

'Student'

*A Statistical
Biography of*
WILLIAM
SEALY
GOSSET



E. S. PEARSON

Edited by
R. L. PLACKETT
and
G. A. BARNARD

'Student'

A Statistical Biography of
William Sealy Gosset

Based on writings by
E.S. Pearson

Edited and augmented by
R.L. Plackett

With the assistance of
G.A. Barnard

CLARENDON PRESS · OXFORD
1990

Contents

1. Introduction	1
2. Background	4
2.1 Introduction	4
2.2 Combination of observations	5
2.3 Biometry	8
2.4 Books used by Gosset	10
2.5 Karl Pearson's lectures	13
3. William Sealy Gosset	15
3.1 Introduction	15
3.2 Early statistical career	16
3.3 Correspondence	17
3.4 Professional activities	18
3.5 Character and personality	21
4. Karl Pearson	23
4.1 Introduction	23
4.2 Correlation	23
4.3 Time series	29
4.4 Discrete distributions	34
4.5 Comment and criticism	39
5. Ronald A. Fisher	45
5.1 Introduction	45
5.2 Probability integral of t	45
5.3 <i>Statistical methods for research workers</i>	53
5.4 Design of experiments	54
5.5 Advice and argument	65
6. Egon S. Pearson	70
6.1 Historical introduction	70
6.2 Difficulties about z and χ^2	78
6.3 Order statistics and range	84

6.4 Random numbers and sampling experiments	88
6.5 Review of <i>Statistical methods for research workers</i>	95
6.6 Effects of non-normality	101
6.7 Differences on z-test and design of experiments	104
6.8 Final comments by E. S. Pearson	107
7. General commentary	111
Notes on the text	120
Bibliography	127
Subject index	137
Name index	140

Preface

'Student' was the pseudonym which William Sealy Gosset used when publishing his scientific work. He was employed for the whole of his working life as a Brewer with Arthur Guinness, Son & Co. Ltd., as the firm was then called, but his publications over thirty years during the early part of the twentieth century, and his friendship with other statistical pioneers, have had a profound effect on the practical use of statistics in industry and agriculture. The account of his work and correspondence which follows has been developed from the writings of Egon Sharpe Pearson, who knew him well. After Gosset's death in 1937, Pearson published an essay on 'Student' as statistician, and towards the end of his own life in 1980 he commented in typescript on his correspondence with Gosset, as well as on the earlier correspondence between his father Karl Pearson and Gosset. We have attempted to collate and edit all this material with the objective of presenting a rounded biography of a distinguished statistician, whose attractive personality shone out in everything that he did.

Acknowledgements

We are indebted to the following for permission to publish the material specified: Bertha Gosset, for extracts from the letters of her father W.S. Gosset, and for her comments on his outlook; Sarah Pearson, for commentaries taken from a typescript of her father E.S. Pearson, and for extracts from letters to Gosset from Karl Pearson and E.S. Pearson; Joan Fisher Box, for information given in some material towards a biography of Gosset; the Biometrika Trust, for material taken from papers in *Biometrika*; the University of Adelaide, which holds the copyright on the publications and papers of the late R.A. Fisher, for extracts from *Statistical methods for research workers* and his letters to Gosset; the Library of University College London, for extracts from the papers and correspondence of Karl Pearson, list numbers 525 and 704; the editors of *Annals of Human Genetics*, for the photograph of 'Student' in 1908; the American Statistical Association, for extracts from papers in their *Journal*; the Royal Statistical Society, for a quotation from their *Journal*, Series A; and the International Statistical Institute, for extracts from a paper in their *Bulletin*.

On behalf of E.S. Pearson, we thank Janet Abrahams for reproducing his earlier drafts, and finally turning out an excellent piece of craftsmanship; and we also thank the Finance Committee of University College London for in part supporting the expenses involved from a fund left to the College in the 1830s by his great-uncle Samuel Sharpe. E.S. Pearson's typescript, and the originals of his correspondence with Gosset, are in the possession of Sarah Pearson. On our own behalf, we are grateful to Anthony Watkinson, formerly of the Oxford University Press, whose comments made clear the need for editorial work; to Gitta Faulkner, for typing successive drafts of this book with unfailing cheerfulness; to Anders Hald, for information about Kirstine Smith; to Oscar B. Sheynin, for information about Oskar Anderson; and to the Libraries of the University of Newcastle upon Tyne and University College London, for their help when consulting manuscripts, special collections, and books kept elsewhere.

1. Introduction

Egon Sharpe Pearson has a secure place in any account of the development of statistical methodology during the twentieth century. Between 1925 and 1938, his collaboration with Jerzy Neyman established the Neyman–Pearson theory of testing hypotheses. The continuing importance of this feature of statistical inference owes much to his interests in the connection between theory and practice, which are also shown by his work on editing statistical tables. His enthusiasm for the use of quality control in industry led to the Royal Statistical Society forming an Industrial and Agricultural Research Section in 1933, and greatly assisted the introduction of control charts in wartime. He was Managing Editor of *Biometrika* from 1936 to 1966, in which role the subject of statistics was immeasurably helped by his conscientious editing and kindly advice to contributing authors. His many honours attest to the esteem in which he was widely held.

However, Egon Pearson was not only a distinguished statistician but also an outstanding historian of statistics. His historical work was informed by a lucid style free from polemics, imbued with a deep understanding of the flow of ideas, and supported by knowledge and experience of exceptional range. The historical aspect of his career began in 1938–9 when events led him to compile biographies of his father Karl Pearson and ‘Student’ (William Sealy Gosset). From 1955 onwards, a series of studies in the history of probability and statistics appeared in *Biometrika*, where Egon Pearson and others examined a remarkable diversity of topics. A collection of these studies, together with earlier historical papers by Karl Pearson, was published in 1970. After relinquishing the post of Editor of Auxiliary Publications for *Biometrika* in 1975, Egon Pearson produced an edited version of his father’s lectures given at University College London during the academic sessions 1921–33 on the history of statistics in the seventeenth and eighteenth centuries.

When he died in 1980, Egon Pearson had been gathering material for what he called his ‘*magnum opus*’. He started from the fact that he possessed some forty letters exchanged between Gosset and himself, and also about the same number written by Gosset to Karl Pearson. When putting the letters in order for typing, he was inevitably led to comment on and to link up many experiences. His correspondence with Gosset and related comments were collected under the title

THE GROWTH OF MODERN MATHEMATICAL STATISTICS
The part played by Student

He then had the idea of including much else of an autobiographical nature: summer vacations with his parents; memories of Yorkshire and Oxford farms around the turn of the century; sailing experiences with his Courtauld cousins in the 1920s; his fascination with Italian art, particularly black and white sculpture and church architecture; pencil drawing of the coast of Scotland; and an account of his five months' visit to the USA in 1931. He collected such information as he could about Gosset's family, the project grew with time, and he pondered long about what title to give 'this amorphous account which I think no publisher would accept'. As a temporary measure, he decided to call it

ALL THIS—AND STUDENT TOO

But he knew that at 84 the project might never be completed, and his account ends in April 1919.

We think that only Egon Pearson could have brought to a close the venture on which he had embarked, but also that nearly all the material which he had assembled is of great value for biographical purposes. Accordingly, we have been led to prepare an account of Gosset's life and statistical work which integrates what Egon Pearson published in 1939 with his '*magnum opus*' and other relevant material. We begin in Chapter 2 by describing the background to Gosset's statistical work, with particular emphasis on methods used in astronomy and geodesy for the combination of observations, and on the advances made in the Biometric School under the direction of Karl Pearson. Chapter 3 is an outline of Gosset's life based on Launce McMullen's account, with additions from other personal recollections and Gosset's letters. Chapter 4 is a much altered version of what Egon Pearson had written about his father's correspondence with Gosset, where changes have been made so as to identify and explain the principal topics considered. The letters between Gosset and Ronald A. Fisher (the Guinness collection) were privately circulated by Launce McMullen in 1962, and were used by Joan Fisher Box in her biography of Fisher. Chapter 5 presents and assesses the main statistical aspects of this correspondence.

Egon Pearson divided the letters which he exchanged with Gosset into six groups, and within each group they are arranged in chronological order, so that an individual letter can be identified by group and number. He supplied a historical introduction, a discussion for each group, brief or discursive comments on successive letters, closing remarks, and copies of all the letters. His method of approach for the letters to Karl Pearson, in which the classification is by subject matter rather than 'date' order, has been preferred as

giving a more readable version. Pearson had already associated some groups with defined topics, and we have extended this process within groups. Chapter 6 retains his general commentary and division into six groups, but his individual comments and the letters have been replaced by accounts of the main topics in each group illustrated by informative passages from the correspondence. Finally, we have added in Chapter 7 an appreciation of Gosset's achievements, and a discussion of his personal outlook and professional relationships.

Much else can be found in the rich source which Pearson has bequeathed to posterity. His autobiographical memories, although not as extensive as he had planned, will doubtless be valuable for a future biography. Constance Reid has already described the personal side of the Neyman-Pearson joint work in her biography of Neyman. The details of Gosset's family and colleagues, and the letters and notices written immediately after his death, add little to the 1939 biography. Most of Pearson's material concerns Gosset the statistician, and that is the topic on which we have chosen to concentrate.

2. Background

2.1 INTRODUCTION

Statistics is a scientific discipline concerned with the collection, analysis, and interpretation of data obtained from observation or experiment. The subject has a coherent structure based on the theory of probability, and includes many different procedures which contribute to research and development throughout the whole of science and technology. Statisticians engage in specialist activities worldwide, working for government, industry, higher education, and research institutions. They are organized in professional bodies, both national and international, and meet regularly to discuss progress, in groups which range in size from small seminars to large conferences. Their interests extend from theoretical studies which examine methodology to practical applications in a wide variety of fields. Textbooks which describe every aspect of the subject and journals which report on progress over the spectrum of activities have both steadily built up into large collections of scientific material.

The statistical world of a century ago was quite different. Assessment of numerical data had long been the core of the subject, but the data were almost exclusively concerned with social and economic questions. Statistical methodology consisted of a set of techniques which would now be regarded as fairly basic, but which nevertheless can still yield useful conclusions: counting and classification, averages and index numbers, maps and diagrams, descriptive analysis, and numerical laws. Some of the professional bodies which exist today had already been founded, but they were much smaller in size, their active members were usually government statisticians, and publications associated with them reflected a narrow range of official interests. The normal law of error and the method of least squares had been established for the analysis of data from astronomy and geodesy, and these techniques were collected into books on the combination of observations. But such activities were separate from the subject of statistics as generally perceived, and nothing was available from either source which could provide appropriate procedures for the statistical study of heredity and evolution. However, a small group of outstanding individuals, who worked on the solution of these practical problems, gradually emerged during the latter part of the nineteenth century. When at last the main streams of technical advance were united, there ensued a change in the scope and meaning of statistics which has continued ever since to enlarge and diversify the subject.

2.2 COMBINATION OF OBSERVATIONS

By the middle of the eighteenth century, enough was known about the theory of probability to permit the reconsideration of methods for the combination of observations in astronomy and geodesy. During the following seventy years, the methods were gradually endowed with a theoretical structure in which three principal features can be distinguished. They consisted of a procedure for calculating the probabilities of possible causes given an event, a probability distribution for the errors of observation, and an algorithm for estimating the unknown constants in linear relationships.

A theorem for calculating the probabilities of causes was proved by Bayes in 1764, with little impact on contemporary thought. Laplace stated a similar procedure in 1774, and his words were translated as follows by W. Stanley Jevons in *The principles of science*, published exactly one hundred years later.

If an event can be produced by any one of a certain number of different causes, all equally probable a priori, the probabilities of the existence of these causes as inferred from the event, are proportional to the probabilities of the event as derived from these causes.

Early in the nineteenth century, this came to be known as the inverse application of the theory of probability, or briefly as inverse probability, and would now be described as Bayes' theorem with a uniform prior distribution. The first major application of inverse probability was made in 1809 when Gauss discussed the combination of observations in his *Theoria motus corporum coelestium*. He proved that the normal law is the unique distribution for which the arithmetic mean of a random sample of observations is the value where the posterior distribution of the population mean attains a maximum. When there are several population means, all linear combinations of unknown constants, an assumption of normality for the errors, together with an application of inverse probability, led Gauss to conclude that the unknown constants are best estimated by the method of least squares. Another approach to the normal law of errors appeared in 1810, when Laplace showed that the sum of a large number of errors is approximately normally distributed. This extension of what de Moivre had proved for a binomial distribution was an intermediate form of what is now called the central limit theorem. In 1811, Laplace pursued this asymptotic approach by showing that the method of least squares could be justified without an application of inverse probability, because the mean absolute errors of the estimated constants are minimized in large samples when the method of least squares is used. Finally, Gauss established in 1823 that the method of least squares can be justified by a criterion of minimizing the squared errors of the estimators, without any reference to a normal distribution, and for any sample size.

This was virtually the end of the pioneering phase. The principle of inverse

probability had become a standard method of inference, and in the hands of Fourier and Cournot led to statements closely resembling those derived much later from the theory of confidence intervals. Although several laws of errors had passed under review, enough properties of the normal distribution were known to establish that formulation as clearly preferable. The method of least squares was agreed to be a convenient estimation procedure, and by 1826 was complete in many important particulars. These included the solution of the equations of estimation by triangular decomposition, the inversion of the matrix of the equations to provide measures of precision, and the estimation of the error variance from the sum of squared residuals.

From the middle of the eighteenth century until well into the nineteenth, probability in Britain was chiefly applied to life insurance, and those who wrote on the subject in English directed their attention towards practical questions concerned with the valuation of annuities and reversionary payments. The second edition of Laplace's *Théorie analytique des probabilités*, published in 1814, was accompanied by a long introductory *Essai*, which came as a revelation and stimulated great interest. Those who first responded to the demand for popular exposition and the challenge of rigorous interpretation had mostly been trained in mathematics at Trinity College, Cambridge, and were active in the Society for the Diffusion of Useful Knowledge, an organization which published cheap booklets to educate the working class. In 1830, the Society published an anonymous work *On probability*, known later to have been written by John William Lubbock and John Drinkwater (afterwards Bethune). This book represents a transitional phase in the exposition of the theory, where applications to life insurance, supported by tables of annuities, are discussed side by side with the sections of *Théorie analytique des probabilités* which originated in the eighteenth century. The method of least squares is described only briefly and there is nothing about laws of errors.

A much more influential figure in this transitional phase of probability in Britain was Augustus De Morgan,¹ a logician who was Professor of Mathematics at University College London and author of a very large number of articles on mathematical topics for encyclopaedias. The *Encyclopaedia metropolitana* published a treatise by De Morgan on *Theory of probabilities*, the original edition in 1837 and another in 1845. He firmly supported the use of inverse probability, gave a comprehensive review of *Théorie analytique des probabilités*, and included a derivation of the normal distribution on the lines adopted by Gauss in 1809. But De Morgan was not altogether successful in his presentation of least squares, where the method is justified on the assumption that both the observations and the coefficients in the linear relationships are equally variable. His more popular *Essay on probabilities* appeared in 1838; the first half effectively constitutes a statistical textbook at introductory level, and the remainder is concerned with what De Morgan regards as the most

common applications of the theory of probabilities, namely to life contingencies and insurance offices.

For reasons both mathematical and logical, these advances in the theory and applications of probability did not meet with unquestioning acceptance. Gauss had given two demonstrations of the method of least squares and Laplace a third, so that the relations between the different modes of presentation needed to be explained. Robert Leslie Ellis² scrutinized the demonstrations, reduced the mathematical difficulties of Laplace's analysis, and warmly approved Gauss's second demonstration, based on what are now called linear estimates with minimum variance. However, Laplace's explanation of the normal law of errors continued to pose severe problems. His line of approach could be sustained in terms of the accumulation of a large number of small errors, but ironically could only be made intelligible for most readers by a return to earlier arguments from a binomial distribution. This simplification was adopted by Adolphe Quetelet, but when John F.W. Herschel³ reviewed the poor English translation of *Lettres ... sur la théorie des probabilités* in 1850, he introduced another derivation of the normal law of errors. This began from an assumed statistical independence of errors in any two orthogonal directions, and since Herschel's prestige was great, his explanation of the law of errors became widely accepted, despite speedy and cogent objections from Ellis. Some twenty years later, the field was reviewed again by James W.L. Glaisher,⁴ who was stimulated by a claim that the American Robert Adrain had independently discovered the method of least squares in 1808. Glaisher's opinions generally supported those of Ellis, and his main contribution to the combination of observations lay in giving lectures at Cambridge which were described as critical, constructive, and comprehensive. Practitioners of the subject were confident that the normal law of errors was appropriate, because of empirical support provided by frequency distributions from sources as diverse as the chest sizes of Scottish soldiers and observations on Polaris made over five years.

The concept of inverse probability was challenged on logical grounds, especially by George Boole in his *Laws of thought*, published in 1854.

It has been said, that the principle involved in the above and in similar applications is that of the equal distribution of our knowledge, or rather of our ignorance—the assigning to different states of things of which we know nothing, and upon the very ground that we know nothing, equal degrees of probability. I apprehend, however, that this is an arbitrary method of procedure.

This is the origin of the phrase 'equal distribution of ignorance' which was to resound down the years. But Boole was cautious in opposing the views of his fellow logician De Morgan, who

of English writers, has most fully entered into the spirit and the methods of Laplace.

The question continued to receive careful attention, notably by John Venn, who raised strong objections to a famous consequence of inverse probability—the rule of succession—in his *Logic of chance* a dozen years later. However, when Francis Y. Edgeworth turned to statistical questions in the early 1880s, he was able to reconcile the differing views taken of inverse probability. His article on 'The philosophy of chance' in *Mind* for 1884 came to the following conclusion.

The preceding examples, especially the first, may show that the assumptions connected with 'Inverse Probability', far from being arbitrary, constitute a very good working hypothesis. They suggest that the particular species of inverse probability called the 'Rule of Succession' may not be so inane as Mr. Venn would have us believe.

Throughout the nineteenth century, the ideas of Laplace prevailed at the foundations of the theory of the combination of observations, so that the law of error and the method of least squares came to be inextricably linked. The justifications of least squares offered by Gauss were less influential: the first was derived from a disputed principle, while the second was regarded as a minor variation on Laplace which failed to show that the result was most probable. Among British mathematicians, Ellis was virtually alone in his warm approval of Gauss's second demonstration. However, Gauss gave practical expression to the use of inverse probability, and the methodology of least squares which he developed and used gained general acceptance in astronomy and geodesy. The subject achieved maturity in about twenty years, and thereafter refinements were steadily added, for example Benjamin Peirce's criterion for the rejection of doubtful observations and Peters' formula giving an estimate of the probable error in terms of the absolute values of the residuals. Meanwhile, statistics in Britain began to take a completely different direction.

2.3 BIOMETRY

The statistical study of heredity began with Francis Galton, a Victorian polymath. He was an extraordinary person, curious about everything, rich in ideas, and tireless in pursuing them. His interests were catholic: psychology, heredity, and anthropology are relevant here because statistical ideas were derived from them, but he is also remembered for contributions to meteorology, decimalization, fingerprints, photography, and exploration. The bibliography of Galton's published work fills almost fifteen printed pages. Much of his influence on statistics can be attributed to a fairly late statement of his ideas, the book *Natural inheritance*, published in 1889 when Galton was 67, and to his discovery of correlation which came immediately afterwards.

Among the first readers of *Natural inheritance* was Karl Pearson, another

Victorian polymath. He was at this time 32 and had for five years been Goldsmid Professor of Applied Mathematics and Mechanics at University College London. Whereas Galton's interests were essentially scientific, Pearson had already displayed a professional competence in matters far removed from mathematics, for example in philosophy, history, and literature. Along with Olive Schreiner, the famous novelist of Africa, and Maria Sharpe, whom he was later to marry, he was an active member of the 'Men's and Women's Club', which was concerned with improvement of the relations between men and women. To this society he read a paper on *Natural inheritance* and although his initial reaction was somewhat critical, his subsequent career soon became exclusively devoted to statistical matters. During the last decade of the nineteenth century, a stream of papers appeared, by himself or with collaborators, which put forward his work on frequency curves, correlation, and regression analysis in the guise of contributions to the theory of evolution.

