of de Méré's questions Pascal was in his so-called "worldly period" (1652–1654). I am indebted to my father-in-law, the late Professor F.T.H. Fletcher, for many insights into the life of Pascal.
Pascal’s triangle, as it is known in the West, is a tabulation of the binomial coefficients that may be obtained from the expansion of \((P + Q)^n\) where \(P = Q = \frac{1}{2}\). The expansion was developed by Sir Isaac Newton (1642–1727), and, independently, by the Scottish mathematician, James Gregory (1638–1675). Gregory discovered the rule about 1670. Newton communicated it to the Royal Society in 1676, although later that year he explained that he had first formulated it in 1664 while he was a Cambridge undergraduate. The example shown in Fig. 1.1 demonstrates that the expansion of \((\frac{1}{2} + \frac{1}{2})^4\) generates, in the numerators of the expression, the numbers in the fifth row of Pascal’s triangle. The terms of this expression also give us the five expected frequencies of outcome (0, 1, 2, 3, or 4 heads) or the probabilities when a fair coin is tossed four times. Simple experiment will demonstrate that the actual outcomes in the "real world" of coin tossing closely approximate the distribution that has been calculated from a mathematical abstraction.

During the 18th century the theory of probability attracted the interest of many brilliant minds. Among them was a friend and admirer of Newton, Abraham De Moivre (1667–1754). De Moivre, a French Huguenot, was interned in 1685 after the revocation by Louis XIV of the Edict of Nantes, an edict which had guaranteed toleration to French Protestants. He was released in 1688, fled to England, and spent the remainder of his life in London. De Moivre published what might be described as a gambler’s manual, entitled *The Doctrine of Chances or a Method of Calculating the Probabilities of Events in Play*. In the second edition of this work, published in 1738, and in a revised third edition published posthumously in 1756, De Moivre (1756/1967) demonstrated a method, which he had first devised in 1733, of approximating the sum of a very large number of binomial terms when \(n\) in \((P + Q)^n\) is very large (an immensely laborious computation from the basic expansion).

It may be appreciated that as \(n\) grows larger, the number of terms in the expansion also grows larger. The graph of the distribution begins to resemble a smooth curve (Fig. 1.2), a bell-shaped symmetrical distribution that held great interest in mathematical terms but little practical utility outside of gaming.

It is safe to say that no other theoretical mathematical abstraction has had such an important influence on psychology and the social sciences as that bell-shaped curve now commonly known by the name that Karl Pearson decided on—*the normal distribution*—although he was not the first to use the term. Pierre Laplace (1749–1827) independently derived the function and brought together much of the earlier work on probability in *Théorie Analytique des Probabilités*, published in 1812. It was his work, as well as contributions by many others, that interpreted the curve as the *Law of Error* and showed that it could be applied to variable results obtained in multiple observations. One of the first applications of the distribution outside of gaming was in the assessment of errors in
FIG. 1.2 The Binomial Distribution for \( N = 12 \) and the Normal Distribution

\[
\binom{12}{x} \left( \frac{1}{2} \right)^{12} = 1
\]
astronomical observations. Later the utility of the "law" in error assessment was extended to land surveying and even to range estimation problems in artillery fire. Indeed, between 1800 and 1820 the foundations of the theory of error distribution were laid.

Carl Friedrich Gauss (1777–1855), perhaps the greatest mathematician of all time, also made important contributions to work in this area. He was a consultant to the governments of Hanover and of Denmark when they undertook geodetic surveys. The function that helped to rationalize the combination of observations is sometimes called the Laplace-Gaussian distribution.

Following the work of Laplace and Gauss, the development of mathematical probability theory slowed somewhat and not a great deal of progress was made until the present century. But it was during the 19th century, through the development of life insurance companies and through the growth of statistical approaches in the social and biological sciences, that the applications of probability theory burgeoned. Augustus De Morgan (1806–1871), for example, attempted to reduce the constructs of probability to straightforward rules of thumb. His work An Essay on Probabilities and on Their Application to Life Contingencies and Insurance Offices, published in 1838, is full of practical advice and is commented on by Walker (1929).