The opening phase of Pearson's statistical career was influenced by two other readers of *Natural inheritance*: Francis Y. Edgeworth, who worked on the quantification of economic science, and W.F. Raphael Weldon, a biologist concerned to develop Darwin's theory of natural selection. Edgeworth, twelve years senior to Pearson, had trained himself to become an expert on statistical theory, and was inspired by Galton's work to publish a series of papers on correlation. His main result was expressed by Edgeworth's theorem on the multivariate normal distribution, an important element in some of Pearson's work during this period, and Edgeworth also gave sound advice to Pearson by means of correspondence. But Edgeworth had a reserved and complex personality, and was not the sort of man with whom others could have a close working relationship.

Weldon, three years younger than Pearson, joined the staff at University College London in 1891 as Jodrell Professor of Zoology. He was convinced that studies of animal and plant populations would provide support for Darwin's theory, and he turned to Pearson for help on the statistical problems involved. Pearson became deeply interested and a firm friendship developed, as a result of which the science of biometry was born, the Biometric School was established, and the journal *Biometrika* was founded in 1901. The Biometric School was committed to Galton's 'law of ancestral heredity', formulated in terms of regression, and became involved in controversy after Mendel's work on plant hybridization was rediscovered. William Bateson was the most notable in Britain of the many biologists who were attracted by the simplicity of Mendel's hypothesis. But the possibility of cooperation between the Biometric and Mendelian Schools, never strong, was removed by Weldon's sudden death in April 1906. It had been the custom of the Weldon and Pearson families to take their Easter holidays together so that Weldon and Pearson, along with helpers such as Alice Lee, could push forward their

researches. Weldon went to London for dental treatment, expecting to return in a few days; but instead there came a telegram saying he was dead. In the opinion of the younger Pearson, his father never fully recovered from the shock.

During the early 1890s, Pearson gave public lectures on graphical statistics and the laws of chance, with as much experimental evidence and as little theory as possible. His first course on the theory of statistics was given at University College London for two hours per week in the session 1894–5, and there were just two students: Alice Lee and George Udny Yule. After that session, the statistical course became annual, and the group led by Pearson was gradually enlarged. In 1903, he obtained the first of a long series of annual grants from the Worshipful Company of Drapers which enabled him to fund and continue a Biometric Laboratory. Training and research in mathematical statistics, and the computation of new mathematical tables were associated with this laboratory. After Weldon's death, Pearson's chief interest and research programme shifted from pure biometry to eugenics, and a gift from Galton led to the foundation of a Eugenics Laboratory in 1907. The many memoirs and lecture series on social and eugenic problems issued before the First World War and to some extent afterwards were based on research work carried out in the second laboratory. These laboratories together formed the base of Pearson's Biometric School, and were combined in 1911 when the Department of Applied Statistics was established. Between 1894 and 1930, University College London was the only place in the UK for advanced teaching in statistics, and from 1906 to 1936 Pearson was the sole editor of *Biometrika*, which for most of that period was the main outlet for research work on statistical theory. The reputation of the Biometric School for teaching and research is the reason why Gosset made contact with Pearson when he felt the need for statistical advice soon after the beginning of his career, and why virtually all his papers were published in *Biometrika*.

2.4 BOOKS USED BY GOSSET

When writing to Egon Pearson on 11 May 1936, Gosset recalled the time before his first visit to Karl Pearson.

I had learned what I knew about errors of observations from Airy and was anxious to know what allowance was to be made for the fact that a 'modulus' derived from a few observations was itself subject to error. . . .

The first English textbook on the combination of observations was written by Sir George Biddell Airy, a member of the Victorian scientific establishment. He was chiefly notable for bringing order to British astronomy during the

nineteenth century, and his outlook was strongly practical. His book *On the algebraical and numerical theory of errors of observations and the combination of observations* was first published in 1861, and the third and last edition of 1879 is a duodecimo volume of some 120 pages.

No novelty, I believe, of fundamental character, will be found in these pages ... the work has been written without reference to, or distinct recollection of, any other treatise (excepting only Laplace's *Théorie des probabilités*); ...

There are four parts to the book. Part I concerns the law of probability of errors and various associated constants, such as the probable error and modulus (i.e. $\sqrt{2}$ times the standard deviation). The law was justified in the first edition by an argument which occupies eight of the 103 pages and closely follows Laplace's version of the central limit theorem. Airy later substituted an alternative proof based on Herschel's reasoning, because Laplace's 'final steps are very obscure and difficult'. Values of the mean error and probable error of linear combinations are derived in Part II, and these quantities are 'determined' from a series of n observations. In particular,

Probable error of the mean

$$= 0.6745 [\text{Sum of squares of apparent errors}/n(n-1)]^{1/2},$$

where the apparent error is the difference between an actual error and the mean of the actual errors. Part III is a treatment of the method of least squares for up to three unknown parameters. Airy translates 'moindres carrés' as 'minimum squares',⁵ a term seldom used after 1825, and Laplace's influence is evident in the way that the method is established. The final part is entitled 'On mixed errors of different classes, and constant errors'. Airy gives an example of a series of observations made day after day affected by a daily 'constant error' which follows a different law from the ordinary error. He describes at length how to calculate the 'mean discordance of each day's result' and 'probable error of mean discordance'. The two quantities are compared 'and now it will rest entirely in the judgement of the computer' whether or not a constant error exists.

The reviewer of *Errors of observations in the American Journal of Science and Arts*⁶ for November 1861 affected to find that Airy's method of choosing the linear combination with minimum probable error differed from the method of least squares, used by most men of science. In so far as a judgement by contemporary standards can be made more than a century after publication, Airy's book appears to be a sound introduction to the analysis of observational data, with a good balance of topics clearly described. There is, however, too much description. The theory is illustrated by applications to problems in astronomy and geodesy, but the results are always expressed in algebraic terms. There are no numerical examples, except for an appendix added in the second edition, where the theoretical law for the frequency of errors is verified

by a comparison of observed and expected frequencies, in order to establish 'the validity of every investigation in this treatise'.

Gosset also used books by Merriman and Lupton. *A textbook on the method of least squares* by Mansfield Merriman was published in 1884, and ran to eight editions, the last completed in 1911. Merriman was an American civil engineer who explored the history of methods for combining observations. His book reflects these interests and was evidently recognized as a useful exposition of theory and practice. Two deductions are given for the law of probability of error, the precision of quantities calculated from the observations is judged by probable errors, and the uncertainty of the probable error is considered. The method of least squares is explained in detail, using many examples from surveying. Topics such as propagation of error, rejection of doubtful observations, social statistics, and Galton's median, all make brief appearances. The section on 'Uncertainty of the probable error' outlines a paper published by Gauss in 1816.

Gauss presented in 1809 an exposition of least squares based on the law of error of observations defined by the formula

$$(h/\pi^{1/2}) \exp(-h^2 d^2),$$

where d is an error and h is the measure of precision. In modern terms, the distribution is normal with mean zero and variance $1/2h^2$. Gauss proceeded in 1816 to discuss the 'determination' of h from a sample of m errors d_1, d_2, \dots, d_m . He assumed a uniform prior distribution for h and showed that the posterior distribution is then normal, with mean $H = (m/2\sum_{i=1}^m d_i^2)^{1/2}$ and variance $H^2/2m$. For a standard normal distribution, the probability is $\frac{1}{2}$ that the variate lies between $\pm\rho\sqrt{2}$, where $\rho = 0.4769363$. Gauss inferred that the odds are one to one for the true value of h to lie inside the interval $H(1 \pm \rho/m^{1/2})$, and that the corresponding interval for the probable error r is $R/(1 \pm \rho/m^{1/2})$, where $R = \rho/H$. He then extended his investigation to consider the interval estimation of r from the sum of the n th powers of the errors taken positively, and gave a table of the probable limits of r for $n = 1, 2, \dots, 6$ when m is large. Gauss denoted the corresponding formulae by I, II, \dots , VI, respectively, and remarked that formula II is the most advantageous: 100 errors of observation treated by this method give as reliable a result as 114 errors by formula I, 109 errors by formula III, \dots , and 251 errors by formula VI. But formula I is easy to calculate and only slightly inferior to formula II. The formulae corresponding to I and II using residuals instead of errors are associated with the names of Peters and Bessel, respectively.

Notes on observations by Sydney Lupton was published in 1898. This is a slim volume containing twenty-two chapters, most of them only a few pages in length. The subtitle describes the contents as 'an outline of the methods used for determining the meaning and value of quantitative observations and experiments in physics and chemistry, and for reducing the results obtained'.

Less than half of the material is statistical in nature, and the treatment is mostly superficial by the standards of 1898. However, there are over forty references to assist the reader, and this guidance would have been the main virtue of the book for 'those who desire to pursue the subject more thoroughly'. The list of writers who are cited includes Laplace, Bertrand, Galton, Edgeworth, and Karl Pearson, as well as Airy, Chauvenet, and Merriman.

2.5 KARL PEARSON'S LECTURES

Glaisher took an active interest in the combination of observations from 1870 to 1880, and lectured on the topic at Cambridge. But the only relevant lectures attended by Gosset were those given by Karl Pearson. By the end of 1891, Pearson was conversant with the English school of political arithmetic, the German school of state science, and the French school of probability, the last represented by Déparcieux, Laplace, and Quetelet. The contents of his popular lectures at Gresham College on 'The geometry of statistics and the laws of chance' reflect his changing outlook on statistics between 1891 and 1894. He began by describing all kinds of diagrams, proceeded to an elementary account of the theory of probability, and finally gave a treatment of questions arising from biometry. In particular, he discussed the problem of a priori probabilities with reference to the principle of 'the equal distribution of ignorance' and the contribution of Bayes, Laplace, Boole, Venn, and Edgeworth. At least two of the lectures were concerned with 'normal curves', but, as the title of the course suggests, there was no place for the method of least squares. The books and articles to be consulted were described as follows in November 1892.

For those of you who may have time for reading I would strongly recommend a comparison of Chaps. X–XII of Stanley Jevons's *Principles of science* with Chaps. VI–XI of Dr Venn's *Logic of chance* and Prof. Edgeworth's *Philosophy of chance* published in *Mind* for 1884. . . . While dealing with the subject of books I may also refer to:

- De Morgan: *Formal logic* (1857). Here Chaps. IX–XI are closely connected with the topics of our first two lectures.
- De Morgan: *An essay on probabilities* (1838). This is still a useful and suggestive little book, although it requires some mathematical knowledge.
- Whitworth: *Choice and chance* (3rd ed. 1878). An excellent book with which to approach the elements of the mathematical theory.
- Westergaard: *Die Grundzüge der Theorie der Statistik* (1890). By far the best textbook on the relation of *statistics* and probability for those *who read German*.

Rapid developments in the theory of statistics ensured that lecture courses on the subject changed considerably over the years. But Pearson's emphasis

after 1893 was always on techniques likely to be useful for the study of heredity and evolution, in preference to those for other applications. The material which he presented in his university courses is known from lecture notes taken by Yule for the sessions 1894–5 and 1895–6, Gosset for the session 1906–7, Isserlis for 1913, and Egon Pearson for the first-year and second-year courses of the session 1920–1. Yule's comments over forty years later confirm Pearson's awareness of the continental direction in statistics.

The first course opened with a brief outline sketch of history, leading up to a 'Kollektivmass' definition of statistics. Among the works bearing on theory to which we were referred those of Zeuner, Lexis, Edgeworth, Westergaard and Levasseur might be expected: but would any other lecturer have thought of suggesting the study of Marey's *La méthode graphique dans les sciences expérimentales* (1878, 1885)? Karl Pearson was an enthusiast for graphic representation and thought in graphic terms.

An answer to Yule's question is that Edgeworth listed Marey's book among the references given for the second of his Newmarch lectures at University College London in 1892.

Some features remain throughout all the syllabuses: Bayes' theorem with an 'equal distribution of ignorance', the Pearson system of frequency curves, binomial and normal distributions, correlation, and multivariate normality. Topics at first excluded or overlooked are subsequently inserted: the method of least squares, polynomial regression, the Poisson distribution. Otherwise the contents change with the progress of research or rediscovery, mainly in the Biometric School, and to a lesser extent elsewhere: contingency tables, the chi-squared test for goodness of fit, the variate difference correlation method, the distribution of the standard deviation in normal samples. The *t*-test only appeared after 1921, perhaps in response to submissions from younger members of staff.

3. William Sealy Gosset

3.1 INTRODUCTION

The Gossets were an old Huguenot family who left France at the Revocation of the Edict of Nantes in 1685, and became middle class with army and clerical traditions. Among them was Frederic Gosset, a Colonel in the Royal Engineers, who in 1875 married Agnes Sealy Vidal. There were five children, the eldest being William Sealy Gosset, born on 13 June 1876 in Canterbury. He was bright enough to win scholarships, which must have provided welcome additions to the funding of his education, since the family had to live frugally. Gosset was a Scholar of Winchester College from 1889 to 1895. The school had been founded in 1382 by William of Wykeham, and towards the end of the nineteenth century was with other public schools beginning to go through an intellectual renaissance. Unable to follow his father into the Royal Engineers because of poor eyesight, Gosset took up a scholarship at New College, Oxford, another Wykehamite foundation. Here he obtained a First in Mathematical Moderations in 1897 and left in 1899 with a First Class degree in Chemistry.

In October 1899, Gosset became a Brewer with Arthur Guinness, Son & Co. Ltd, manufacturers of stout at the St James's Gate Brewery in Dublin, and he remained with the firm for the whole of his working life. Guinness had recently decided to introduce scientific methods into brewing, and the new policy was effected by appointing men with First Class science degrees at Oxford and Cambridge to positions in junior management. The next Brewer to be appointed after Gosset was Geoffrey S. Phillipotts, and the two young men became close friends through their common interest in outdoor pursuits, being keen on walking in the Wicklow mountains south of Dublin and yachting along the coast between Carlingford and Wexford. As a result of this friendship, Gosset met his future wife Marjory Surtees Phillipotts, sister to Geoffrey, and the couple were married on 16 January 1906 in Tunbridge Wells. They lived first in Dublin, and then rented a furnished house in Wimbledon from September 1906 to the spring of 1907 while Gosset was attending lectures and tutorials given by Karl Pearson at University College London. A son was born during this period, two daughters followed, and after 1913 the family lived in a house with a large garden at Blackrock on Dublin Bay.

3.2 EARLY STATISTICAL CAREER

Gosset was a practical man, and his statistical methods were developed in response to the needs of the brewery. Research work at Guinness on the selection, cultivation, and treatment of barley and hops led to an accumulation of data affected by variability of materials, susceptibility to temperature change, and necessarily short series of experiments. The problems of interpretation posed by small samples in which the measurements were not independent soon became apparent, and the Brewers found that Gosset was ready to help in analysing the data. At this stage, his advice was based on consulting books concerned with the theory of errors, in particular Airy's *Theory of errors of observations* and Merriman's *Method of least squares*. Gosset's advisory work formed the background to his report to the Board of Guinness in November 1904 on 'The application of the "law of error" to the work of the brewery'. Here, he set out the case for using statistical methods, discussed the error curve,¹ applied results on the addition or subtraction of random variables, and by comparing the difference between $\Sigma(A+B)^2$ and $\Sigma(A-B)^2$ showed some awareness of the consequences of correlation. The summary of the report included a suggestion that questions concerning 'the degree of probability to be accepted as proving various propositions' should be referred to a mathematician. Soon afterwards, the introduction of Gosset to Pearson was effected through a letter from Vernon Harcourt, an Oxford chemistry lecturer, and arrangements were made for Gosset and Pearson to meet.

The meeting took place on 12 July 1905, when Pearson was spending his long vacation in a house at East Ilsley in Berkshire, which he had rented from July to September to be within cycling distance of Weldon in Oxford. Gosset was on holiday in the house to which his father had retired at Watlington in Oxfordshire, about twenty miles from East Ilsley, and he cycled over to discuss a list of questions concerned with the cost effectiveness of experiments, limits of error from repeated measurements, how to establish a relationship between sets of observations, and what books would be useful. Pearson's notes for the interview include formulae for the variance of $A \pm B$ and the probable error of a correlation coefficient, and references to papers on the theory of statistics. After Pearson's death in 1936, Gosset wrote² about this interview to Egon Pearson.

He was able in about half an hour to put me in the way of learning the practice of nearly all the methods then in use before I came to London a year later.

Immediate consequences of the visit were a supplement to the brewery report of 1904, and a second report (1905) on correlation, both of which reflect Pearson's influence. However, the report on correlation notes that, by contrast with the Biometric School, the brewery was working with small

samples, and Gosset came close to appreciating the dangers of using large-sample methods without adjustment.

Guinness had a policy of sending staff away for specialized study, and Gosset derived benefit from two such visits. He attended a course on brown beers³ at Birmingham University, where he learned how to use a haemocytometer and began to study the question of how many squares should be counted in order to estimate the concentration of yeast cells with sufficient accuracy. Later he spent the first two terms of the session 1906–7 in close contact with the Biometric Laboratory, another consequence of the meeting with Pearson, and he recalled his time there after Pearson's death.

I am bound to say that I did not learn very much from his lectures; I never did from anyone's and my mathematics were inadequate for the task. On the other hand I gained a lot from his 'rounds': I remember in particular his supplying the missing link in the error of the mean paper—a paper for which he disclaimed any responsibility.

For most of this period, Gosset worked on pieces of research in which he derived and applied various sampling distributions, and his wife helped by copying out measurements and tables. His first published paper, in which counts of yeast cells are compared with the exponential limit of the binomial distribution, had been in preparation before visiting Pearson and appeared in the February 1907 issue of *Biometrika*. Publication had been agreed by Guinness provided that a pseudonym was used, and that none of their data appeared. This obsession with secrecy was no unusual whim; secrecy was and is widely practised in industrial and commercial circles in the UK. The ban forbidding staff to use their own name was not lifted until just before the Second World War. Thus, Gosset became 'Student' and later colleagues were similarly cloaked in pseudonymity.

3.3 CORRESPONDENCE

Gosset had many correspondents, and much of his spare time was devoted to writing letters, mainly to experimenters.⁴ This book is concerned with his contributions to the advance of statistical methods for industry and agriculture, which are greatly clarified by his correspondence with Karl Pearson, R.A. Fisher, and Egon S. Pearson. These letters enable Gosset's papers to be connected with the gradual development of his ideas, permit the establishment of relationships within his published work, and show how he both influenced and was influenced by the work of other statisticians.

There are about forty letters from Gosset to Karl Pearson, but unfortunately only three in the reverse direction are preserved. Launce McMullen, who succeeded Gosset as head of the statistical section at the St. James's Gate Brewery, suggested that around 1934, when the Gossets were leaving Ireland for England, there was a 'holocaust' of much valuable material. This included

letters from Pearson to Gosset, as well as the typewritten draft of three or four chapters of an elementary textbook on the use of statistical methods in experimentation. The loss of this material is, historically, a great pity. Gosset's letters show that here was a penetrating and objective critic, who, when his interests were aroused, was prepared to explore scientific fields unconnected with his duties as a Guinness Brewer. It would have been illuminating to read Pearson's reaction and see how far he accepted certain very mildly expressed criticisms.

In round figures, there are 150 letters from Gosset to Fisher, but only fifty in the reverse direction. Thus, about a hundred of the letters from Fisher to Gosset have been destroyed. Although the loss is much to be regretted, the misfortune is not so great as with the letters from Karl Pearson to Gosset, because other considerations apply. The letters from Fisher which survive give a sufficient indication of the terms on which the correspondence was conducted, and much can be inferred about the questions to which Gosset replied. Moreover, Fisher made a summary of the letters, perhaps when writing his obituary notice of Student published in 1939. His brief comments are sometimes informative, but they often overlook important issues and must be considered less reliable when in conflict with the letters themselves.

The letters from Egon S. Pearson to Gosset were filed at the brewery, probably because many were shown to his assistant Edward Somerfield with a request for comment. Consequently their correspondence survives largely intact, and it is the subject of Chapter 6.

3.4 PROFESSIONAL ACTIVITIES

Gosset reported to Karl Pearson on 24 October 1907 that he had been definitely put in charge of the Experimental Brewery, a job he was temporarily at before, and that this involved him in a certain amount of statistical work. Pearson realized Gosset's potentiality and in 1907 offered to look out for a post which would do justice to a man of his training and ability. But Gosset replied on 11 September 1909 that for a man with a growing family the pay would have to be very good for him to be able to afford to give up his present job.

Please do not consider yourself bound to look out for anything for me: I am very far from being dissatisfied [*sic*] with my present billet (after all £800 per year jobs are none too common) but it is easy to imagine some better use being made of my time.

The average annual salary of a British university professor in 1910–11 was £600, while those for a lecturer and assistant lecturer were £250 and £150, respectively.⁵ In any case, it was the experimental problems arising at the brewery which gave Gosset—mathematician, chemist, now with some statistical knowledge—the appropriate field in which to exercise his talents.

While in the course of discovering the pitfalls of experimental sampling and how to avoid them, Gosset must have mentioned to Pearson, probably on a visit, how he thought he should write a textbook on the subject. The book appears to have been planned on the basis of joint trials arranged with the maltster Edwin S. Beaven, who was on commission to Guinness. Gosset returned to the topic in his letter of 8 December 1910.

Re the text book for experimenters idea I am inclined to write an introduction, a contents bill & a couple of chapters & forward them to you (typed!) for your opinion as to whether the book is worth writing. It is certainly wanted, but by so few people, I fear, that one could hardly expect a publisher. I expect that I am rather favourably situated from the practical point of view, a greater *number* of correlation coefficients (of sorts) etc. etc. must pass through my hands than through those of anyone else in the wide world even including yourself. But it is mainly a question of time, & that is lacking to all of us.

He began his next letter on 6 February 1911 with comments on 'the great alcoholic controversy', in which Pearson was then engaged for reasons explained below in §4.5. The letter ends with brief comments on progress.

The text book is progressing, slowly as I am not at present on duties which leave me with a typewriter during meal hours, but I hope to send the first three chapters in a month or two.

Do not trouble to reply to this. You must have quite enough to do to keep up your correspondence with the teetotallers without having brewers on your hands too!

This seems to be the last reference to the textbook, and the unfinished typescript may have been destroyed in 1934. If Pearson had realized the importance of the draft, he would have encouraged its completion, perhaps publishing it among the *Drapers' Company Research Memoirs*. But perhaps the rather secretive Board of Directors of Guinness did not or would not allow such publication. The forward-looking Managing Director La Touche who had been largely responsible for the decision to appoint scientifically trained brewers died rather suddenly in 1914.

When war began in 1914, Gosset was quick to offer his services to Karl Pearson. He wrote on 26 August to say that he could spare about three hours per day and Sunday for any checking or computation work, and could probably borrow a Brunsviga from the brewery without much trouble. By 1 September of the following year, he seemed resigned to the contribution he could best make to the war effort.

My own war work is obviously to brew Guinness stout in such a way as to waste as little labour and material as possible, and I am hoping to help to do something fairly creditable in that way. All the same I wish government would double the tax again, it's such an obvious waste of pig food now!

During much of 1916 the Biometric Laboratory was working on torsional strain in the blades of aeroplane propellers, and on 7 May Gosset had evidently been discussing the matter with Pearson.

It occurred to me that if your difficulty with the aeroplane propeller is that the blade as a whole turns more at high speed than at low, both being required, some system of sliding the blade in a suitably rifled socket attaching to the hub by a spring might work. Then centrifugal force might be made to twist the blade as a whole forward when the increased resistance was twisting the working part back so as to keep the angle of the working part constant.

This is a long letter, the statistical aspects of which are considered in §4.5. Gosset concludes with a brief expression of disappointment at his rejection from the armed forces.

The doctors at the War Office thought me too short-sighted to serve.

In 1922, Gosset acquired his first regular statistical assistant Edward M. Somerfield, and correspondence with Fisher in November is concerned with arrangements for Somerfield to be a voluntary research worker at Rothamsted for three months. The financial side is set out in Gosset's letter of 15 November.

I quite understand that there is no sort of charge or fee, but it seems to us that if a firm of our standing sends a man for educational purposes we should pay for it. That being so we thought an unofficial hint from you as to the amount and the particular fund to which we could pay it would be a help. Would you like to suggest £25? If you think this too little I'll try to get some more. I don't know the Board's ideas and can't guarantee even £25, but I should hope to get it.

Similar arrangements were made in October 1930 for Somerfield's assistant A.L. Murray to visit Rothamsted for about six months, and the firm was then prepared to give a donation of £50 for services rendered.

Gosset's membership of statistical societies was influenced by the Pearsons, father and son. When Karl Pearson reorganized the programme of the Department of Applied Statistics after the First World War, he founded the Society of Biometricians and Mathematical Statisticians,⁶ a kind of seminar largely devoted to discussions of departmental research. As is general nowadays, outside authorities were asked to contribute. Gosset read papers on 13 December 1920 and 28 May 1923, concerned respectively with Spearman's correlation coefficients and testing varieties of cereals, both later published in *Biometrika*. The society was dissolved in 1927, but Egon Pearson's activities in the next few years led the Royal Statistical Society in 1933 to form an Industrial and Agricultural Research Section. These were the fields with which Gosset was particularly associated; he was elected to the Society in 1934, and contributed to meetings of the Section.

In October 1934, Gosset began to give all his attention to the new Guinness Brewery at Park Royal in north-west London, and at the end of 1935 he left

Dublin to take up his appointment as Head Brewer, in charge of the more scientific part of the production. He died from a heart attack on 16 October 1937, aged 61.

3.5 CHARACTER AND PERSONALITY

Gosset was a modest, kindly, and tolerant man,⁷ who disliked controversy and was absolutely devoid of malice. His three children were brought up rather frugally, as he had been, and he detested extravagance. But he did not mind spending money on education, and was often generous to those in need. His knowledge of world affairs was considerable, and the range of his domestic and sporting interests was exceptionally wide. He was a keen fruit-grower who specialized in pears, and an able gardener who enjoyed experimenting. At one time, he crossed a raspberry with a loganberry and tried to market the results—but the fruit did not catch on. His enthusiasm for walking, cycling, fishing, skating, and skiing began when he was still at school. He was a good carpenter and built a number of boats, including a collapsible wooden punt for fishing on lakes. An article in the *Field* of 28 March 1936 describes his boat for fly-fishermen, equipped with a rudder at each end by means of which the direction and speed of drift could be adjusted. He was a sound though not spectacular shot. His thigh was broken when his Model T Ford, known as the 'flying bedstead', overturned in July 1934. Before the accident, he was a regular golfer, using a remarkable collection of old clubs dating at least from the beginning of the century. After the accident, he took up bowls with great keenness. An appropriate last glimpse of Gosset is the photograph taken in April 1936 which shows him clad in tweed jacket, shorts, and boots, wearing a rucksack, carrying a walking stick and a fishing rod, and gazing across a snowy Dartmoor.

His daughter Bertha Gosset provided Egon Pearson with early personal memories.

Like many of his contemporaries he was devoted to Gilbert and Sullivan operas and never failed to attend the performances of the D'Oyley Carte Company when he had a chance—e.g. in Oxford and later in Dublin. He played the penny whistle and after he was married sang us songs as children. It was a great joy to him when we had our first wireless, made by my brother Harry about 1922. He constantly listened to good music afterwards for the rest of his life and developed a knowledge and appreciation, especially of Beethoven, his favourite composer, which we all shared.

The only 'don' I know he admired tremendously was Dr Spooner of New College; lately I read his *Life* and realised that Dad was there just before Dr Spooner was unanimously elected to be the Warden—obviously greatly respected by everyone. We had a reproduction of his portrait in our sitting room—I feel sure Dad must have been influenced by him because the two men had much in common, especially a deep integrity, wide interests, humility, and a capacity for taking infinite trouble. In fact I

enjoyed reading his *Life* mainly because he reminded me so much, in essentials, of my father.

I find it difficult to answer about religion, for my father was very reticent; he supported my mother who took us to church; and attended himself when we were young; but not later, explaining his absence by saying he had attended so often when young (at Winchester?). He was extremely careful never to say anything which might undermine our faith, and we all grew up as practising Christians.

4. Karl Pearson

4.1 INTRODUCTION

Nearly all Gosset's papers originated in problems arising from work at the St James's Gate Brewery on the manufacture of stout and the production of barley, so that his attention was directed to statistical methods for both industry and agriculture. During the period from 1906 to 1919, he was in regular contact with Karl Pearson (hereafter also referred to by his initials) by both personal meetings and correspondence, with the result that the questions which attracted his interest were strongly connected with activities of the Biometric School. The period accounts for about half of Gosset's published research, and for practically all his work on correlation, time series, and discrete distributions. Gosset's letters to K.P. contain not only the first signs of future papers, often years before publication, but also his comments and criticisms on a variety of matters which could be expected to interest his old teacher.

4.2 CORRELATION

When Galton made known the idea of correlation, he also supplied a graphical procedure based on medians and probable errors to derive a sample value of the correlation coefficient. His method was reconsidered by Edgeworth, who moved towards the definition now generally accepted, by how much is not altogether clear. Much of the theory of multivariate normal correlation was developed in 1896 by K.P., and in particular he showed that 'the best value' r of the correlation coefficient in samples from a normal bivariate distribution is given by the product-moment formula. He also obtained an incorrect expression for the standard error of r , but this was corrected when his joint paper with Louis Napoleon G. Filon appeared in 1898. At about this time, Sheppard¹ published his method of estimating the correlation coefficient by doubly classifying at the medians, and Yule established the connection between the theory of correlation and the method of least squares. Such were the principal known results concerning normal bivariate correlation when Gosset observed in his report of 1904 that the rule for the probable error of sums or differences sometimes failed in practice.

The results of Pearson and Filon seem to have been derived from an asymptotic posterior distribution of the correlation coefficient based on

uniform prior distributions for all the parameters. This standpoint of inverse probability is reflected in the main objective of Gosset's paper 'Probable error of a correlation coefficient', forwarded to Pearson on 24 October 1907, and published in the issue of *Biometrika* for September 1908.

We require the probability that R for the population from which the sample is drawn shall be between any given limits.

It is clear that in order to solve this problem we must know two things: (1) the distribution of values of r derived from samples of a population which has a given R , and (2) the a priori probability that R for the population lies between any given limits. Now (2) can hardly ever be known, so that some arbitrary assumption must in general be made; when we know (1) it will be time enough to discuss what will be the best assumption to make, but meanwhile I may suggest two more or less obvious distributions. The first is that any value is equally likely between $+1$ and -1 , and the second that the probability that x is the value is proportional to $1 - x^2$: this I think is more in accordance with ordinary experience: the distribution of a priori probability would then be expressed by the equation $y = \frac{3}{4}(1 - x^2)$.

Gosset began with samples of 4, 8, and 30 from a large body of data in which the true value R of the correlation coefficient was known to be 0.66; and using the same data, with x values taken from one sample and y values taken from another, to economize arithmetic, he obtained sample values r for the case when R is zero. These, together with mathematical considerations for samples of 2, led him to guess the Pearson type II curve with equation

$$y = y_0(1 - x^2)^{(n-4)/2}$$

for which he calculated the moments. Finding these to be in broad agreement with his sample moments in the case $R = 0$, he concluded that his equation 'probably represents the theoretical distribution of r when samples of n are drawn from a normally distributed population with no correlation'. However, he was unable to suggest an equation when there is correlation, and gave reasons for saying that the distribution could not be represented by a Pearson curve unless $R = 0$. Since the solution for (1) was obtained only for $R = 0$, no posterior distribution for R could be given at this stage.

Gosset had already used similar methods in his paper on 'The probable error of a mean', published in *Biometrika* for March 1908, to suggest an equation for the distribution of the sample standard deviation. He now supplied the impulse which led Pearson to draft the lines of an investigation by Herbert E. Soper² 'On the probable error of the correlation coefficient to a second approximation', which appeared in *Biometrika* for March 1913. Earlier approximations for the mean and standard deviation of r were improved by the use of asymptotic expansions, theoretical values were compared with observed, and Soper 'hoped that further experiments would be carried out which will ... show definitely ... whether the application of the standard types of frequency curves to the distributions of statistical constants in small

samples is justified'. This was the work which drew Fisher's attention to the problem. His paper 'Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population' [CP4]* was sent to Pearson in September 1914, and published in the issue of *Biometrika* for May 1915. Fisher began with appreciative comments on the work of Gosset and Soper, in the course of which he confirmed that Gosset's predicted distributions of the standard deviation and of the correlation coefficient when $R=0$ were indeed correct, and he found that Soper's expressions were most accurate for large samples when the exact formulae become most complicated. The distribution of the correlation coefficient was then obtained by a brilliant piece of geometrical reasoning, and Fisher proceeded to use his 'absolute criterion' to derive an approximation for what he later termed the maximum-likelihood estimate of the correlation coefficient, describing the estimate as 'the most probable value of the correlation of the whole population'.

Gosset acknowledged the receipt of an offprint on 15 September 1915, and went on to say that the question of a posterior distribution was still open. He perhaps intended to write $1-x^2$, and not $1-x$, in his suggested form of prior distribution.

I am very glad that my problem is a step nearer solution. (I never really liked Soper's approximation though of course it was colossal) but there still remains the determination of the probability curve giving the probability of the real value (for infinity population) when a sample of x has given r . Of course, this would have to be worked out for two or three a priori probabilities and if otherwise convenient I would try $y=y_0(1-x)^{(m-4)/2}$ (giving m the values 3, 4 and 6 in succession) as the a priori distribution of the probability of x being the real value of r .

The letter continues with a description of the background to his work on correlation.

I don't know if it would interest you to hear how these things came to be of importance to me but it happened that I was mixed up with a lot of large scale experiments partly agriculture but chiefly in an Experimental Brewery. The agricultural (and indeed almost any) Experiments naturally required a solution of the mean/S.D. problem and the Experimental Brewery which concerns such things as the connection between analysis of malt or hops, and the behaviour of the beer, and which takes a day to each unit of the experiment, thus limiting the numbers, demanded an answer to such questions as 'If with a small number of cases I get a value r what is the probability that there is really a positive correlation of greater value than (say) .25?'

When Pearson wrote to Fisher on 26 September 1914 provisionally accepting CP4, he indicated that he would 'like to see your paper extended with graphs of some of the curves, and tracing as n increases the change of the

*For explanation of [CP4] etc., see p. 125.

frequency form towards a normal distribution', and on 30 January 1915 he declared his intention, although 'quite aware that this will be very laborious', to tabulate the ordinates of the frequency curves for r , 'as soon as opportunity offers ... unless you want to do them yourself'. There were now further reasons for Pearson to explore the connection between small-sample distributions and the large-sample expressions of the Biometric School. The first results were presented in an editorial 'On the distribution of the standard deviation of small samples: Appendix I to papers by "Student" and R.A. Fisher', which immediately followed Fisher's paper in the same issue of *Biometrika*. Here, Pearson examined the rapidity with which the distribution of the sample standard deviation approached normality, and he concluded that the theory of probable errors could safely be applied for $n \geq 25$. He stated that the most reasonable value for σ , the population standard deviation, is obtained when the observed value of s , the sample standard deviation as defined on p. 46 below, is the mode of the frequency curve, whence

$$\hat{\sigma} = s[n/(n-2)]^{1/2}.$$

According to a footnote in the 'Cooperative Study' described below, Gosset pointed out that the best value of σ is obtained by maximizing the frequency curve with respect to variation in σ , whence

$$\tilde{\sigma} = s[n/(n-1)]^{1/2}.$$

His proposal was evidently based on taking the mode of the posterior distribution of σ when the prior distribution is uniform, but Pearson regarded the assumption of a uniform prior as not 'in accordance with experience'. Gosset's letter to Pearson on 1 September 1915 shows that they had continued to differ on this question.

But if I didn't fear to waste your time I'd fight you on the a priori probability & give you choice of weapons!

But I don't think the move is with me; I put my case on paper last time I wrote & doubt I've much to add to it. It was roughly this:

If $y = \varphi(x)$ be the distribution of the a priori probability that x is the S.D. then (I am in bed remember) *you* can easily deduce the mode of the a posteriori [sic] distribution when s has been given in one trial & in fact I wrote down what I supposed to be the value though I forget it for the present.

But it all depends on $\varphi(x)$ & my feeling is that $y = \varphi(x)$ is generally a pretty flat curve, if not then the value s which you have got doesn't really carry much weight.

I should like to have a shot or two, giving $\varphi(x)$ different forms from c to (say) $e^{-(x-s)^2/2\sigma^2}$ & see what it amounts to but I fear that my analysis wouldn't be able for it.

Of course I quite see that given a definite series you stand to hit the mode most often but the point is that there may be another series whose chance of being the

author of the trouble is so much greater that some value *not* the mode may have a better chance than the mode of the first.

But then, that's all clear to you without my writing it over again.

As we shall see, Gosset reasserted his argument for the uniform prior in a letter to Pearson of 6 July 1917, and again in a letter to Fisher dated 3 April 1922.

The extension of Fisher's paper which Pearson had proposed was indeed very laborious, and Pearson assembled a team³ consisting of Herbert E. Soper, Andrew W. Young, Beatrice M. Cave, Alice Lee, and himself to undertake the task. Progress was reported to Fisher on 4 November 1915, and by 13 May 1916 Pearson could write to him that 'the *whole* of the correlation business has come out quite excellently'. After this tremendous amount of work, done in the intervals of time that could be spared from assisting the war effort, publication was delayed by financial problems affecting *Biometrika*. However, the results eventually appeared in the issue for May 1917 as a paper 'On the distribution of the correlation coefficient in small samples. Appendix II to the papers of "Student" and R.A. Fisher. A Cooperative Study'. The introductory section concludes as follows, in which equation (iv) is Fisher's expression for the frequency distribution of r .

Clearly in order to determine the approach to Soper's approximations, and ultimately to the normal curve as n increases we require expressions for the moment coefficients of (iv), and further for practical purposes we require to table the ordinates of (iv) in the region for which n is too small for Soper's formulae to provide adequate approximations. These are the aims of the present paper.

Section 8 is concerned to determine the 'most likely' value of the correlation in the sampled population, identified as the mode of a posterior distribution. But the authors argued that a uniform prior distribution is not valid for correlations, and this was the aspect of the paper which Gosset addressed when he wrote to Pearson on 6 July 1917.

Just a line to say that *Biometrika* has come & to convey my respectful admiration of the cooperative paper: what a landslide of work I did start when I began to play with my small numbers! I am not altogether sure that I quite agree with all you say about Bayes & would like to put up three points.

(1) As to your note on p.353, it's all very well to poke fun at infinity & zero but as a matter of fact they don't come into it at all. In finding the maximum the only parts of the scale which are practically considered are those likely to give a maximum* all of which obviously lie quite close to Σ unless of course n is 2: & by the way what happens to your formula when n is 2!

Observe too that although you talk about your previous knowledge you make no

*All you really require to give this result is a curve of distribution of ignorance very platykurtic with its mode around Σ .

use of this knowledge when you take $\sigma = [n/(n-2)]^{1/2}\Sigma$. I don't see what philosophical basis you have at all for inverting the thing.

(2) What you say about the a priori probability of getting low correlation rather than high is first class but I am not sure that it is easy to make use of knowledge of similar correlations without destroying the independence of the result in question (*vide infra*). But would it not be possible to compare various a priori distributions of ignorance not equal. I take it we would all rule out U curves but possibly if we wrote for $\varphi(\rho)$ $1 - \rho^2$ or $(1 - \rho^2)^2$ or even $(1 - \rho^2)^3$ we should get distributions of ignorance more appropriate for correlations in general i.e. of ignorance concerning the particular subject but not of correlations in general.

(3) The disadvantage of using actual knowledge concerning similar work is that you destroy the independence of the work before you. When you have the result what does it mean? & what do you want it to mean? Surely it is better to be able to say 'From our general experience of the correlation coefficient the population of which this is a sample probably had a correlation coefficient of .58 but this is much higher than that found from similar populations which have a mean of about .40' than to say 'Combining our knowledge of similar populations with the actual result before us the population in question probably had a correlation coefficient of about .45'. In the latter case you allow so very little to the work which you are particularly investigating & so much to the body of work which is not before you.