THE NORMAL DISTRIBUTION

The normal distribution was so named because many biological variables when measured in large groups of individuals, and plotted as frequency distributions, do show close approximations to the curve. It is partly for this reason that the mathematics of the distribution are used in data assessment in the social sciences and in biology. The responsibility, as well as the credit, for this extension of the use of calculations designed to estimate error or gambling expectancies into the examination of human characteristics rests with Lambert Adolphe Quetelet (1796–1874), a Belgian astronomer.

In 1835 Quetelet described his concept of the average man – l'homme moyen. L'homme moyen is Nature's ideal, an ideal that corresponds with a middle, measured value. But Nature makes errors, and in, as it were, missing the target, produces the variability observed in human traits and physical characters. More importantly, the extent and frequency of these errors often conform to the law of frequency of error – the normal distribution.

John Venn (1834–1933), the English logician, objected to the use of the word error in this context: "When Nature presents us with a group of objects of every kind, it is using rather a bold metaphor to speak in this case also of a law of error" (Venn, 1888, p. 42), but the analogy was attractive to some.

Quetelet examined the distribution of the measurements of the chest girths
of 5,738 Scottish soldiers, these data having been extracted from the 13th volume of the *Edinburgh Medical Journal*. There is no doubt that the measurements closely approximate to a normal curve. In another attempt to exemplify the law, Quetelet examined the heights of 100,000 French conscripts. Here he noticed a discrepancy between observed and predicted values:

> The official documents would make it appear that, of the 100,000 men, 28,620 are of less height than 5 feet 2 inches: calculation gives only 26,345. Is it not a fair presumption, that the 2,275 men who constitute the difference of these numbers have been fraudulently rejected? We can readily understand that it is an easy matter to reduce one’s height a half-inch, or an inch, when so great an interest is at stake as that of being rejected. (Quetelet, 1835/1849, p. 98)

Whether or not the allegation stated here – that short (but not too short) Frenchmen have stooped so low as to avoid military service – is true is no longer an issue. A more important point is noted by Boring (1920):

> While admitting the dependence of the law on experience, Quetelet proceeds in numerous cases to analyze experience by means of it. Such a double-edged sword is a peculiarly effective weapon, and it is no wonder that subsequent investigators were tempted to use it in spite of the necessary rules of scientific warfare. (Boring, 1920, p. 11)

The use of the normal curve in statistics is not, however, based solely on the fact that it can be used to describe the frequency distribution of many observed characteristics. It has a much more fundamental significance in inferential statistics, as will be seen, and the distribution and its properties appear in many parts of this book.

Galton first became aware of the distribution from his friend William Spottiswoode, who in 1862 became Secretary of the Royal Geographical Society, but it was the work of Quetelet that greatly impressed him. Many of the data sets he collected approximated to the law and he seemed, on occasion, to be almost mystically impressed with it.

> I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error." The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement amidst the wildest confusion. The huger the mob and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along. (Galton, 1889, p. 66)
14 1. THE DEVELOPMENT OF STATISTICS

This rather theological attitude toward the distribution echoes De Moivre, who, over a century before, proclaimed in *The Doctrine of Chances*:

Altho' chance produces irregularities, still the Odds will be infinitely great, that in the process of Time, those irregularities will bear no proportion to the recurrency of that Order which naturally results from ORIGINAL DESIGN... Such Laws, as well as the original Design and Purpose of their Establishment, must all be from without... if we blind not ourselves with metaphysical dust, we shall be led, by a short and obvious way, to the acknowledgement of the great MAKER and GOVENOUR of all; Himself all-wise, all-powerful and good. (De Moivre, 1756/1967 p. 251-252)

The ready acceptance of the normal distribution as a law of nature encouraged its wide application and also produced consternation when exceptions were observed. Quetelet himself admitted the possibility of the existence of asymmetric distributions, and Galton was at times less lyrical, for critics had objected to the use of the distribution, not as a practical tool to be used with caution where it seemed appropriate, but as a sort of divine rule:

It has been objected to some of my former work, especially in *Hereditary Genius*, that I pushed the application of the Law of Frequency of Error somewhat too far.