Unfortunately, the authors of the 'Cooperative Study' misunderstood Fisher's method of deriving an approximate 'absolute criterion' estimate of the correlation coefficient. They wrongly asserted that his equation was deduced from Bayes' theorem with a uniform prior distribution, and they criticized this choice of prior. The mistake could have arisen from their conviction that the most reasonable value of a parameter is the mode of a posterior distribution, or because of the wording employed by Fisher. He described his estimate as 'the most probable value', and when he responded to arguments by Kirstine Smith against the maximization of probability densities, the Draft of Note⁴ which he sent to Pearson in June 1916 described his 'absolute criterion' as 'derived from the Principle of Inverse Probability'. Whatever the reason, Fisher made his position quite clear in 1921 [CP 14], when he stated that his method of estimation involved no assumption whatsoever as to the probability distribution of the true value of the correlation coefficient⁵. One of the illustrations in the 'Cooperative Study' had concerned a sample of 25 with a product-moment estimate of 0.6, reduced to 0.59194 by a uniform prior distribution, and reduced further to 0.46225 by a prior considered to be more appropriate. Fisher used the same data to illustrate his transformation $r = \tanh z$, and remarked that 0.462 exceeded the prior mean by only 0.002. Gosset wrote on 3 April 1922 to acknowledge the receipt of offprints of CP 14 and CP 19, and to argue once more against non-uniform prior distributions.

When I was in the lab. in 1907 I tried to work out variants of Bayes with a priori probabilities other than $G=C$ but I soon convinced myself that with ordinary sized samples one's a priori hypothesis made a fool of the actual sample (as the co-operators

found) and since then have refused to use any other hypothesis than the one which leads to your likelihood (where I could deal with the Mathematics). Then each piece of evidence can be considered on its own merit.

It is typical of Gosset's insight that he should have seen that independent likelihoods can, but independent posteriors cannot, be combined, unless the latter are equal to the former.

Gosset's letter of 6 February 1911 refers to the value to him of coefficients of rank correlation, suggested by Charles Spearman in 1904, and which K.P. described as a 'method of assay' in the *Drapers' Company Research Memoir*, IV, of 1907.

I have been doing a lot of work with the 'Rank' correlation method. My samples are small (8 cases) but fairly numerous (12) so that the averages give first approximation answers. The characters are of varying degrees of preciseness from percentages given to two places of decimals to the difference between two estimates of stickiness in hops or two sets of opinions on the bitterness of beers. Some of the results (averages) are very good compared with results obtained from larger samples properly so that one can trust the others pretty well. I have 17 characters but am not working out all the combinations and I find the great advantage of the method is not that you save time on the individual correlations but having once set out a few characters one can add column after column of new characters doing all the difference and squaring on other pieces of paper. Perhaps the ideal would be to have little slips with the rank numbers on them & then the slips could be set against one another for correlation.

I find that when there are many 'ties' the formula requires modification as the S.D. of a set of ranks with a tie is lower than the ranks without a tie. The following correction may not be right but at all events it gives the same result if you reverse the order of one of the characters.

This correction appeared in a brewery report of 1911, and eventually led to his substantial paper concerned with 'An experimental determination of the probable error of Dr Spearman's correlation coefficients', where the material of his sampling experiment of 1907 is used again. The paper was read to the Society of Biometricians and Mathematical Statisticians on 13 December 1920 and published in the issue of *Biometrika* for July 1921.

4.3 TIME SERIES

The problem of relating the simultaneous movements of two time series emerged towards the close of the nineteenth century, and after the discovery of correlation precise methods of analysis could be developed. Two papers on the subject by Reginald H. Hooker, which appeared in the *Journal of the Royal Statistical Society* for 1901, each foreshadow one of the principal methods subsequently used. In the first paper, Hooker related marriage rate and trade by correlating their deviations from trends established by averages of nine

years. His second paper contains the suggestion of correlating differences between prices on consecutive days, and the same idea was put forward in connection with barometric readings by Frances E. Cave-Browne-Cave in *Proceedings of the Royal Society*, Series A, for 1904. This approach, which became known as the variate difference correlation method,⁶ was studied by Hooker in the *Journal of the Royal Statistical Society* for 1905, and he concluded that the correlation of differences was useful when investigating the similarity of rapid changes with no apparent periodicity.

Gosset visited Pearson at East Ilsley in July 1905, and after returning to the Guinness Brewery armed with references on the subject of correlation, he proceeded to examine the connection between analysis of malts and acidity of beer, taking first differences to eliminate temperature effects. During his sabbatical leave at the Biometric Laboratory, Gosset discussed the correlation of first differences with Pearson, who argued that the method was only appropriate for linear relationships. Further details of Gosset's relevant activities at this time are given in the autobiographical section of his letter to Fisher on 15 December 1918.

2. When the vagaries of laboratory determinations and of large scale experiments drove me into investigating the theory of errors and so to Karl Pearson, there was a period of about a year during which I was in charge of our Experimental Brewery when I worked on what I could pick up from a study of various papers to which K.P. referred me. Now it happened that one of the points to be investigated was the connection between the Laboratory Analysis of malts and the length of time the resulting beer remained potable as measured by acidity (if you have ever drunk Guinness in England you will understand why). But one of the chief factors in acidity production is the temperature (both at brewing and during storage) and at that time our arrangements for stabilising temperature in the Experimental Brewery were rather primitive. Hence I was forced to take first differences between successive brewings (i.e. days, for the plant would only produce one a day) to eliminate the large temperature effect.

3. After that the firm sent me to Gower Street for a year during which I went through Pearson's usual course and also made those tentative efforts to solve my own little problems of small numbers which you subsequently tackled so successfully. During that time I talked to K.P. about the first difference method and he showed me that with random observations $r_{\Delta a \Delta b} = r_{ab}$ but denied that you could eliminate time effect as if you have anything but a linear relation with the time you still have time effect in the 1st Differences.

Gosset had realized from the study of field plot observations and laboratory determinations how in practice there tended to be a correlation between observations made near together in space or time. The elimination of secular changes was a topic which continued to engage his attention for about seven years after his return to Dublin. His letter to Pearson on 9 November 1908 describes what he has been doing.

The first way of dealing with the difficulty was to take time as a third variable & having correlated it with such of my variables to take the partial

$$\rho_{12} = (r_{12} - r_{13}r_{23}) / [(1 - r_{13}^2)(1 - r_{23}^2)]^{1/2}.$$

But as my correlations with time are rarely linear & can certainly never be considered random I don't always get very useful results.

Next I tried Hooker's way (or mine as far as that goes) of correlating successive differences between the variables finding in fact $R_{(x_1 - x_2)(y_1 - y_2)}$. This very nearly gives you the partial when the regression is linear though of course any progressive change must come into the correlation. The difficulty is rather in manipulation. I have all my results on cards & it's very easy to make mistakes when taking differences. (Also the prob: error is rather large).

A third way I have tried: it is easy to carry out & gives what may some times be wanted, if I am not mistaken, but it always gives a lower value than the partial. The way is this. . . . I fear I am not very clear. In any case Hooker's method is the best for me but I would like to know if I'm right about the other.

On 8 December 1910, another approach to the problem is suggested.

Now in general the correlation weakens as the unit of time or space grows larger & I can't help thinking that it would be a great thing to work out the law according to which the correlation is likely to weaken with increase of unit. Of course some arbitrary assumptions must be made as to the distribution of correlation so to speak & I can't get hold of any reasonable supposition.

The effect of spatial correlation is then explored using figures collected by A.D. Hall⁷ of Rothamsted 'an agricultural experimenter of some notoriety'.

Towards the end of 1912, Gosset's activities in this area intensified, and his letters of 12 September, 18 September, and 13 October record the results. Here the variate difference method is applied to study the correlations of tuberculosis death rate with infantile mortality, marriage rate with average wages, and bankers' clearing house returns per head with Sauerbeck's index numbers. The letters display a natural pleasure in the progress made, and explain Gosset's search for illustrations in the economic field, where the figures were supplied by his colleague Edward G. Peake.⁸

If one now took 2nd differences & correlated *them*, one would have a criterion of whether the first differences had achieved their object. They should in that case give the same result. I will try. That's rather a find! . . .

You will recollect that the method was first published by Hooker in the Journal of the Statistical Society [*sic*]. He called it correlating differences from the 'instantaneous' mean. It is rather an obvious method, I used it, off my own bat, within six weeks of your first telling me about correlation. Anyhow I have to use something of the sort over and over again. . . . The elimination of secular changes in experimental work is of course the point that interests me, but apart from changes from point to point in a field none of my work could be published: that is why I am attacking these economic figures.

Pearson's friendly letter of 17 September 1912 contains thoughtful comments on spurious correlation, and poses questions intended to clarify Gosset's analyses of the data on tuberculosis and phthisis, but his other replies have gone forever. A further indication of Pearson's views can be gleaned from Gosset's letter to Fisher on 15 December 1918, which passes quickly over this period but shows Gosset taking an independent line.

4. There I left it for some years, using the method when necessary in a qualitative kind of way until the publication of Peake's paper in the *Banker's Magazine*, for the methods of which I was distantly responsible, led me to investigate the matter more fully so as to be able to defend the method against K.P. and the result was the paper in *Biometrika* to which you refer.

Gosset's short paper on 'The elimination of spurious correlation due to position in time or space' was sent to Pearson with a covering letter on 26 January 1914 and published as *Miscellanea (iv)* in the issue of *Biometrika* for April 1914. The letter contains a tantalizing comment on Pearson's role.

Also note that it is really Hooker: my part has merely been to be the anvil on which you have hammered Hooker.

Gosset showed that if $\{X_n\}$ and $\{Y_n\}$ are time series which each consist of the sum of (a) a polynomial in time with constant coefficients and (b) a random error independent of time, then the correlation of the random errors can be estimated by differencing both series until a steady value for the correlation of the differences is obtained. Publication was delayed because 'n + 1 other things have interfered'.

Meanwhile Oskar J.V. Anderson, a pupil of Chuprov at St Petersburg (Leningrad), had in 1911 submitted his diploma thesis on the correlation analysis of time series, which independently established the variate difference method in its general form. His paper, using mathematical expectation to determine the variances of differences and correlation coefficients, was sent to Pearson on 9 June 1914. Although Anderson, with the assistance of his brother, had compiled the covering letter in English, Pearson replied in German, and incidentally made the following assessment of Gosset.

'Student' ist nicht ein Fachmann, doch glaub' ich daß Sie zu vieles Gewicht an seine Worter liegen ...

Anderson's paper appeared in the issue of *Biometrika* for November 1914, together with a paper by Beatrice M. Cave and Pearson, which gave numerical illustrations relating to ten economic indices of Italian prosperity for the years 1885 to 1912. All this activity attracted Fisher's attention, and in 1916 he wrote a review article [CP 7] appreciative of Student, where he opined that 'the Variate Difference Correlation Method has evidently a great future'. However, the method as then used took no account of the difficulties arising from the existence of lag correlations, and was for that reason criticized by

Warren M. Persons in *Publications of the American Statistical Association* for 1916–17. Late in 1918, Fisher sent Gosset a letter which seems to have been concerned with these difficulties. Gosset replied on 15 December with the autobiographical details already quoted, and went on to express his doubts about the use of the method when applied to vital or economic statistics.

When K.P. produced his Italian paper I was, I confess, a little uneasy, he seemed to me to sound rather too high a note; and when subsequently he sent me two papers attacking the method to review for *Biometrika*, I felt that the thing had got beyond me and would have refused to do so even had my anonymity permitted me to indulge in controversy.

One of these papers is I think apropos. The statistics in question concerned, if I remember right, the death rates of infants during the 1st 2nd . . . years after birth and this sportsman maintained that owing to epidemic diseases there was a two year period in the death rates at all low ages and went on to talk about sine curves and amplitudes and other things of which I have had no adequate experience. Anyhow I wrote to K.P. and told him that I didn't see how you get rid of a two year period by taking differences and left it at that.

Since your letter came I've been thinking about it and there seem to me to be two weak points in the application of the difference method to ordinary vital statistics so that I doubt myself whether it will ever be used except for the obvious 1st difference which I believe to be both legitimate and useful.

They both depend on the assumption which is made that $r_{a_1a_2}$, $r_{b_1b_2}$, $r_{a_1b_2}$, $r_{a_2b_1}$ etc. are negligible or zero.

Clearly yearly (or even monthly) vital or economic statistics are not like laboratory observations (where also there are usually time effects) for the former are the result of summing periods while the latter are individual results.

Now the variation in such sums can only be due to 'causes' operating to a greater or lesser extent over the period in question and it would only be by chance that they would alter suddenly at the end of such a period. Hence successive periods (years) are correlated quite apart from the general trend which the difference method was invented to dispose of. Hence $r_{a_1a_2}$, $r_{b_1b_2}$ are in general by no means negligible nor are $r_{a_1b_2}$ etc.

He continued with suggestions for the analysis of time series which had wave form, and his brief comments in a letter to Fisher a fortnight later mark the point at which the variate difference correlation method disappears from his correspondence.

At a meeting of the Royal Statistical Society in 1921, G. Udny Yule read a paper in which he criticized the contributions made by Gosset and Anderson, and favoured a return to Hooker's method of correlating deviations from trend. Although Major Greenwood supported Gosset's work in the discussion on the paper, Fisher pointed out the complications that followed when lag correlations exist. When in 1924 Fisher examined 'The influence of rainfall on the yield of wheat at Rothamsted' [CP 37], he argued that correlating the

residuals of time series was best effected by first fitting polynomials, and that was the advice he gave in §37 of *Statistical methods for research workers*.

4.4 DISCRETE DISTRIBUTIONS

Gosset's list of publications begins with his paper 'On the error of counting with a haemocytometer', which appeared in the issue of *Biometrika* for February 1907. The paper is concerned with the distribution of yeast cells or blood corpuscles when a liquid containing such particles is spread in a thin layer over a grid, and the number of particles per unit area is counted. Gosset obtained what he termed the exponential series as a limit of the binomial distribution when the number of unit areas is large. He compared the two distributions, and showed that the exponential series approached normality as the number of particles per unit area became infinite. The theoretical work was tested on four frequency distributions derived from counts over the whole 400 squares of the haemocytometer, and the agreement was found to be satisfactory. Gosset noted that binomials fitted better than exponentials, and he pointed out that this was only to be expected because there was one more constant to estimate; his observation was made long before the chi-squared test was modified by the introduction of 'degrees of freedom'. But he was unaware that his exponential series was in fact a Poisson distribution, and that the theory and application of this distribution had a history extending over the previous seventy years.⁹

Gosset's interest in discrete distributions was renewed following a discovery made in the laboratory of Sir Almroth Wright that the power of white blood cells to ingest foreign organisms depends on the presence of opsonin in blood plasma. Opsonic power was determined by the number of bacilli per leucocyte, and the ratio of the value for the patient's serum to the corresponding value for normal serum defined the opsonic index. The estimation of opsonic indices was first discussed from a statistical viewpoint by Greenwood and White in *Biometrika* for March 1909. They presented several frequency distributions of the number of bacilli in phagocytic cells, and derived distributions of the means of small samples. Pearson curves of types I and V were fitted, and led to the conclusion that not only were the populations skew, but that the means also had skew distributions. Gosset found the skewness of the means surprising, and he attempted in his paper on 'The distribution of means of samples which are not drawn at random', published in the same volume of *Biometrika*, to explain the form of the distribution on the basis of homotyposis. This hypothesis, according to which the individuals composing the sample are more like each other than the rest of the population, had been put forward by Pearson in 1901 to explain aspects of inheritance. Consider a sample of size n when each value has variance σ^2 and the correlation coefficient between

each pair of values is ρ . Denote by M_2 the variance of the sample mean. Gosset showed that

$$M_2 = \sigma^2 [1 + (n-1)\rho]/n.$$

He concluded from the values of the third and fourth moments that the distribution of the sample mean would tend to normality less rapidly than when $\rho = 0$.

A second paper by Greenwood and White appeared in *Biometrika* for October 1910, and was mainly concerned to study the distributions of the means for subsamples of 25, 50, and 100 taken consecutively from a sample of 20 000 phagocytic cells. The distributions were markedly skew, but the authors considered that this skewness could not be due entirely to homo-typosis. Their view was rejected by Gosset when writing to Pearson on 8 December 1910 about the problem of correlation between individuals which are successive in time or adjacent in space.

The same sort of problem occurs in the opsonic index work where, *pace* Greenwood, the eccentricity of the distribution of the means of 25, 50 & 100 *must* be due to correlation of *some* kind in space or time or both. It is not of course the simple kind of correlation which I assumed in my paper but you cannot account for the facts except by some sort of correlation between the members of a sample. It is curious and noteworthy that it is some kind of correlation which affects the skewness & kurtosis more than the variability. Sudden lumps of similar phagocytes on the slides, or half hours of observation at the microscope might have this sort of effect perhaps.

Around 1912, the work of Poisson and von Bortkiewicz¹⁰ became known in the Biometric School. In the issue of *Biometrika* for April 1914, Herbert E. Soper published a table giving the terms of the series

$$e^{-m}(1 + m + m^2/2! + m^3/3! + \dots),$$

and Lucy Whitaker explored the fitting of binomials (positive and negative) and Poisson series to many sets of data. Miss Whitaker thanked Pearson 'for his aid at various stages', and her trenchant style proclaims his influence. Previous approximations of the Poisson distribution were criticized, and in particular Whitaker found that some of the material which Gosset used for illustration in 1907 is consistent with negative values of the binomial parameter q . His letter of 26 August 1914 shows that since April there must have been some discussion (verbal or by letter) between Gosset and Pearson arising out of her paper.

What you say about the Poisson distribution is of course right enough but a Poisson distribution which doesn't start at zero is a queer bird & I don't quite see how it's to arise: anyway it doesn't get up out of a haemocytometer. Of course I had no business to call the negative q a criterion yet actually the majority *are* negative & positive ones are to some extent natural for you get them when n is not 'large' from an unmixed population.

In the November 1915 issue of *Biometrika*, Pearson published his paper 'On certain types of compound frequency distributions in which the components can be individually described by binomial series.' He began with a discrete mixture of binomial distributions $\text{Bi}(n, q_s)$ for $s = 1, 2, \dots, u$, fitted the binomial $\text{Bi}(\kappa, Q)$ by equating moments, and gave expressions for Q and κ in the limit as $q_s \rightarrow 0$ with $nq_s = m_s$, finding that $Q < 0$.

Thus, if two or more Poisson's series be combined term by term *from the first*, then the compound will always be a negative binomial. This theorem was first pointed out to me by 'Student' and suggested by him as a possible explanation of negative binomials occurring in material which theoretically should obey the Law of Small Numbers e.g. 'Student's' own Haemacytometer counts.

However, Pearson seems to have realized that Gosset's contribution had not been fully presented, and must have written to apologize for the omission. Gosset's letter of 15 February 1916 refers to this and says in a typical Gosset manner that he had not been hurt in any way, only interested in seeing the job 'properly done'. The letter continues with brief comments on his 1907 formula $0.67449N^{1/2}$, and proceeds to discuss Pearson's paper in relation to Gosset's exploration of various forms $y = f(m)$ for the probability distribution of the Poisson parameter.

All I wanted to say was that there is a large class of people, to whom I belonged at one time & even now do to some extent, who wouldn't be bothered with point binomials but who might find it very useful to know that if they had counted N individuals, a rough measure of their prob error is given by $\frac{2}{3}\sqrt{N}$, & that it might be worth while putting this out again for their benefit.

Next though I have quite good reasons for supposing that a type III curve is generally what you might expect for a distribution of that character, yet I'm bound to confess that I've also tried V & VI without success, the integrals soon getting out of hand. Anyhow it is quite beside the point, for you showed that the form of $f(m)$ doesn't matter before I took it up at all.

But what I was chuckling at, & in fancy I saw you sharing my amusement, was that I had been toiling away at a particular form of $y = f(m)$ when all the while you had done the thing for *all* forms of $y = f(m)$, to wit your $Q \dots$ which comes down to $-\sigma^2/M$ in my notation.

And now I'm not sure that you did quite see it after all, or at least the full significance of it. For if you allow, as you say in your letter that 'an indefinitely large number of perfectly mixed Poisson populations following any law of frequency might practically be used to describe any frequency whatever', it follows absolutely that when you take uniform samples in space or time from any population of the kind, you *must* get *negative* binomials, which was what I wanted to show, except in the rare case when σ of the $y = f(m)$ is 0, when you get the Poisson.

You will probably tell me that in fact you have got positive binomials, which of course would show that the sentence I quoted in inverted commas is not universally true, but I don't think *significantly* positive binomials are at all common.

Now if you consider how mixtures arise either in flasks or (shall we say) towns, it

is by gradual thinning of strata (you have observed the striae in a syrup which is being thinned with water), & as they thin they gradually approach mixture, but I take it that the image of a mixture of mixed populations is not inapt & I think one has thus a very good explanation of negative binomials: if you prefer to put it that way, negative binomials are what we *ought* to expect, the Poisson being a limiting case of the *negative* binomial & the $-Q$ being a measure of the lack of perfection of mixture of the population in question—imperfection of technique if you prefer it. We are, I think, altogether at one—except perhaps that I regard the negative binomial as being a necessity & you a convenience—but naturally my way of putting it gives *me* a better idea of things just as yours gives *you* the best idea.

Some of Gosset's ideas on the subject eventually took shape in his paper on 'An explanation of deviations from Poisson's law in practice'. This was published in *Biometrika* for November 1919, but seems to have been sent to Pearson at the end of 1918: in Gosset's letter of 1 February 1919, the proofs are awaited, and with his letter of 28 February they are returned. Several key assumptions are listed for the Poisson distribution to hold, and Gosset pointed out that 'if the different divisions have different chances of containing individuals', or 'if the presence of one individual in a division increases the chance of other individuals falling into that division', then a negative binomial will fit the figures best. There is nothing in the paper about a continuous mixture of Poisson distributions, but his letter to Fisher on 30 December 1918 shows that Gosset was investigating the choice of a Pearson type I curve for $f(m)$.

Briefly I've been bringing Poisson's Law of small numbers up to date. Miss Whitaker was the last I think, and she showed that all the statistics (including even Student's yeast cells!) were hopelessly unpoissonic and had a pesky way of giving negative binomials. Well, the thing depends on either mixture or correlation. Mixture in the individuals, i.e. that the chance of the individual reaching a particular division is not always the same doesn't matter much unless some individuals have a really appreciable chance when you get of course a positive binomial. On the other hand mixture of the divisions, i.e. that the chance of a division absorbing individuals varies, is of importance and gives you a negative binomial. The figures are, using the notation $m = nq$ for each division and $v_1 \mu_2$ moments of the frequency distribution of divisions with 0, 1, 2 ... individuals.

$$v_1 = \bar{m}$$

$$\mu_2 = m + \sigma_m^2 \text{ etc.}$$

Correlation will of course give you positive or negative binomials according as it reduces or increases the spread.

Having got so far it occurred to me to see whether you can explain the distribution of the phagocytes of the opsonists by supposing that the phagocytes are divisions with an unequal chance of wolfing bacteria.

I therefore took the monumental work of Greenwood and somebody who counted 20,000 phagocytes and from the moments of their distribution worked out the first four moments of the ' m 's which might be supposed to lead to their results. These

moments are so far reasonable that they give a β_1 of about .9 and β_2 of about 3.2 and a range with no negative m in it (type I).

Then I come to reconstruct their phagocyte count from my m curve and at once I fetch up to an integral which is beyond me though I fancy that it is something pretty in Γ functions it is


$$\int_{0.1991}^{9.4572} e^{-m}(m + 1.5676)^r(1 + m/.1991)^{.0492}(1 - m/9.4592)^{2.3383} dm$$

r is integral and the thing, multiplied by a suitable constant which I can get out of a book, should give the number of phagocytes in the original distribution which contain r bacteria. Could you tell me whether the thing has a reasonable solution and if so what? It would be some satisfaction to me even if it could only be done for $r=0$.

Gosset's letter to Pearson on 1 February 1919 presents a further analysis of the Greenwood-White data, but reports a lack of progress with the theory.

Since I wrote last I have been having a shot at that 20,000 leucocyte distribution of Greenwood's. You will recollect that he tried to fit a continuous curve to it & didn't succeed too well χ^2 being 83 & P a very small zero. He thereupon proceeded to 'explain' his want of success which always seemed to me an odd thing to do: all he had tried to do was to get a curve with the same moments as his observed distribution & even apart from the fact that he was fitting a continuous curve to discontinuous observations there was no reason why he should crab his observations as he did.

I assume that the distribution is obtained by a mixture of leucocytes having different ' m 's pointing either to different capacity for absorbing bacteria or to different opportunities for absorption or to both.

I then found the first four moments of the ' m 's, constructed the (type I) curve, read off the frequencies in .3 groups & by means of the Poisson tables (Soper's I think) in *Biometrika* reconstructed the original leucocyte distribution. It was not too good, χ^2 only dropping 10 to 73, but it occurred to me at once (I had been reading the literature) that as my curve, which was just not a J , rose sharply at about 1.7 there was no possibility of dead leucocytes which were elsewhere stated to occur though no numbers were given. I then assumed 500 dead, subtracted from the zero group & repeated the performance whereupon χ^2 fell to 27 and P rose to .04. No doubt there would have been a similar improvement in the continuous curve. I made two further efforts assuming 350 & 400 to be dead & screwed down χ^2 to 25.3 in the last one giving a P of .065 but as over 9 of the χ^2 was in one group & as P rises rather rapidly just about there I feel sure that somewhere between 400 & 450 dead leucocytes would give a very fair fit. The two later ' m ' curves were clubfooted J s a type I had not known of before  with the infinite part just a little over 2 & the 'toe' at about 8 & it seemed to me consistent with the supposition that during the preliminary circulation all the leucocytes which were alive got a chance of collecting bacteria represented by an m of just over 2 but that those which settled at once got no more while the later chaps which came down on top ran up to an m of 8.

It would be much simpler if one could integrate

$$\int_{a_1}^{a_2} e^{-x} x^r (1 + x/a_1)^{m_1} (1 - x/a_2)^{m_2} dx$$

but I understand from Fisher to whom I mentioned the matter in the course of a letter on his difference correlations that the thing won't integrate. I should have hoped myself for something natty in Γ functions.

The matter is discussed again in Gosset's letter to Pearson on 28 February 1919, which comments on the practical details of experimentation, but adds nothing with regard to the statistical aspects. It was left to Greenwood and Yule to round off this particular problem in the *Journal of the Royal Statistical Society* for 1920 by using a Pearson type III curve to represent $f(m)$, and giving an exact derivation of the negative binomial distribution.

4.5 COMMENT AND CRITICISM

Gosset became involved with agricultural experiments in about 1905, when he was approached for advice by the maltster Edwin S. Beaven, who was on commission to Guinness, and who carried out breeding experiments in 'cages' at a barley growing nursery near Warminster in Wiltshire. This was the beginning of a lifelong friendship between Gosset and Beaven, who met and corresponded regularly over the next thirty years, mainly on matters arising from the design and analysis of barley experiments.

When Gosset returned to Dublin in the spring of 1907, he could hardly fail to carry back with him an interest in biometry, which his work gave ample opportunity to extend. His letter of 20 September 1908 describes at some length a recent visit to Beaven's nursery. The visit was primarily undertaken in connection with brewery work, but Gosset had in mind the possibility that Beaven might be persuaded to undertake some barley crossing experiments for K.P. Gosset also thought that he himself might get some insight from Beaven's data into 'pure line' inheritance, and if so he hoped that K.P. would accept a paper for *Biometrika*.

The controversy between biometricians and Mendelians which occurred at the British Association meeting in Cambridge in 1904 quietened down somewhat after Weldon's death in 1906. Gosset attended some of the talks on the mechanism of heredity given at the Dublin B.A. meeting in 1908. Here, he met K.P.'s Irish assistant, Miss Amy Barrington, whom he knew from his University College visit. William Bateson spoke on Mendelian genetics more than once, but both Gosset and Miss Barrington thought that Bateson's data were not consistent with the simple Mendelian theory so far enunciated. Gosset's letter of 20 September 1908 gives his reasons for rejecting Bateson's views on colourblindness, and he wrote again on 9 November with details of his colourblind pedigree.

Another interest of the biometricians was in wasps and bees. The solitary Edgeworth had published two papers on wasps in the *Journal of the Royal Statistical Society* for 1885 and 1897. He was trying by an original method to estimate the distribution of the length of absence from their nest of individual insects and how this varied with the time of day, using observations made in Edgeworthstown (Ireland), Oxford, and Hampstead (on the golf links). Edgeworth was perhaps encouraged by Pearson to submit material on this topic for *Biometrika*. The issue for June 1907 contains two papers on wasps and bees, one by Edgeworth and the other by Alexandra Wright, Alice Lee, and Karl Pearson. Gosset's letter of 11 September 1909 refers to sending 'more wasps' to K.P. At the end of his letter of 24 April 1910, he discusses the habits of bees, and finally, with his letter of 16 June 1911, he reports the despatch of 'a small sample of wasps to University College'.

After Weldon's death, K.P.'s chief interest and research programme shifted from pure biometry to the work of the Eugenics Laboratory, financially supported by Francis Galton. Gosset received copies of the memoirs published from University College and read of the controversies which spread into the Press. His letter of 24 April 1910 refers to the paper by Ethel M. Elderton¹¹ and K.P. 'On the measure of resemblance of first cousins', published in 1907 as *Eugenics Laboratory Memoir*, IV. They compared coefficients of resemblance for characters based on continuous variables with those based on data classified only in 'broad categories'. K.P. dealt with the latter by introducing a variety of coefficients, e.g. based on a correlation ratio η^2 and on the coefficient C_2 of mean-square contingency. Gosset pointed out that it was doubtful whether the degrees of relationship given by these measures could be easily compared in tables with different numbers of categories and different sample sizes. He continued for several pages to discuss probable errors, with reference to the paper by John Blakeman and K.P. published in *Biometrika* for October 1906.

Much controversy was aroused by the publication in 1910 of *Eugenics Laboratory Memoir*, X: 'A first study of the influence of parental alcoholism on the physique and ability of the offspring' by Ethel M. Elderton with the assistance of K.P. They found no marked relation between the intelligence, physique, or disease of the offspring and parental alcoholism in any of the categories investigated. This conclusion was a blow for temperance reformers, who were advised to replace energetic but untrained philanthropy by real knowledge, while economists and medical men were likewise displeased. A review by J.M. Keynes in Vol. 73 of the *Journal of the Royal Statistical Society* was followed by an exchange between Keynes and Pearson in Vol. 74 under the title 'Influence of parental alcoholism'. Letters also appeared in *The British Medical Journal* and *Archiv für Rassen- und Gesellschafts-Biologie*. Gosset told K.P. on 6 February 1911 that he had 'been reading a small fraction of the

great alcoholic controversy', and he suggested further points on which the data might provide answers.

In the first place it seems possible that drinkers who have insane, tubercular or other tendencies may come to an end before they produce many children so that drinking parents may be to that extent selected healthy parents. Is this possible? The second point concerns the teetotal thesis that 'one very frequently sees the elder members of a family who were born before the parents took to drink, quite healthy while their younger brothers & sisters are progressively degenerate'.

Would your material furnish an answer to the question is this fact or imagination. For it seems to me that although the teetotalers have no right to use the argument that the children were some of them born before the drinking began (in view of their own poor record in the matter) yet in view of the fact that your 'drinkers' children are on the whole slightly older than the 'sobers' children it would seem possible that there is some, probably not much, weight in the criticism.

I do not forget your argument that the 'drinkers' children are as numerous between 11 & 12 as between 5 & 6 but it seems to me that the drinking & the prolificacy may be results of the same set of circumstances, temperament etc. & that possibly the prolificacy may precede the drinking. In any case a priori one would expect that *if* it is true that the placenta is permeable to alcohol, drinking by the mother might affect her children for the worse, & also in such cases as the mother suckling her children; & it would be interesting to see whether the correlation between place in family & mentality & health is of the same sign in sober & drinking families.

Tuberculosis was another social problem of the time, and the relevant statistics were examined by the Eugenics Laboratory. Gosset's letter of 12 September 1912 comments on this issue.

About the tuberculosis I find that the English rural counties also have their death rate from phthisis at its maximum at an earlier age than the English urban counties just as Ireland which is mainly rural has its maximum earlier than England. Herewith the figures from the Registrar General's report for 1907. I suppose this would support your provisional theory, for infection must be less prevalent in country districts.

Deaths per 100,000 1902-1906

	Urban male	Rural male	Urban female	Rural female
0-	437	265	367	241
5-	158	113	209	164
10-	174	149	384	445
15-	758	776	916	1,171
20-	1,406	1,874	1,130	1,617
25-	2,024	2,238	1,479	1,712
35-	2,953	2,051	1,812	1,481
45-	3,627	1,951	1,572	1,155
55-	3,030	1,730	1,195	1,000
65-	1,682	928	735	626

I daresay this is quite familiar to you, but it seems to bear on the Irish case.

The letter then proceeds to discuss the connection between tuberculosis deaths and infantile mortality, one of the examples which Gosset used to illustrate the variate difference correlation method.

With the advent of war, biometry and eugenics disappear from the correspondence between Gosset and Pearson. One of the letters to survive from this period, when the burden of work on K.P. was greater than at any other time in his career, shows that he was still capable of opening up a new field of enquiry. Kirstine Smith¹² was a graduate student in the Biometric Laboratory during the winter of 1915, when she completed her paper 'On the "best" values of the constants in frequency distributions', published in *Biometrika* for May 1916. She then began work on 'the distribution of experiments problem', and Pearson must have mentioned the project to Gosset, who wrote as follows on 7 May 1916.

(1) It depends on the kind of equation which you are going to fit to your observations.

E.g. if you have reason to believe that an equation of the form $y = ax^n$ is to fit your observations then a single point determined as accurately as possible where x is as large as possible will give you the best value of a . This follows, I think, from your assumption that the error of observation of y remains constant throughout the range consequently the error is relatively smallest when x is largest. Actually in most cases the error would probably increase with x & have a 'horn' distribution.

(2) I propose as a working hypothesis that the proper way is to divide your experiments over as many points as there are unknowns in the equation you propose to fit. Less points you obviously cannot fit but more means that you do not put the weight in the important places. E.g. in a simple parabola $y = a + bx + cx^2$ it looks at all events as if the two ends and the middle are the important points & that points between would not help much.

The weakness of this is of course that you don't very often know what sort of equation you are going to get.

Gosset's anticipation of some features of optimal design is noteworthy, but the algebraic details which follow are inconclusive because 'when I tried the method with more than two constants the silly thing seemed to give the same result which was absurd so I can't vouch for it'. Kirstine Smith's pioneering paper of 85 pages 'On the standard deviation of adjusted and interpolated values of an observed polynomial function and its constants . . .' was published in *Biometrika* for November 1918, and there was a gap of over thirty years before the resurgence of interest in optimal regression designs during the 1950s.

Gosset was now forty years old, a friend who could gently chide K.P. when he observed bearish behaviour in his former teacher. Pearson had criticized

an article by Leonard Darwin in *Eugenics Review* for July 1913, not only privately and in that journal, but also in a paper 'On certain errors with regard to multiple correlation occasionally made by those who have not adequately studied the subject', published in *Biometrika* for April 1914. Fisher supported Darwin's position,¹³ but neither of them challenged Pearson. However, Gosset took an opportunity to express his views to K.P.

I was looking up the partial correlation paper* for another purpose the other day & came across a sentence about half way down p.29 which seems to show that before & up to the writing of that paper your use of the word 'correlation' was not very different to Major Darwin's use of it at present, & I thought perhaps it might incline you to take a more lenient view of Major Darwin's paper if you were to read that page.

I believe that I was so rude as to omit to write a 'roofer' (hospitable roof etc.) after my stay with you: please let this atone. . . . I hope you got that finished up without further excitements. I worked away for a time with the hope that I might get some trick for smoothing third differences but without any success.

*Math. Cont. XI.

This letter, written from St James's Gate, Dublin, on 26 September 1916, was answered by Pearson from 7 Well Road, Hampstead, on the same day. While appreciative of Gosset's visit—and taking his help for granted—Pearson's views on Darwin's statements were unchanged.

I have been the person who should have written—not indeed to thank you for your aid as it seems absurd to thank a man for helping in what is national work—but to show you that I had not just let you come and go without leaving an impress on my mind. But up to the present I have been continuing the work—about 10 hours daily & there is still more to be done. . . .

I have looked at the passage to which you refer and I should agree with every word of it now! I have certainly never held the view that *selection* does not affect correlation, in fact I think I first gave the formulae for determining its influence, but Major Darwin originally never said a word about *selection*. He baldly stated that when environment was uniform then the correlation between environment & a character must be zero. As I said then & say now you cannot possibly determine the correlation from such data. If he had said you are determining a partial correlation in your environment coefficients with *zero variability* the absurdity of his statement would have been obvious because the variability is all *practically* available variability.

After February 1919, when Gosset returned the proofs of his second paper on Poisson's distribution, there is a gap of almost six years in what is preserved of his letters to Pearson, which never again comment on statistical work at University College. When the correspondence resumes late in 1924, K.P. had evidently been making enquiries about the availability of brewery data. Gosset replied on 30 November.

I am sorry to say that I could not let you have such figures as you ask for, even if they would suit your purpose, without the leave of the board.

A month later, however, he enclosed the issue of the *Journal of the Institute of Brewing* for March 1924, containing a lengthy paper on barley experiments.

Another gap from 1927 to 1931 leads to the final phase of the correspondence between Gosset and Pearson, most of which is concerned with the design of experiments and criticism of the *t*-test. These matters fall more appropriately into Chapters 5 and 6, respectively.

5. Ronald A. Fisher

5.1 INTRODUCTION

Although Gosset's work on the probable error of a mean appeared at the beginning of his career, and when his association with Karl Pearson was close, that line of research was far more influential for R.A. Fisher and is therefore the first of the topics to be considered here. Gosset's relationship with Fisher developed strongly after 1919, when Fisher went to Rothamsted. About half of the letters which passed between them date from the next ten years, and during this period much of Gosset's published research is concerned with experimental design, to which Fisher was then contributing new and fundamental ideas. Another great statistical event of the 1920s was the publication of *Statistical methods for research workers*, and Gosset's assistance to Fisher can be examined with the aid of correspondence. Much attention has been given to the final difference of opinion between them with regard to experimental design, but in fact other differences had already occurred, and were interspersed with friendly advice on both sides.

5.2 PROBABILITY INTEGRAL OF t

Gosset's paper on 'The probable error of a mean', published in *Biometrika* for March 1908, examines the following problem.

The usual method of determining the probability that the mean of the population lies within a given distance of the mean of the sample is to assume a normal distribution about the mean of the sample with a standard deviation equal to s/\sqrt{n} , where s is the standard deviation of the sample, and to use the tables of the probability integral.

But as we decrease the number of experiments, the value of the standard deviation found from the sample of experiments becomes itself subject to an increasing error, until judgements reached in this way may become altogether misleading.

... The aim of the present paper is to determine the point at which we may use the tables of the probability integral in judging of the significance of the mean of a series of experiments, and to furnish alternative tables for use when the number of experiments is too few.

Denote by x_1, x_2, \dots, x_n the values of a random sample of size n from a normal distribution with mean zero and standard deviation σ . Write x for the sample

mean. In accordance with the practice of the Biometric School, Gosset defined the sample standard deviation by

$$s = \left[\sum_{i=1}^n (x_i - \bar{x})^2 / n \right]^{1/2}.$$

He obtained the first four moments of s^2 , showed that they correspond with a Pearson type III curve and, assuming that a Pearson curve did indeed fit, derived the distribution of s . After noting that the symmetry of the distribution of x implied that its correlation with s was zero, he went on to show that the correlation of x^2 with s^2 was also zero. Inferring on this basis that x and s are statistically independent, he found the distribution of $z = x/s$. He investigated the properties of the distributions of s and z , confirmed the theoretical results for $n = 4$ by a sampling experiment, and showed that the z -distribution tends to normality with variance $1/(n - 3)$ when n becomes large. The solution of the problem was completed by the calculation of a table of the probability integral of z for values of n from 4 to 10 inclusive, thus complementing the standard normal tables to which the introduction refers.

We see then that if the distribution is approximately normal our theory gives us a satisfactory measure of the certainty to be derived from a small sample in both the cases we have tested; but we have an indication that a fine grouping is an advantage. If the distribution is not normal, the mean and the standard deviation of a sample will be positively correlated, so that although both will have greater variability, yet they will tend to counteract each other, a mean deviating largely from the general mean tending to be divided by a larger standard deviation. Consequently I believe that the table given in Section VII below may be used in estimating the degree of certainty arrived at by the mean of a few experiments, in the case of most laboratory or biological work where the distributions are of a 'cocked hat' type and so sufficiently nearly normal.