I may have done so, rather by incautious phrases than in reality; ... I am satisfied to claim the Normal Law is a fair average representation of the observed Curves during nine-tenths of their course; ... (Galton, 1889, p. 56) 6

**BIOMETRICS**

In 1890, Walter F. R. Weldon (1860–1906) was appointed to the Chair of Zoology at University College, London. He was greatly impressed and much influenced by Galton's *Natural Inheritance*. Not only did the book show him how the frequency of the deviations from a "type" might be measured, it opened up for him, and for other zoologists, a host of biometric problems. In two papers published in 1890 and 1892, Weldon showed that various measurements on shrimps might be assessed using the normal distribution. He also demonstrated interrelationships (correlations) between two variables within the individuals.

But the critical factor in Weldon's contribution to the development of statistics was his professorial appointment, for this brought him into contact with Karl Pearson, then Professor of Applied Mathematics and Mechanics, a post Pearson had held since 1884. Weldon was attempting to remedy his weakness in

---

6 Note that this quotation and the previous one from Galton are 10 pages apart in the same work!
mathematics so that he could extend his research, and he approached Pearson for help. His enthusiasm for the biometric approach drew Pearson away from more orthodox work.

A second important link was with Galton, who had reviewed Weldon's first paper on variation in shrimps. Galton supported and encouraged the work of these two younger men until his death, and, under the terms of his will, left £45,000 to endow a Chair of Eugenics at the University of London, together with the wish that the post might be offered first to Karl Pearson. The offer was made and accepted.

In 1904, Galton had offered the University of London £500 to establish the study of national eugenics. Pearson was a member of the Committee that the University set up, and the outcome was a decision to appoint the Galton Research Fellow at what was to be named the Eugenics Record Office. This became the Galton Laboratory for National Eugenics in 1906, and yet more financial assistance was provided by Galton. Pearson, still Professor of Applied Mathematics, was its Director as well as Head of the Biometrics Laboratory. This latter received much of its funding over many years from grants from the Worshipful Company of Drapers, which first gave money to the University in 1903.

Pearson's appointment to the Galton Chair brought applied statistics, biometrics, and eugenics together under his direction at University College. It cannot however be claimed absolutely that the day-to-day work of these units was driven by a common theme. Applied statistics and biometrics were primarily concerned with the development and application of statistical techniques to a variety of problems, including anthropometric investigations; the Eugenics Laboratory collected extensive family pedigrees and examined actuarial death rates. Of course Pearson coordinated all the work, and there was interchange and exchange among the staff that worked with him, but Magnello (1998, 1999) has argued that there was not a single unifying purpose in Pearson's research. Others, notably MacKenzie (1981), Kevles (1985), and Porter (1986), have promoted the view that eugenics was the driving force behind Pearson's statistical endeavors.

Pearson was not a formal member of the eugenics movement. He did not join the Eugenics Education Society, and apparently he tried to keep the two laboratories administratively separate, maintaining separate financial accounts, for example, but it has to be recognized that his personal views of the human condition and its future included the conviction that eugenics was of critical importance. There was an obvious and persistent intermingling of statistical results and eugenics in his pronouncements. For example, in his Huxley Lecture in 1903 (published in Biometrika in 1903 and 1904), on topics that were clearly biometric, having to do with his researches on relationships between moral and
intellectual variables, he ended with a plea, if not a rallying cry, for eugenics:

The mentally better stock in the nation is not reproducing itself at the same rate as it did of old; the less able, and the less energetic, are more fertile than the better stocks. . . . The only remedy, if one be possible at all, is to alter the relative fertility of the good and the bad stocks in the community. . . . intelligence can be aided and be trained, but no training or education can create it. You must breed it, that is the broad result for statecraft which flows from the equality in inheritance of the psychical and the physical characters in man. (Pearson, 1904a, pp. 179-180).