Gosset gave four illustrations, the first concerned with the different effects of optical isomers in producing sleep, and the others with experiments published in the *Journal of the Agricultural Society*. His first data set gave the additional hours of sleep obtained with a treatment (1) as compared with (2). The mean of 10 observations was +0.75 and the standard deviation was 1.70, giving $z = 0.44$. From his table, he found the probability $P = 0.887$ for z to be less than this and deduced that

the odds are 0.887 to 0.113 that the mean is positive. That is about 8 to 1, and would correspond in the normal curve to about 1.8 times the probable error. It is then very likely that (1) gives an increase of sleep, but would occasion no surprise if the results were reversed by future experiments.

By expressing his ideas within a framework of inverse probability, and by translating his probability into the then traditional terms of a 'probable error', Gosset was following standard practice. But he differed from the Biometric School in using different symbols to denote population parameters and sample

estimates; their custom had been to use identical symbols, leading to much confusion. The paper is also notable for what is perhaps the first use of empirical sampling.

Gosset's friend Beaven was in touch with agricultural work at Cambridge, and his report of Gosset's keen interest explains how Gosset came to make contact with Frederick J.M. Stratton, an astronomer who lectured on the combination of observations. A paper by Thomas B. Wood and Stratton on 'The interpretation of experimental results' was published in the *Journal of Agricultural Science* for 1910, in circumstances which Gosset described when writing to Pearson on 18 September 1912.

If I'm the only person that you've come across that works with too small samples you are very singular. It was on this subject that I came to have dealings with Stratton, for in a paper setting up to teach agriculturalists how to experiment he had taken as an illustration a sample of 4! I heard about it, wrote to the man whom I supposed to be writing the paper with him and he forwarded my letter to the guilty pair. They sent me their papers to correct the day before the proofs were sent in and I mitigated some of it! A high handed proceeding, but all for the good of the cause.

Stratton took great pleasure in giving encouragement to younger men, and so Fisher was fortunate to have him as tutor at Caius. When he was still an undergraduate, Fisher wrote his paper 'On an absolute criterion for fitting frequency curves' [CP 1], in which he applied what he later termed [maximum-likelihood to estimate the mean and variance of a normal population. Stratton made him send a copy of the paper to Gosset, who queried with Fisher his expression $[\sum(x-m)^2/n]^{1/2}$ for the estimated standard deviation, on the grounds that $[\sum(x-m)^2/(n-1)]^{1/2}$ was long established within the theory of errors. Fisher replied 'with two foolscap pages covered with mathematics of the deepest dye in which he proved, by using n dimensions that the formula was, after all $[\sum(x-m)^2/(n-1)]^{1/2}$...', as Gosset reported to Pearson on 12 September 1912 when sending him Fisher's next letter.

I am enclosing a letter which gives a proof of my formulae for the frequency distribution of $z (= x/s)$, where x is the distance of the mean of n observations from the general mean and s is the S.D. of the n observations. Would you mind looking at it for me: I don't feel at home in more than three dimensions even if I could understand it otherwise.

... It seemed to me that if it's all right perhaps you might like to put the proof in a note. It's so nice and mathematical that it might appeal to some people. In any case I should be glad of your opinion of it.

Pearson replied five days later from North Yorkshire ('home on Sept. 23rd') to say with some repetition that Fisher's proof baffled him.

I do not follow Mr Fisher's proof & it is not the kind of proof which appeals to me. His paper on 'A new criterion etc.' he sent to me and asked me something about reprinting in *Biometrika*. I did not think it of any importance at the time & had some

communication with him on the subject. Of his tutor Straton [*sic*] I saw a good deal at one time. He was puzzled if I recollect rightly because if x & y were independent & $z = x - y$ so that $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$ then surely $x = y + z$ and $\sigma_x^2 = \sigma_y^2 + \sigma_z^2$ and not $\sigma_z^2 - \sigma_y^2$! I don't understand Fisher's proof, for I see no reference anywhere in it to the Gaussian distribution which he starts by assuming. I should not have thought that any such relation as mean $(\bar{x} - m)^2/\mu^2 = 1/(n - 3)$ was true generally, but I do not see what the writer is doing at all. My failure may very likely only be evidence of my density. What is μ in his geometry, the radius of his sphere or what? He never condescends to tell you, nor show the links between each stage in his thought. Whether the proper formula for the S.D. is $[S(x - m)^2/n]^{1/2}$ or $[S(x - m)^2/(n - 1)]^{1/2}$ seems to be of very little practical importance, because only naughty brewers take n so small that the difference is not of the order of the probable error of the summation! Of course, if Mr Fisher will write a proof, in which each line flows from the preceding one & define his terms I will gladly consider its publication. Of his present proof I can make no sense.

Nothing further on the frequency distribution of z was published until 1915 when Fisher confirmed, in his paper on the distribution of the correlation coefficient, that Gosset's formula was correct.

By this time, an interest in small samples had become established within the Biometric School. A table of the distribution of the standard deviation in normal samples was calculated by Andrew W. Young and appeared in *Biometrika* for May 1916. He made comments on the occasional occurrence of samples of two and three which persuaded Gosset to enlarge his original table of the probability integral of z . The extended version, in which the sample sizes range from 2 to 30, was published in *Biometrika* for May 1917, after which z again disappears from view for a few years.

On 3 April 1922, Gosset enquired of Fisher about the distribution of a regression coefficient.

But seriously I want to know what is the frequency distribution of $r\sigma_x/\sigma_y$ for small samples, in my work I want that more than the r distribution now happily solved. If you cared for it I could run out my old samples of 4 on the slide rule to give an illustration to your solution.

Just over a week later, he asked about the probable errors of partial correlation and partial regression coefficients for small samples, and reported on the conflicting views of leading authorities of the time.

I know that Yule has proved that they are the same as the ordinary 'total' coefficients for large numbers, but in conversation with me Prof. Edgeworth once said that though he could understand correlation coefficients being useful to the likes of us, he doubted whether we should get much practical use from partials. He then expressed a feeling which I have 'in my bones' that the prob: error of a partial derived from small numbers is of a higher order almost than that of the corresponding 'total'.

Fisher's replies have been lost, but they must have contained tests of significance for the difference between two means as well as for regression,

partial regression, and partial correlation coefficients, all expressed in terms of z , because Gosset wrote as follows on 5 May 1922.

Anyhow as to the regression factor the net result seems to be that if you use the accepted formula for S.D. you must do so exactly as you use the accepted S.D. of a mean, and with small numbers should use Student's Tables: which is of course very satisfactory for Student!

The ordinary formula must have been deduced on the supposition that the S.D. for the given values of x was required: it is obvious that that is so when it is pointed out to one.

It had not occurred to me to test the significance of two means of different sized samples by my Type VII Curve, nor have I had time (and inclination together) to try and find out how, since your letter came, but I could probably get the work of tabulating your integral done easily enough though whether I should be allowed to publish the work of the man I'd get to do it is another matter.

I am surprised that the effect of taking a partial correlation or regression is only to diminish the weight by one case, but I see that it is in line with other phenomena of the kind.

The reference here to 'tabulating your integral' suggests that a change from z to t was indicated, where

$$t = z\sqrt{v}$$

is the now standard criterion defined using the number of degrees of freedom v appropriate to the problem considered.

Gosset visited Rothamsted in September 1922 and met Fisher for the first time, afterwards sending him a copy of Student's tables 'as you are the only man that's ever likely to use them!' On 12 October, Gosset accorded the type VII a few lines in a letter mostly concerned with answering an enquiry about his use of Macdonell's data to find empirical sampling distributions of the correlation coefficient.

I haven't yet had time to do anything with the type VII, apples at home and business in the Brewery but hope to get on to it soon.

The hope was fulfilled and his letter of 7 November gives the details, together with two columns of probabilities calculated by Gosset, one using his trigonometrical series, and the other using Fisher's expansion formula [CP44].

I have recently been working a little at the Type VII and in the absence of my wife in England, I put the office Baby Triumphator in my rucksack and have been messing about with it at home, partly in the hope of understanding Tract II for computers.

In the course of that study I calculated all the values for $t = 1$ from $n = 2$ to $n = 30$ to seven places (accurate to 6 places) . . .

Last night I checked your values for $x = 1$ (discovering a slight slip) from your correction formulae and calculated the same values to seven places. As I used the sum of at least four numbers of 7 places they also have an error in the seventh place due to approximations, but the correspondence is quite wonderfully close down to about $n = 10$. Then your formulae go high, either because the omitted later terms are

not negligible or because the fourth correction should be less than you make it or of opposite size or indeed from a mixture of these causes. I hope to get some light on this from an examination of the figures.

Gosset worked hard on the tables throughout the winter of 1922–3, but then the pace slowed down for reasons given in his letter of 6 February 1923.

I will work away now, as opportunity offers at the remaining values of C_3 and C_4 , but people are getting querulous about the machine and I really cannot spare daylight to work on at the Brewery so I fear that I shan't do much more till next winter. Don't hesitate to put someone else on it if you are in a hurry.

Work on *Statistical methods for research workers* began in the summer of 1923, and this could be the reason why Fisher enquired when the table of t would be completed and whether he could quote the table of z already published. Gosset replied on 12 July.

I think you have all the completed work on the table, but I expect to finish it sometime next winter. I should say that it is certainly in course of preparation. As to 'quoting' the table in *Biometrika* it depends just what you mean by quoting. I imagine that they have the copyright and would be inclined to enforce it against *anyone*. The journal doesn't now pay its way though it did before the war and they are bound to make people buy it if they possibly can. I don't think, if I were Editor, that I would allow much more than a reference!

After an absence from calculation of over six months, he resumed work on 15 October using the asymptotic series, and was under the impression that the results were intended for Fisher's book.

The tabulating season having now commenced I took a calculating machine home on Saturday and began work last night on it. It took me practically the entire evening to pick up the threads and finally I only computed $\cdot 1$ for all values between $n = 5$ and $n = 21!$... I will finish C_4 and run out that part of the table which can be computed from your coefficients ... I take it that this table is, if it gets finished in time, to be published in your book. I'm not sure that I have any other method of publication open to me.

However, when Fisher made his summary of letter 34, he said he did not believe he ever seriously contemplated the reproduction of Gosset's table of t .

Both the tables of z , and all but one of Gosset's publications, had appeared in *Biometrika*, and loyalty to his old Professor is evident in his letter of 2 November.

Re publishing the table I've been thinking about it and have come to the conclusion that I must offer it to K.P. first. I rather doubt his wanting to publish a third table on the same subject especially as you would have to write the explanatory notes. Have you any objection to my offering him a table on our behalf on these lines? If he were to accept it would be all to the good and if he refuses it won't do much harm.

Gosset broached the subject on a visit to London soon afterwards, when his discussions with K.P. revealed errors in the tables of z and thus raised doubts about whether the table of t would be acceptable. He assessed the position on 23 November.

Whether he will have anything to do with our table I don't know. I rather doubt it, but personally I feel I could hardly put it before him unless you are prepared to do quite a lot of checking either yourself or per Miss McKenzie. Just as well you didn't take that table from *Biometrika*!

A fortnight later, Pearson had warmed to the proposal, but Fisher was resistant about checking, doubtless busy with his book.

K.P. again wrote that he would be glad to consider our table, for the second volume of *Tables for Biometers* and therefore presumably for *Biometrika* on the way. The same plates would be used.

It seems rather a shame to burden you with checking after what you say, but I think I may fairly put your own tables up to you.

Notwithstanding the decision to offer the table to *Biometrika*, Fisher remained keen to retain the right of publishing elsewhere. Gosset agreed on 20 December to make this point clear to K.P., while also stressing the difficulties that K.P. had experienced from breaches of copyright.

Re your postscript about publication, I quite agree: when the thing is put together I will either send it or take it to K.P. and will make it clear that you wish to have the right of publication in case you wish to include it in any book you may be bringing out.

Perhaps I may have been wrong in what I said to you: if I recollect right it was that you should not take the table without K.P.'s permission. I do know that K.P. made me get permission to reprint the table from Beaven's paper not only from E.S.B. but also of the Ministry of Agriculture and he himself is very sore with the Americans who have pirated both from *Biometrika* and from the *Tables for Biometricians*. The fact is that these things are either printed at a loss or at so small a profit that every effort has to be made to sell copies in order to make both ends meet.

The tables were 'long ago finished' when Gosset sent a copy on 20 May 1924 with a request for Fisher's account of them. He gave a description of the methods used, and added a P.S.

If you could let me have your account quite early next month I can probably take it to K.P. when I next get over. I'll get it typed as he is finding it more and more difficult to read manuscript.

Fisher was unable to complete his notes on the uses of the table and on his approximation formula until 17 July, so that Gosset would have only the table to take to K.P. in June, and he admitted on 31 May 1925 that it was left at his father's house when he visited University College.

I have at last taken the table to K.P. together with your explanatory notes: I am not at all sure that he won't publish it. On the other hand I gave him every chance not to and it may come on to you yet. I have I fear been very slack about it, I brought it over last year but unfortunately left it at home when I went up to the laboratory.

By the middle of 1924, *Statistical methods for research workers* was almost complete, and since Fisher was going to be away in Canada from the end of July to the beginning of September, he asked Gosset to read the proofs of the book.¹ A long list of notes and corrections accompanied Gosset's letter of 20 October. One of his suggestions was that the tables could be folded out of the book when in use, and this idea was implemented in earlier editions. The table of χ^2 had presented a problem to Fisher because Elderton's table in the first volume of *Biometrika* could not be reproduced without infringing copyright restrictions. Fisher therefore prepared 'a new table (Table III, ...) in a form which experience has shown to be more convenient'. He gave the values of χ^2 for selected values of P , the complement of the distribution function, instead of P for arbitrary χ^2 , and thus introduced the concept of nominal levels of significance. The footnote to Table III, giving the rule that $(2\chi^2)^{1/2} - (2n - 1)^{1/2}$ has a unit normal distribution for large values of n , follows another of Gosset's suggestions. Fisher used the same method of presentation for Table IV—the table of t .

The necessary distributions were given by 'Student' in 1908; fuller tables have since been given by the same author, and at the end of this chapter ... we give the distributions in a similar form to that used for our Table of χ^2 .

Presumably much of Table IV was derived by inverse interpolation, either from 'Student 1917', which appears next to 'Table IV' in Gosset's notes on the proofs, or from the table finished in May 1924 and referenced in subsequent editions of *Statistical methods for research workers*.

Work on the proofs continued until March 1925, at which time Gosset's assistant Somerfield was engaged in preparing an index. The project concerning the probability integral of t surfaced again in Gosset's letter of 31 May already mentioned, and he disclosed on 12 June that Pearson had liked only one of Fisher's two contributions.

K.P. is very anxious to publish your note about the use of the table, but doesn't like the binomial approximation which he considers requires a proof of convergence. It was in vain that I pointed out that converging or diverging the proof of the pudding lies (to me, doubtless not to you) in the fact that you get about seven places the same with $n = 21$ up to $t = 6$.

Anyhow he returns both and I send them herewith, his idea being I think that if you can prove convergence he would like to publish both and that if you can't you might prefer not to let him have the other though as I say he would like to publish it. I hope you will send it back to me however, whether the other consents or dissents as I'm sure that K.P. means to be conciliatory and in any case it ought to go into *Biometrika*.

Although Pearson was still expecting Fisher's explanatory notes at the end of September, the decision to submit tables, notes and formulae to *Metron* had been taken when Gosset and Fisher agreed early in October on how their work was to be presented. Gosset read the proofs in January 1926, returned them to Fisher early in February, and received offprints in June. Jack W. Dunlop wrote to Fisher from Stanford University on 5 July 1927 with a list of corrections, but they seem to have been misprints because Gosset was 'pretty sure the proofs left me corrected'.

5.3 'STATISTICAL METHODS FOR RESEARCH WORKERS'

Gosset received his copy of the first edition in June 1925, and a year later he gave an exposition of the art of reviewing.

I sent on an attempt at a review of Statistical Methods to the Sec. Eugenics Society and have had no reply I suppose it arrived all right. I found it very difficult to write, in fact if I hadn't happened to have a train journey, which somehow facilitates composition, in the middle I'd have been at it yet.

He noted on 22 October 1927, with his usual disregard for proper names, that the book 'gets a very good review in the Journal of American Statistical Society'.²

The comments which Gosset and Somerfield made on the proofs of the first edition in October 1924 included the following.

Suggest that to start off with such a technical example as Ex. 1 is a bit heavy. The non biologist is faced in the very first example with the following undefined jargon 'heterozygous, linked factors, dominance, viable, allelomorphs, genes, crossover ratios, gametes', besides having to take a certain amount of mathematics on trust ... In any case I should reduce the question to one of plants only as that is all you really deal with. But is there no problem of more general interest?

When the second edition came out in 1928, there was a new chapter IX, superseding Section 6 and Example 1 of the first edition. Gosset wrote on 1 April to express thanks 'for letting us see the additions to the book', and Fisher replied in 4 April.

I had rather hoped that you would have liked Chapter IX; it was in fact partly your suggestion, and I thought I had made rather more of it than you would have expected. I imagine the first sentence is my best reply to the question of what the practical research man wants it for. Do you not like the way χ^2 behaves? I was delighted with it, and I had fancied that the various formulae might save even you some time.

Egon Pearson reviewed the second edition in *Nature* on 8 June 1929, and his remarks led to an exchange of letters over the next four months which forms the subject of §6.5.

On 31 December 1934, Fisher received a letter from Isidor Greenwald of the University and Bellevue Hospital Medical College, New York University.

On pages 112–114 of the fourth edition of your book *Statistical methods for research workers*, I found a discussion of the results of some experiments by Cushny and Peebles. I was curious to see just what conclusions Cushny and Peebles had drawn from their observations and examined their paper (*Journal of Physiology*, 32, 501). I found that they stated that the levo- and racemic (not dextro-) forms of hyoscine (not hyoscyamine) had about the same influence in inducing sleep. The figures in their table justify this conclusion, the differences between the length of sleep after the levo and racemic forms of hyoscine being positive in 6 cases and negative in 4 and the mean being 0.05 hour. What 'Student' and you have done, apparently, is to misread their column 'L-hyoscyamine' as 'D-hyoscyamine' and their column 'L-hyoscine' as 'L-hyoscyamine'. I am greatly surprised that this error should not have been corrected long ago.

Somerfield found that all these statements were true, and Gosset confirmed the mistakes on 7 January 1935.

That blighter is of course perfectly right and of course it doesn't really matter two straws. The rummy thing about it is that I have no recollection at all of having selected two columns out of a four column table and if I had not such a genius for making slips I should be inclined to think that I had taken the figures from a notice of the paper. I fear you will have to alter the headings in your next edition and I give you full leave to slang me as much as you please in a footnote.

P.S. I remember I had a good deal of difficulty in getting any figures to illustrate with but I haven't the faintest recollection of how I managed to run across Cushny and Peebles. Of course it is not surprising that no one discovered the blunder for in the pre Fisher days no one paid the slightest attention to the paper.

Fisher decided that the drugs would be unnamed in future editions, although this was a matter of some regret, as he explained to Greenwald on 10 January.

I am rather sorry, as physiological differences between optical isomers have a certain interest in themselves, and I am at the moment engaged in testing some of their taste differences.

5.4 DESIGN OF EXPERIMENTS

5.4.1 Introduction

The scientific revolution of the seventeenth century was founded on planned experiments, repeated measurements, and the analysis of data by mathematical models. An agricultural revolution followed in the eighteenth century, and this section is concerned only with the design and analysis of agricultural field experiments. The four-volume study of experimental

agriculture published by Arthur Young in 1771 stressed the need for comparative experiments as a means of allowing for differences in climate and soil fertility, and introduced replicated trials in order to minimize the effects of environmental variation. James Johnston recommended in 1849 that the replications of each treatment should be as far removed from each other as was convenient. Thus the positions of different replicates could be related by systematic use of the knight's move in chess. He also proposed that two fertilizers should always be tested not merely alone but in combination. During the 1890s, Edwin S. Beaven began his experiments on barley at Warminster, and he developed two other systematic designs. One was the chessboard, an extension of the knight's move described by Egon Pearson (1939). The other was the half-drill strip method for comparing two varieties, which is considered in detail below.

Statistical methods were first applied to agricultural field experiments in two papers published early this century. Both made extensive use of data from uniformity trials, in which all the plots are treated alike. Wood and Stratton (1910) gave frequency distributions, calculated probable errors, applied tests of significance, and estimated the number of replications for a specified treatment to have a significant effect. Mercer and Hall (1911) made recommendations on plot size, number of replications, and experimental plan. An appendix to the second paper by Student gave a systematic layout for comparing two varieties, with the property that the standard error of the estimated difference between varieties was reduced by the correlation between half-plots. Both papers passed through Gosset's hands before publication, and the contact between Gosset and Fisher was effected by Stratton in his capacity as Fisher's tutor at Gonville and Caius College, Cambridge.

5.4.2 Chessboard plans

In 1923, there was an exchange of letters between Gosset on the one hand and Beaven, Fisher, and Yule on the other, concerned with the standard error of an estimated difference between varieties in a chessboard plan. The correspondence between Gosset and Beaven is reviewed by Egon Pearson (1939), and that between Gosset and Fisher is now interleaved. Gosset wrote to Beaven on 29 March with a note on the error, also to Fisher on the same day enclosing the memorandum reproduced by Pearson. Beaven told Gosset that he thought Yule was working at chessboards, whereupon Gosset thought what Yule would be likely to do, and sent him another note on the error. Gosset reported this action to Beaven on 9 April and to Fisher on 16 April. His letter to Beaven on 20 April summarized the replies from Fisher and Yule, and his letter to Fisher on 27 April gave the information that he had discovered

the mistake in the memorandum arising from the omission of $-\sigma_e^2/mn$. Fisher replied on 2 May.

I am glad the error estimate is straight now. A great beauty of splitting the sum of squares into fragments is that each fragment has independent sampling errors appropriate to the number of degrees of freedom. This greatly simplifies tests of significance; for instead of calculating say intraclass correlations, with some misgivings as to cross relationships and performing my transformations and corrections appropriate to such correlations one only has to make a direct comparison.

The remainder of this letter foreshadowed the distribution of F and its connections with the normal, t -, and χ^2 distributions. Letters from Gosset on 21 and 27 June concluded the interchange on the chessboard plan error formula, and his 1923 paper in *Biometrika* acknowledged Fisher's help in a lengthy footnote giving two derivations of the residual sum of squares.

5.4.3 Half-drill strip method

In the course of writing this paper, Gosset spotted a fallacy in work on the half-drill strip method which he had allowed Beaven to publish. The method compared two varieties, say A and C, by replicating the 'sandwich' ACCA, with the consequence that local linear trends in fertility were eliminated. When the strips were divided into sub-plots, the experimental plan was as follows:

```

A A A ... A
C C C ... C
C C C ... C
A A A ... A
A A A ... A
C C C ... C
C C C ... C
A A A ... A
. . . . .

```

Here, the strips correspond to rows, and pairs of strips were sown together as either AC or CA in a single drill.

Gosset stated the problem on 21 June.

The fact is that the sub-plots making up a half drill strip are correlated and therefore cannot be used to give you an estimate of the prob: error of a number of half drill strips. I think however that you can get a *minimum* error in that way: i.e. if by chance the (necessarily few) half drill strips give an error below that which would be calculated on a random basis from the sub-plots the latter should be taken.

He discussed the matter at length in four letters written in July, the contents of which are closely related to the second part of his 1923 paper. Gosset realized that yields from the sub-plots of a strip were positively correlated, partly because of faulty technique, but also because of changes in the rate of change of fertility. As a result, the standard error of (A-C) estimated from a pair of strips was greater than the same quantity estimated from sub-plot differences.

By the way, can you imagine a case where you could really gain in knowledge by dividing up your units into parts? My illustration to Beaven, not a very good one is that if you have a yellow and a green banana and wish to tell whether yellow or green bananas are sweetest (and have a good palate!) you don't gain much by cutting each into six pieces and tasting six pairs rather than eating two bananas!

He proposed to overcome this difficulty by taking as a unit the sandwich ACCA, estimating the correlation coefficient r_{12} between adjacent pairs of 'subsandwiches' along the rows, and assuming that the correlation fell off in accordance with $r_{pq} = r_{12}^q$. However, further difficulties then arose.

All the same I don't seem to have much luck in any attempt to get the correlation from a consideration of adjacent subsandwiches and an a priori theory of diminution in correlation for in fact, probably owing to technical difficulties such as you pointed out, the adjacent sub sandwiches are barely correlated while the next but one's are quite respectably so.

(average correlation adjacent	+ 0.06)
(average correlation next but one	+ 0.27)

Fisher replied with comments which appear to have concerned the existence of correlation in all directions, the underestimation and the exaggeration of errors, and the possibility of waves of fertility. Gosset ended his account of the problem by taking up the question of whether correlation along strips was compensatory.

Of course there is theoretically a length of subplot compared with the length of drill strip, where the positive and negative correlation will cancel and give you values for the S.D.s of the subplots and of the sandwiches which will give you the same p.e. [probable error] for your result, but in general I feel that to hit it off would be a queer coincidence; obviously it would vary not only with every field, but with the direction of your drill strips in the field and I'd lay long odds that even with these fixed it would vary from year to year.

5.4.4 Canons of experimentation

When Gosset was making his enquiries about the error formula for a chess-board plan, Fisher and Mackenzie had just completed their paper in *Journal of Agricultural Science* [CP32].³ They analysed the results from a factorial experiment on potatoes, where twelve varieties were 'planted in triplicate on

the "chessboard" system' and treated with combinations of dung and potash. This paper is sometimes described as the first with an analysis of variance,⁴ but the claim has been disputed.

Gosset wrote on 25 July, soon after publication.

I have come across the July J.A.S. [*Journal of Agricultural Science*] and read your paper. I fear that some people will be misled into thinking that because you have found no significant difference in the response of different varieties to manures that there isn't any. The experiment seems to me to be quite badly planned, you should give them a hand in that; you probably do now.

He proceeded to give his views on what the experiment could tell, and responded at length on 30 July after an enquiry from Fisher.

(3) How would I have designed the exp? Well at the risk of giving you too many 'glimpses of the obvious' I will expand on the subject; you have brought it on yourself!

The principles of large scale experiments are four.

(a) There must be essential similarity to ordinary practice and when I say *essential* I mean any departure whatever from ordinary practice which hasn't been proved to be inessential.

(b) Experiments must be so arranged as to obtain the maximum possible correlation between figures which are to be compared.

(c) Repetitions should be so arranged as to have the minimum possible correlation between repetitions (or the highest possible negative correlation).

(d) There should be economy of effort; i.e. all the experimental material should be concentrated on the decision point and no more experiment should be made than is certain to be enough to give a decision.

Gosset then examined whether each of these four canons was obeyed or violated in the potato experiment, and he found that (a) was violated, (b) obeyed, (c) not altogether obeyed, while with (d) the category depended on the object of the experiment. His remarks closed with general advice based on experience.

You will probably think many of my objections trivial and that experiments planned by me must be very stodgy. So they are, but my experience is that very often the silliest objections turn out to have enough in them to spoil your experiment. If I'm planning an experiment now I am careful to fit it symmetrically into the days of the week and hours of the day and every blessed thing I can think of: I've been had too often by 'trivialities'.

Lastly when you have planned your experiment show the plan to someone that you haven't said anything about it to and let him pull it to bits: things that seemed quite certain and obvious when you were planning may not be so to him and it will perhaps put you on your guard. Don't necessarily do what he says, but see that your reasons are better than his before you turn him down. The fact that he doesn't know what he is talking about is sometimes an advantage, he is in a better position than you to apply general principles whereas you are unconsciously biased by feasibility, practice and even opportunity.

Within the next year, Fisher began to develop his own canons of experimentation, and the first results appeared in 1925 at the end of *Statistical methods for research workers*. In §48, he rejected systematic arrangements, stated the principle of allocating treatments to plots at random, and showed that increased accuracy could be obtained by blocking. The Latin square was introduced in §49 and illustrated by an artificial example based on the uniformity trial data of Mercer and Hall. When Gosset returned the proofs of the book on 20 October 1924, he pointed out that the later pages showed less notes and corrections than the earlier, possibly because of his understanding less of the subject matter, but nevertheless he expressed firm views on Latin squares.

(2) I don't expect to convince you but I don't agree with your controlled randomness. You would want a large lunatic asylum for the operators who are apt to make mistakes enough even at present.

I quite agree that such an experiment as the 6×6 of the Irish plots is not at all good when systematically arranged but when you replicate the sets of six often enough the thing becomes random again. If you say anything about Student in your preface you should I think make a note of his disagreement with the practical part of the thing: of course he agrees in theory.

The same point was made when Gosset wrote on 30 November 1925 to say that he had taken some of the classical crop figures, from Mercer and Hall and others, and arranged them in five-sided squares divided into 'varieties' (a) by controlled randomness à la Fisher and (b) by a diagonal system similar to the Ballinacurra plots.

Then I have found hitherto that the variance of the means of the 'varieties' is much the same whether they are chosen on the A or the B system, at the present time the B variance is I believe slightly less than the A variance and I don't expect that there will be any appreciable difference between them.

I am going to infer that though the Latin Square has obvious theoretical advantages, yet for those who are apt to make mistakes in practice the other system has practical advantages which do not carry any great danger of actually departing from practical randomness.

In 1926, Fisher explained his canons of experimentation in a keynote paper [CP 48]. However, Gosset continued to regard the new designs as an extension of systematic arrangements, and from that viewpoint he preferred the Latin square to the randomized block as a device for regularizing the distribution of fertility. The paper by Eden and Fisher on winter oats in *Journal of Agricultural Science* for 1927 [CP57] gave him an opportunity to test this view by reference to their experiment, which was the subject of two letters in April 1928.

[13 April]

I have been looking into the winter oats paper and it appears that there is such a marked fertility slope from left to right in your diagram that your error is quite perceptibly larger than it would have been had you regularised the distribution, as you might have done without loss of randomness if not by Latin squaring at least by including one of each treatment in each column. In addition to a smaller error you would, apparently, have had a rather more consistent set of results, though there would not have been many more significant differences.

As this question of regularising has been a question on which our points of view have not entirely coincided I should like, *if you have no objection*, to defend the Latin Square against the Randomised block with your experiment as a test either in a note in the J.A.S. [*Journal of Agricultural Science*] or as part of a paper which I have in mind on the lack of randomness in things in general.

[18 April]

The fact is that there are two principles involved in the Latin Square of which I attach the greater importance to the balancing of the error and you to the randomisation. It is my opinion that in the great majority of cases the randomisation is supplied to any properly balanced experiment by the soil itself though of course where the ground has been used for experimenting before or for any other reason has met with a 'straight edged' lack of uniformity in recent years it is better to supply it artificially. (I *don't* consider the arrangement in the Irish chessboards a *properly* balanced experiment.) . . .

Lastly why do I propose to defend the Latin square against the randomised block? Because I cannot call to mind that you have published any results from Latin Squares and to my mind that is a much more damaging attack on it than any that have been made from other quarters.

Eight years later, what had been a private difference of opinion about the relative merits of experimental plans became a public controversy⁵ on randomized versus systematic designs. The course of events is considered in §6.7.

5.4.5 Lanarkshire milk experiment

A report on *Milk consumption and the growth of schoolchildren* by Gerald Leighton and Peter L. McKinlay was published in 1930, and concerned a nutritional experiment involving 20 000 children in 67 Lanarkshire schools. For four months, 5000 children received $\frac{3}{4}$ pint daily of raw milk, 5000 the same amount of pasteurized milk, and 10 000 acted as controls. Some schools were provided with raw milk, and others with pasteurized milk, but no school got both. The selection of children was made in certain cases by ballot and in others on an alphabetical system, but modified as follows.

In any particular school where there was any group to which these methods had given an undue proportion of well-fed or ill-nourished children, others were substituted in order to obtain a more level selection.

At the beginning and end of the experiment, all the children were weighed and their height was measured. Controversy ensued from the final conclusion of the report.

In so far as the conditions of this investigation are concerned the effects of raw and pasteurized milk on growth in weight and height are, so far as we can judge, equal.

This conclusion was challenged by Stephen Bartlett (1931), and on 18 April 1931 by Fisher and Bartlett [CP92], who argued that pasteurized milk has less value than raw milk for both boys and girls, although the doubt introduced by providing schools with either raw or pasteurized milk could not be wholly eliminated. Gosset reported to K.P. on 14 July 1931 that he was presently engaged in criticism of the experiment, and of the note in *Nature* by Fisher and 'a man at Reading'. Nine days later he sent K.P. a draft of his paper on the Lanarkshire milk experiment.

I hope you will find it interesting, though its chief merit to the likes of me (that there is no d---- mathematics in it) will hardly commend it to you.

K.P.'s reply of 26 July is reproduced in full to show his detailed remarks on Gosset's criticisms and proposals.

Your paper to hand. You seem to prove that little can be deduced from a very elaborate and expensive experiment! So far, so good, or rather so bad.

I do not know whether any statistical advice was given before the experiments were started, but at least one factor of growth seems omitted by you all. Namely that when you have a child, which by circumstances of birth or early environment is deficient in growth it tends to 'pick up' in later years. If your 'controls' consisted of age for age heavier and stouter children than the 'feeders' then I should anticipate that the older children of both groups would be closer together without any differential feeding. Does that bear on these results?

Now your practical proposals are (1) either to repeat the experiments on the same scale with more safeguards and more nutritional observations or (2) to experiment on identical twins.

With regard to (1), I see no need for a second experiment. The original schedules must still be accessible and what is more the School Medical Officers' cards with the several data as to nutrition, teeth, etc., of the whole or at least the bulk of these children. Now there is nothing to prevent anyone having access to that information, making a selection of pairs equal in age, and very closely in weight and height. It would clearly mean throwing out a certain number of the 'controls' and 'feeders' at each age, but enough would be left at ages 6 to 11 to get reasonable results. If the School Medical Officers' cards for nutrition, etc., were available, as they probably are at least for 'entrants' and 'leavers', some idea of the state of the three groups could be ascertained. I think, therefore, it would be well to point something of this kind out, rather than suggest a repetition of the experiment. If the selection were really random, then probably some of the children in all groups were getting adequate milk at home, and additions to this or even without this in the 'feeder' group might produce increased

weight without that weight being an advantage really to the child. The growth in weight, without it is in proportion to size, does not necessarily mean a gain in physical fitness. A simple muscular test applied to both 'controls' and 'feeders' might be of greater value than a weight test.

Next as to your (2) method—'Identical' twins. How are you going to determine them to be sure they are 'identical'? Quite a variety of methods have been suggested, but none appear very conclusive, or indeed satisfactory to me. Furthermore (a) is it possible to argue from twins to non-twins? The average weight of twins at birth is very considerably less than that of normal children and it may remain so a good way into the school age, but the principle I have referred to of accelerated growth comes into play, and they may largely approach normal non-twins.* To be sure of 'identical' twins you can only get information from those with adequate knowledge present at their birth. To choose them from apparent likeness and then demonstrate that such twins have higher correlation than unlike twins and are therefore like twins is somewhat circular.

(b) However let us suppose you have got your fifty, are you going to break them up into sex and age groups, and what will the probable errors be with or without such grouping? If you don't group them on *what* are you going to calculate your probable errors? On some other observations giving the standard deviation of other children at that age, and this although twins probably do not grow as other children? and doing so, how will you get a combined probable error for all your twins to ascertain whether the differences of your groups are significant? Or do you mean to assume that your like twins would have no differences in growth except for the milk? I think that is an assumption which needs proof and I don't think you will find it easy to establish. Let us suppose your identical twins start with somewhat unequal weights, that may mean unequal growth rates, and if the milk brings A up to B are you going to attribute it to unequal growth rate or to milk? With large numbers As and Bs would possibly be equally distributed among 'controls' and 'feeders' but I cannot imagine that this would necessarily be the case in any number of 'identical' twins you can get hold of. I am only making suggestions, but your constructive proposals seem to me open to criticism.

* When mice are killed at the 'same age' and when they are treated as adults, it is still found that the size of their bones is correlated with the size of the litter in which they were born!

When Gosset wrote next on 30 July, his response was likewise careful and detailed, and he concluded with a suggestion which was shortly to bear fruit, although not perhaps of the variety he intended.

(1) As you say, neither the authors of the Report, nor I, mentioned the fact that children deficient in growth 'pick up' but, though I was not aware of it as a fact, I think we both had the possibility in mind.

They tested the correlation between the weight (and height) before the experiment and the gain for all the 42 groups and found the coefficients small though some were significant: Boys' weight and girls' height negative on the whole, I think, and Girls' weight positive for higher ages.

I, on the other hand, considered that the difference between 'controls' and 'feeders' at the beginning was *not* due to a selection by height and weight as such, but by

selection of 'feeders' by poverty, and I certainly thought it probable that under-size due to such a cause would be made up when the chance of getting more, or better, food came. But what you have written certainly strengthens my argument.

(2) That I am not in sympathy with a large scale repetition of the experiment I intended to be inferred by the use of the word 'spectacular' in reference to it. Yet I did wish to point out that in my opinion they would have been well advised to choose pairs of children and toss for 'feeder'.

It had not occurred to me that even now they could do anything of the sort with the existing records, excepting only, as I suggested, by sorting out the 'controls' appropriate to the Raws and the Pasteuriseds. But of course this should be done, and, if you see your way to put in my paper, I suggest that Editorial footnotes should be made on both these points (1) and (2). Obviously they might be persuaded to go over their records by you while they probably wouldn't pay much attention to anything I might say. But are they to be trusted to make a proper selection after the event?

(3) I expect I should have developed my twin proposal at somewhat greater length.

In the first place I had in mind taking as many pairs of twins of the same sex as I could get, and really hope that Lanarkshire might produce 200-300 pairs between 5 and 11.

These I should divide *before the experiment* into pairs likely to be identical and those not likely to be identical. The Medical Officer would doubtless give an opinion on appearance.

The first group would probably contain 90 per cent identicals and the second practically none.

The diluents of non-identicals would put up the error, but not by *very* much because they would anyhow be very similar brothers or sisters; the effect of the identicals in the second group would be negligible.

(a) While admitting a theoretical possibility of not being able to argue from twins to other children, it is a little difficult for me to see the practical form of the disability in this particular case. If Raw milk is better for twins than is Pasteurised milk, what can prevent its being so with other children?

Of course I see that in some cases neither might have any effect, but, given that milk of sorts is a good thing for children, and I rather gather that there is a good deal of evidence in favour of this (nothing to do with Lanarkshire), you will differentiate most easily between the Raw and the Pasteurised by experimenting on those likely to benefit most by it, and twins would, according to you, be favourable subjects.

(b) I am going to compare the difference (in weight, height and any muscular exercise which you may suggest) between the two halves of a pair of twins. Such differences will give their own probable error. Of course sex and age will be tested to see how they affect it, and it may be well to group them in four groups by sex and age or we may find that there is no significant difference in the effects in the groups. If we put them all together we shall at the worst increase the error beyond what it would be if we were able to split them up into groups and deal separately with them.

A similar consideration would apply to grouping by social standing, under-nourished appearance at the beginning of the experiment, or any other relevant information.

Of course I don't mean to assume anything so foolish as that all the differences are due to the kind of milk that the child imbibes but, having tossed for which shall be

Raw and which Pasteurised, everything else (except of course the *extent** of the influence of the kind of milk) becomes a random error and can, therefore, be dealt with statistically.

The numbers are small from your point of view, but these nutrition experimentalists would give their eyes to get (say) 20 pairs of twins to experiment with and I should hope to get ten times as many in the two groups together.

If what the Reading people tell me is true there really is quite a considerable difference in favour of Raw milk, enough to show quite clearly in such an experiment.

But besides this ostensible object of the experiment one might hope to get quite a lot of very useful information on the Nature and Nurture question if details of the environment and parentage of the various pairs of twins were collected at the same time. Would it not be worth the while of the Galton Laboratory to father the scheme? and so get it done properly.

*so can this, given information.