Pearson’s contribution was monumental, for in less than 8 years, between 1893 and 1901, he published over 30 papers on statistical methods. The first was written as a result of Weldon’s discovery that the distribution of one set of measurements of the characteristics of crabs, collected at the zoological station at Naples in 1892, was “double-humped.” The distribution was reduced to the sum of two normal curves. Pearson (1894) proceeded to investigate the general problem of fitting observed distributions to theoretical curves. This work was to lead directly to the formulation of the $\chi^2$ test of “goodness of fit” in 1900, one of the most important developments in the history of statistics.

Weldon approached the problem of discrepancies between theory and observation in a much more empirical way, tossing coins and dice and comparing the outcomes with the binomial model. These data helped to produce another line of development.

In a letter to Galton, written in 1894, Weldon asks for a comment on the results of 7,000 tossings of 12 dice collected for him by a clerk at University College:

A day or two ago Pearson wanted some records of the kind in a hurry, in order to illustrate a lecture, and I gave him the record of the clerk’s 7000 tosses . . . on examination he rejects them, because he thinks the deviation from the theoretically most probable result is so great as to make the record intrinsically incredible. (quoted by E. S. Pearson, 1965, p. 11)

This incident set off a good deal of correspondence between Karl Pearson, F.Y. Edgeworth (1845–1926), an economist and statistician, and Weldon, the details of which are now only of minor importance. But, as Karl Pearson remarked, “Probabilities are very slippery things” (quoted by E. S. Pearson, 1965, p.14), and the search for criteria by which to assess the differences between observed and theoretical frequencies, and whether or not they could be reasonably attributed to chance sampling fluctuations, began. Statistical research rapidly expanded into careful examination of distributions other than the normal curve and eventually into the properties of sampling distributions,
particularly through the seminal work of Ronald Fisher.

In developing his research into the properties of the probability distributions of statistics, Fisher investigated the basis of hypothesis testing and the foundations of all the well-known tests of *statistical significance*. Fisher's assertion that $p = .05$ (1 in 20) is the probability that is convenient for judging whether or not a deviation is to be considered significant (i.e. unlikely to be due to chance), has profoundly affected research in the social sciences, although it should be noted that he was not the originator of the convention (Cowles & Davis, 1982a).

Of course, the development of statistical methods does not end here, nor have all the threads been drawn together. Discussion of the important contribution of W. S. Gosset ("Student," 1876–1937) to small sample work and the refinements introduced into hypothesis testing by Karl Pearson's son, Egon S. Pearson (1895–1980) and Jerzy Neyman (1899–1981) will be found in later chapters, when the earlier details have been elaborated.

**Biometrics and Genetics**

The early years of the biometric school were surrounded by controversy. Pearson and Weldon held fast to the view that evolution took place by the continuous selections of variations that were favorable to organisms in their environment. The rediscovery of Mendel's work in 1900 supported the concept that heredity depends on self-reproducing particles (what we now call *genes*), and that inherited variation is discontinuous and saltatory. The source of the development of higher types was occasional genetic jumps or mutations. Curiously enough, this was the view of evolution that Galton had supported. His misinterpretation of the purely statistical phenomenon of regression led him to the notion that a distinction had to be made between variations from the mean that regress and what he called "sports" (a breeder's term for an animal or plant variety that appears apparently spontaneously) that will not.

A champion of the position that mutations were of critical importance in the evolutionary process was William Bateson (1861–1926) and a prolonged and bitter argument with the biometricians ensued. The *Evolution Committee* of the Royal Society broke down over the dispute. *Biometrika* was founded by Pearson and Weldon, with Galton's financial support, in 1900, after the Royal Society had allowed Bateson to publish a detailed criticism of a paper submitted by Pearson before the paper itself had been issued. Britain's important scientific journal, *Nature*, took the biometricians' side and would not print letters from Bateson. Pearson replied to Bateson's criticisms in *Biometrika* but refused to accept Bateson's rejoinders, whereupon Bateson had them privately printed by the Cambridge University Press in the format of *Biometrika*!

At the British Association meeting in Cambridge in 1904, Bateson, then...
President of the Zoological Section, took the opportunity to deliver a bitter attack on the biometric school. Dramatically waving aloft the published volumes of *Biometrika*, he pronounced them worthless and he described Pearson’s correlation tables as: “a Procrustean bed into which the biometrician fits his unanalysed data.” (quoted by Julian Huxley, 1949).

It is even said that Pearson and Bateson refused to shake hands at Weldon’s funeral. Nevertheless, after Weldon’s death the controversy cooled. Pearson’s work became more concerned with the theory of statistics, although the influence of his eugenic philosophy was still in evidence, and by 1910, when Bateson became Director of the John Innes Horticultural Institute, the argument had died.

However, some statistical aspects of this contentious debate predated the evolution dispute, and echoes of them – indeed, marked reverberations from them – are still around today, although of course Mendelian and Darwinian thinking are completely reconciled.

**STATISTICAL CRITICISM**

Statistics has been called the “science of averages,” and this definition is not meant in a kindly way. The great physiologist Claude Bernard (1813–1878) maintained that the use of averages in physiology could not be countenanced:

> because the true relations of phenomena disappear in the average; when dealing with complex and variable experiments, we must study their various circumstances, and then present our most perfect experiment as a type, which, however, still stands for true facts.
> 
> ... averages must therefore be rejected, because they confuse while aiming to unify, and distort while aiming to simplify. (Bernard, 1865/1927, p. 135)

Now it is, of course, true that lumping measurements together may not give us anything more than a picture of the lumping together, and the average value may not be anything like any one individual measurement at all, but Bernard’s ideal type fails to acknowledge the reality of individual differences. A rather memorable example of a very real confusion is given by Bernard (1865/1927):

> A startling instance of this kind was invented by a physiologist who took urine from a railway station urinal where people of all nations passed, and who believed that he could thus present an analysis of average European urine! (pp. 134-135).

A less memorable, but just as telling, example is that of the social psychologist who solemnly reports “mean social class.” Pearson (1906) notes that:

> One of the blows to Weldon, which resulted from his biometric view of life
was that his biological friends could not appreciate his new enthusiasms. They
could not understand how the Museum "specimen" was in the future to be
replaced by the "sample" of 500 to 1000 individuals. (p. 37)

The view is still not wholly appreciated. Many psychologists subscribe to
the position that the most pressing problems of the discipline, and certainly the
ones of most practical interest, are problems of individual behavior. A major
criticism of the effect of the use of the statistical approach in psychological
research is the failure to differentiate adequately between general propositions
that apply to most, if not all, members of a particular group and statistical
propositions that apply to some aggregated measure of the members of the
group. The latter approach discounts the exceptions to the statistical aggregate,
which not only may be the most interesting but may, on occasion, constitute a
large proportion of the group.

Controversy abounds in the field of measurement, probability, and statistics,
and the methods employed are open to criticism, revision, and downright
rejection. On the other hand, measurement and statistics play a leading role in
psychological research, and the greatest danger seems to lie in a nonawareness
of the limitations of the statistical approach and the bases of their development,
as well as the use of techniques, assisted by the high-speed computer, as recipes
for data manipulation.

Miller (1963) observed of Fisher, "Few psychologists have educated us as
rapidly, or have influenced our work as pervasively, as did this fervent, clear­
headed statistician." (p. 157).

Hogben (1957) certainly agrees that Fisher has been enormously influential
but he objects to Fisher's confidence in his own intuitions:

This intrepid belief in what he disarming calls common sense... has led Fisher
... to advance a battery of concepts for thesematic credentials of which neither
he nor his disciples offer any justification en rapport with the generally accepted
tenets of the classical theory of probability. (Hogben, 1957, p. 504)

Hogben also expresses a thought often shared by natural scientists when
they review psychological research, that:

Acceptability of a statistically significant result of an experiment on animal
behaviour in contradistinction to a result which the investigator can repeat before
a critical audience naturally promotes a high output of publication. Hence the
argument that the techniques work has a tempting appeal to young biologists.
(Hogben, 1957, p. 27)

Experimental psychologists may well agree that the tightly controlled ex­
periment is the apotheosis of classical scientific method, but they are not so
arrogant as to suppose that their subject matter will *necessarily* submit to this form of analysis, and they turn, almost inevitably, to statistical, as opposed to experimental, control. This is not a muddle-headed notion, but it does present dangers if it is accepted without caution.

A balanced, but not uncritical, view of the utility of statistics can be arrived at from a consideration of the forces that shaped the discipline and an examination of its development. Whether or not this is an assertion that anyone, let alone the author of this book, can justify remains to be seen.

Yet there are Writers, of a Class indeed very different from that of *James Bernoulli*, who insinuate as if the *Doctrine of Probabilities* could have no place in any serious Enquiry; and that Studies of this kind, trivial and easy as they be, rather disqualify a man for reasoning on any other subject. Let the Reader chuse. (De Moivre, 1756/1967, p. 254)
References


Bernoulli, D. (1966). Part four of the Art of Conjecturing showing the use and application of the preceding treatise in civil, moral and economic affairs (Bing Sung, Trans.). Cambridge, MA: Harvard University, Department of Statistics, Tech. Report No. 2. (Original work published 1865)


REFERENCES

Graunt, J. (1975). *Natural and political observations mentioned in a following index and made upon the bills of mortality.* New York: Arno Press. (Original work published 1662)
Heron, D. (1911). The danger of certain formulae suggested as substitutes for the correlation coefficient. *Biometrika, 8,* 109–122.
REFERENCES


Kendall, M. G. (1961). Daniel Bernoulli on maximum likelihood. Biometrika, 48, 1–18. [This paper is followed by a translation by C. G. Allen of D. Bernoulli’s The most probable choice between several discrepant observations and the formation therefrom of the most likely induction 1777(?);, together with a commentary by Leonard Euler (1707–1783).]


246 REFERENCES


Pearson, K. (1900a). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, 50, 157–175.


Pearson, K. (1911). On the probability that two independent distributions of frequency are really samples from the same population. *Biometrika*, 8, 250–254.


Pearson, K. (1924b). On the difference and the doublet tests for ascertaining whether two samples have been drawn from the same population. *Biometrika, 16*, 249–252.


References


Rhodes, E.C. (1924). On the problem whether two given samples can be supposed to have been drawn from the same population. *Biometrika, 16,* 239–248.


REFERENCES


Simpson, T. (1755). A letter to the Right Honourable George Earl of Macclesfield, President of the Royal Society, on the advantage arising in taking the mean of a number of observations, in practical astronomy. Philosophical Transactions of the Royal Society, 49, 82–93.

Simpson, T. (1757). On the advantage arising from taking the mean of a number of observations, in practical astronomy, wherein the odds that the result in this way is more exact than from one single observation is evinced and the utility of the method in practise clearly made appear. In Miscellaneous tracts on some curious and very interesting subjects in mechanics, physical astronomy and speculative mathematics (pp. 64–75). London: Printed for J. Nourse.


“Student” (1925). New tables for testing the significance of observations. *Metron, 5*, 25–32.