The corrected proofs of Gosset's paper on the Lanarkshire milk experiment were sent to K.P. on 18 August, and the paper appeared in the issue of *Biometrika* for December 1931. Gosset was very chary of drawing conclusions from an experiment in which the groups of children taking raw and pasteurized milk were not random samples from the same population, but selected samples from populations which may have been different. He recommended that, in any repetition of the experiment on a large scale, the children should be formed into pairs of two, balanced with respect to age, sex, height, weight, and physical condition. The pairs should then be divided into 'controls' and 'feeders' by tossing a coin for each pair. He also suggested that 50 pairs of identical twins, divided in the same way, would give more reliable results for small fractions of the expenditure and trouble. Each design would now be described as randomized blocks for two treatments, but that terminology was not used.

In view of Gosset's rejection of the Fisher-Bartlett conclusion, some reply from Fisher might have been expected, but none seems to have come. However, other forces were at work. K.P. had been taken by Gosset's suggestion of an experiment with identical twins, and in their absence he suggested to Ethel M. Elderton that children be paired from the original cards. She submitted a paper to *Annals of Eugenics*, now edited by Fisher in succession to K.P., and Fisher asked Gosset to act as referee. He reported on 2 May 1934.

Here is my review of Dr Elderton's attempt on the Lanarkshire Milk Experiment. I doubt whether you will find much to disagree with in it.

There were 67 schools: I do not know how they were divided but if they were selected at random they should give you the evidence you want. I should not expect them to be so selected, for it seems to me likely that pasteurised milk would be delivered most easily in the town and raw in the country. Such evidence as there is (Elderton's difference between the two sets of controls not significant) is consistent with this hypothesis which would also tend to explain why the difference in favour

of raw milk is so small. I would not put any work into the experiment, if I were you, until satisfied as to the geographical distribution of the schools.

I have looked at my diagrams and it would seem that the loss of clothes amounts to about $\frac{2}{3}$ of the real gain in weight, anyhow in the case of the control girls.

Fisher was unable to see that the paper had thrown the least new light on the difference between pasteurised and raw milk, and he disagreed with Gosset's remarks about town and country, but notwithstanding these doubts Elder-ton's paper was published.

5.5 ADVICE AND ARGUMENT

Soon after Fisher's appointment in 1919 to study the records at Rothamsted Experimental Station, he applied to Gosset for advice on calculating machines. Gosset wrote on 19 September to recommend the Triumphantor 'which is an improved Brunsviga', and the Millionaire 'which another office favours'. He continued in typical Gosset style.

Personally I mostly use slide rule being very rarely able to accumulate enough figures to make it worth while to use a machine, but I always use a Brunswiga [*sic*] when dealing with logs.

This practical approach to his calculating needs was extended a few years later when he told K.P. on 10 December 1924 that he was thinking of making a multislide rule for dealing with multiple regression equations, or in fact any equation without product terms. He wrote to Fisher on the same topic two days later, and illustrated his remarks by a 'finger-drawn circle' with scales for the dependent variable y and explanatory variables x , z , $\cos v$, and $\sin w$.

The circle represents a wooden disc say 1' or 18'' in diameter, to which a paper or cardboard cover can be attached by paste or preferably some system of clipping.

It is free to revolve in a fixed frame, being pivoted on a central axis on which is also pivoted a transparent celluloid cursor. The whole screwed up with the necessary washers so that either disc or cursor can revolve without moving the other and will stay where it is left unless deliberately moved.

The fixed frame is made so as to form a level table with the disc and round the edge is the scale of y the variable to be predicted. (say)

$$y = a + bx + cx^2 + dz + k \log v + p \sin w.$$

Then setting a pointer on the edge of the disc to a on the scale you move the cursor to the zero of x on a suitable scale of x suitably placed on a circle concentric with the disc. Then you move the disc till the particular value of x comes under the cursor and so on. Finally the pointer shows you the answer on the y scale. The covers could have the circles printed on them and would only require scaling and when not in use could be kept in a gramophone [*sic*] record cabinet!

While preparing his talk on 'Errors of routine analysis' to the Society of Biometricians and Mathematical Statisticians, Gosset had another bright idea, essentially a practical aspect of the emergence during the 1920s of quality control charts for the inspection of manufactured product. He gave Fisher some details on 6 December 1926.

In the course of getting things ready I struck a method of producing a lecture diagram which is new to me and seems effective for its purpose. I wished to show how routine analyses are not random in time by putting spots on a time diagram with a straight line to show the mean. Not being a tidy draughtsman I funk'd putting 100 spots at $\frac{1}{2}$ " intervals by inking in circles but a brainwave told me to get black stickybacked paper; it is used to edge photos and by selecting a cork borer of suitable size I found I could cut through three thicknesses of paper at a time and get perfect circles which I could stick on at leisure and which would come off if misplaced without leaving much mark. The straight line was cut off the paper with a photo trimmer and the short lengths pieced together. Quite successful.

Much has been made of the sharp disagreement between Gosset and Fisher concerning balance and randomization which became public at a professional meeting in 1936. In fact, differences of opinion had occurred privately for ten years or more; while usually in a low key, the dispute could for a brief period be expressed *fortemente*. The first volume of *The balance of births and deaths* by R.R. Kuczynski was published in 1928. Fisher accorded the book a friendly review in *Nature* for 9 March 1929, but he was astounded to find in *Eugenics Review* that Gosset was much less favourably inclined, and in consequence Fisher 'had to blow off steam' with a rejoinder. The dispute arising from Egon Pearson's review of *Statistical methods for research workers* (see §6.5) was just drawing to a close with the publication of Gosset's letter in *Nature* on 4 July, and Gosset had this event in mind when he followed Fisher's example on 1 August.

If you really feel called upon to give that tripe another puff there is that in my recent record which prevents my objecting to your method of doing it. But you mustn't suppose I did anything inadvertently. The man has discovered one interesting and novel fact, not one of any *very* great importance, by which I mean that unless economic conditions change it is bound to become obvious sooner or later.

If you or I had spotted it we should have been tremendously excited, worked away at it (I daresay mine would have been quite superficial work but some I'd have done) and written a *short* scientific paper for one of the scientific magazines. That doesn't suit him, perhaps he can't help himself, he must write a book. And a book of that sort won't sell if you are merely scientific; you must be sensational. You may publish tables of crude birth rates, it shows how the effete Europeans are going to the dogs but tables of crude death rates, no it would weaken the case so you must change them into mortality tables which people won't notice compare rather too favourably with our own magnificent American record. When you have to say that at present the English birth rate is such that it is not enough to prevent a decline in population you say that the population of England is bound to die out etc. etc.

The blast continued for several more paragraphs, and, although Gosset ended by declaring 'I've blown off steam now, write what you will, I've said *all* I am going to say about the book', he nevertheless returned to the subject in two further letters. By this time, Fisher had regained his composure, and he closed this correspondence on 13 August, prefacing his remarks with a candid assessment of Gosset's stance:

You are the most persistent man alive, and write quite as though your original ebullition was as calm, considered and rational as could be, instead of being so fervently indignant and prejudiced as an Irish bishop.

After further correspondence touching on fertility contours in field trials, electrically driven calculating machines, and other matters, there is a gap during Fisher's visit to the USA in 1931. Correspondence resumed at the end of 1931. Early in 1932, after Fisher had visited Gosset at Blackrock, he enclosed with his 'thank you' letter of 26 January 1932 a paper (presumably CP 93) on the evolution of dominance. This suggests that their conversation during Fisher's visit had turned to evolutionary genetics. Gosset's next letter (16 July) begins:

It may be merely my ignorance but I get the idea from the little I read about genetics that quite a number of its exponents believe that our various hereditary troubles are conditioned by quite a limited number of 'genes' a piece. That being so it may be worth while to draw attention to a case where it may be difficult to find so simple an explanation.

Would you mind vetting this for me?

Gosset's enclosure drew attention to a paper by Floyd L. Winter on 'Continuous selection for composition in corn', published in the *Journal of Agricultural Research* for 1929. Winter's experiments extended continuously from 1896 to 1924. Selecting for oil content from a foundation stock, Winter produced two strains, one with a mean percentage of oil about *twelve* times the standard deviation of the original population above the original mean, the other about *seven* times below. At the same time, the variance of the oil content was little changed. Gosset concluded:

It does not appear that such steady progress could be obtained with less than hundreds of genes affecting oil content and it seems not unlikely that there may be thousands. . . .

And so we reach the conception of a species patiently accumulating a store of genes, of no value under existing conditions and for the most part neutralised by other genes of opposite sign. When, however, conditions change, . . . the species finds in this store genes which give rise to just the variation which will enable it to adapt itself to the change.

On Fisher's advice Gosset submitted his paper to the *American Naturalist*, but they reacted, as had the Royal Society to Fisher's 1918 paper on 'The correlation between relatives . . .', with rejection. Gosset's paper appeared in

the relatively low circulation *Eugenics Review*. Later, when Fisher became editor of the *Annals of Eugenics*, he invited Gosset to 'produce a paper for the *Annals*', and Fisher's mathematical ability greatly assisted Gosset in preparing his second paper on Winter's experiment, which appeared in the volume for 1934. Fisher meanwhile wished to draw wider attention to Gosset's work with a letter to *Nature*, and on 16 January 1933 Gosset wrote to Fisher:

When I persuaded you to write up the mathematics of myriad gene selection in *Nature* I was so pleased with the idea of having got it done properly, that I overlooked the fact that I have put you into the position of appearing to 'butt in'.

Gosset suggested a pair of opening paragraphs similar to those with which Fisher's letter [CP 106] begins, and went on:

And here I think I hear you murmur 'Damn the man why doesn't he refrain from teaching his granny. He's as fussy about his little bit of stuff as a hen with one chick'.

To which I reply 'I am, curse you; for the very good reason that I'll never have the chance to incubate an egg which interests me so much'.

Cluck. Cluck. Cluck. Cluck.

Fact is that until just recently I was so much taken up with the first part of the thing, 'myriad genes', that I overlooked the fact that the second is really an essential cog in the mechanism of Darwinian selection. For at least twenty five years I've been reading that the continued accumulation of infinitesimal variations can do nothing and all the time I've felt in my bones that Darwin was right.

Cluck. Cluck. Cluck. Cluck.

And now I have been vouchsafed a vision,—and am filled with insufferable conceit—for the nonce I too am among the prophets, a mere Obadiah, but still among the prophets. And if anyone were to offer to make me a Doctor of Divinity on the strength of it I'd accept with conscious pride and flaunt a scarlet gown through the scandalised streets of Oxford without the slightest embarrassment.

Bear with me, Fisher, laugh with me tonight: tomorrow—when I'm sane again—when I know that my little bit was discovered in 1896 and put into better words than mine often since then and when I have been shown that my essential cog will hardly ever fit into the machine and when it does is a clog—then I'll laugh with you—at myself.

Cluck. Cluck.

Yrs. v. sincerely,
W.S. Gosset.

Fisher's letter drawing attention to Gosset's work, and in addition giving the reference to Winter's paper, appeared in the issue of *Nature* for 18 March 1933, where it would have come to the notice of biologists throughout the world. Gosset's cog was indeed an essential cog, not a clog, though acceptance of it in the biological world was slow in coming. In Ernst Mayr's prologue to

The evolutionary synthesis (Harvard, 1980), referring to the 'tremendous impact' of Dobzhansky's *Genetics and the origin of species*—published in 1937—he writes:

Dobzhansky devoted the entire sixth chapter to natural selection. His treatment clearly reflects how strongly natural selection still had to fight for general recognition. Dobzhansky's presentation was particularly effective because he treated selection not merely as a theory but as a process that can be substantiated experimentally. . . . The results of the Illinois maize selection experiments for high protein and oil content were particularly impressive.

The importance of natural selection in evolution was one of the few matters on which K.P. and Fisher were agreed; in Chapter 7 below, it is noted that, about this time, Fisher wished to join with K.P. in proposing Gosset for election to Fellowship of the Royal Society. We may conjecture that, had this occurred, one of the grounds for his election would have been his contribution to our understanding of biological evolution.

6. Egon S. Pearson

6.1 HISTORICAL INTRODUCTION BY E. S. PEARSON

6.1.1 Prefatory remarks

Without doubt the four persons who played the greatest part in shaping my approach to the discipline of mathematical statistics were Karl Pearson (inevitably), R. A. Fisher, W. S. Gosset, and my friend and collaborator, Jerzy Neyman. My mathematical powers were only moderate, for my degree in Mathematics at Cambridge was based on taking Part I of the Mathematical Tripos in 1915, and then, after the First World War using war service at the Admiralty and Ministry of Shipping, plus attendance during 1919–20 at selected Part II lectures to complete my course for the B.A. degree.

This weakness in mathematics undoubtedly had certain compensations: it caused me to thrash out problems with greater thoroughness, and to use an innate capacity for visual presentation. For both these reasons, I was, perhaps, the better teacher of many of the students who came to learn statistics at University College, and better able to give help later on as an adviser over industrial quality control problems and during the Second World War to technical officers in the Services, when three members of my staff and I were attached to the Ordnance Board, Ministry of Supply.

My first serious study of statistical literature began after the First World War, when, at the age of 24, I read K. P.'s memoirs published in the 1890s in the Royal Society's *Philosophical Transactions*; I was particularly fascinated by his development of the system of frequency curves which have since been associated with his name. In 1921, I completed my post-war (1919–21) period at Cambridge, during which I had not only continued my statistical reading but also attended lectures by Sir Arthur Eddington and F. J. M. Stratton on the theories of errors and combination of observations. Further, I went to a short course of lectures on mathematical statistics given by G. U. Yule; this last I attended along with the later well-known agricultural scientist, F. L. Engledow.

In October 1921, with J. O. Irwin, I joined the staff of the Biometric Laboratory of K. P.'s Department of Applied Statistics; we were both Junior Lecturers receiving salaries of either £300 or £350 p.a. Our first year of apprenticeship consisted in

- (a) attending K. P.'s first- and second-year courses on statistics;

- (b) 'demonstrating', i.e. helping K.P.'s students in tackling the numerical examples, illustrating the theory of the lectures;
- (c) undertaking one or two research problems which K.P. suggested to us.

My development from an apprenticeship to an independent line of thought of my own, linked with that of Jerzy Neyman, had for its background the conflict of ideas and techniques between K.P. and R. A. Fisher, between what it is perhaps not inappropriate to term 'modern mathematical statistics Marks I and II'. But before trying to summarize the factors which I believe were most important in shaping my statistical philosophy—a blend of Marks I and II—I must try to set down what seems to me the essential characteristics of the Mark I statistical approach.

6.1.2 Large-sample statistical methods

The line of approach was influenced by the fact that the data to be analysed were collected in 'large samples'. As K.P. and Weldon had written in the first editorial notice at the head of *Biometrika*, 1 (1901), the journal will 'include memoirs on variation, inheritance and selection in Animals and Plants, based upon the examination of statistically large numbers of specimens'. In interpreting 'large-sample' data, no very critical study was necessary of the part played by probability theory in the inferences to be drawn from observations.

Certainly K.P. had read the nineteenth-century literature quite extensively. In the introductory lecture which he gave in November 1892 (reissued in *Biometrika*, 32, 89–100 (1941)) in his Gresham College series of end-of-the-day lectures, he wrote in the syllabus (see E. S. Pearson 1938: Appendix II):

LAWS OF CHANCE. *Being the elements of the theory of probability in its relation to thought and conduct.* Definitions and Fundamental Concepts. Importance of Definition. Relation between the present course and the two earlier ones on *Fundamental Concepts of Science* and on *Statistics*. Statistics and the laws of chance intimately associated with the foundations of knowledge. Controversies, Laplace, Quetelet, De Morgan, Stanley Jevons, Boole, Venn, Edgeworth. Books which may be consulted: Stanley Jevons' *Principles of science*, chaps x–xii; Venn's *Logic of chance*, chaps vi–xii; Edgeworth's 'Philosophy of chance', in *Mind*, 1884; De Morgan's *Formal logic*, chaps ix–xi, and *Essays on probabilities*; Whitworth's *Choice and chance*; Westergaard's *Die Grundzüge der Theorie der Statistik*.

It is noteworthy that, when K.P. was away sick in April 1893 and could not give his group of four lectures, he arranged for Venn and Whitworth, as well as Weldon and Rouse Ball to take his place.

However, when, very shortly afterwards, inspired by Galton and Weldon, he began to make his own contribution to the mathematical techniques needed in the field of 'large-sample' biometry, so exciting was the chase of

discovery that he made very little reference to these 'fundamental conceptions'. Indeed, in 1894, when discussing the fit of a normal curve to a series of dice tossings, we find him writing in a letter to Edgeworth: 'Probabilities are very slippery things and I may well be wrong, but I do not clearly follow your reasoning or illustrations.' (see the letter of 18 February 1894 of K.P. to F.Y.E. which I quoted in *Biometrika*, 52, 14 (1965)).

The biometric investigations involving the analysis of large samples collected in the study of heredity and in the search for traces of natural selection at work chiefly called for estimates from large 'field' samples, of population parameters. These estimates whether of means, standard deviations, correlation, or regression coefficients were associated with standard (or probable) errors which were functions of these parameters; but, when dealing with large samples, the statistician was not let down if he substituted in his expressions for the standard errors the sample estimate, e.g. used $s/n^{1/2}$ for $\sigma/n^{1/2}$, $s/(2n)^{1/2}$ for $\sigma/(2n)^{1/2}$, $(1-r^2)/n^{1/2}$ for $(1-\rho^2)/n^{1/2}$, etc. In his later lectures to statistical classes, he frequently drew attention to this point. However, he did not seem to realize that, to establish that a large sample was not heterogeneous, it was desirable, as Walter Shewhart was later to emphasize, to break it into rational sub-groups and apply small-sample techniques to test for homogeneity.

6.1.3 New theory required to handle small-scale experimental data

It was only when in 1906 W. S. Gosset came to University College with his problems of interpreting the results of small-scale experiments carried out in chemical analysis or in barley breeding, etc., for Guinness's Dublin brewery, that the limitations of the Mark I statistical techniques began to be brought into the open. When faced with such problems, the fundamental concepts concerning the part to be played by the theory of probability in drawing inferences from statistical data needed to be defined on a more logical basis than when the samples were large. Gosset himself whose original reading matter had been Airy's *Theory of errors of observations* and Merriman's *The method of least squares* was clearly under the impression that, had it been possible, he should introduce into his procedures the prior distributions of parameters. For example, see the remarks he makes on p. 36 of his paper on the 'Probable error of a correlation coefficient' (*Biometrika*, 6 (1906)), and in the P.S. of his letter to me of 25 May 1926 (No. I.5), where he suggests taking a uniform prior distribution of σ . It was only when Fisher took up the story that prior distributions were pushed out of sight in the development of 'modern mathematical statistics Mark II'. An excellent account has been given by B. L. Welch (*Journal of the American Statistical Association* 53, 777-8 (1958)) of the relation of Gosset's, Edgeworth's, and K.P.'s work to a theory

of inverse probability. I cannot improve on this and only add that, because Gosset's presentation of this subject was at times a little contradictory, in *this respect* I did not get much help from him.

The fact was that K.P.'s laboratories were not carrying out small-scale experiments, and, as far as I can recollect, no external graduate student came to him (apart from Gosset) asking for help over the interpretation of such data. In K.P.'s fields of interest, sound conclusions appeared to require the analysis of large-scale data. Thus, he jokingly remarked: 'Only naughty brewers deal in small samples!'

But to put a hypothetical question, had during those five years Student come to his Professor, asking him for suggestions as to how to test whether the standard deviations s_1 and s_2 in two small samples of observations—say with $n_1 = n_2 = 10$ —drawn independently from two different normal populations suggested a significant difference between σ_1 and σ_2 , what would have been the result? Would the latter have said 'No answer is possible', or would a mathematical test, analogous to Fisher's variance ratio F -test, have come out as the answer? But this question was not raised, probably because in the brewery work the values of the variances σ_1^2 and σ_2^2 were either known from accumulated past experience, or could safely be assumed to be equal.

For many purposes, the experimental layout using the differences of matched pairs served Student's purpose, and it was only just before Fisher's arrival at Rothamsted that Gosset began to realize the need for what was to become known as the analysis of variance. (See the discussion on pp. 382–90 of my article 'Student as statistician', *Biometrika*, 30 (1939).)

I have indeed wondered how K.P. dealt statistically with the errors which his students in the Department of Applied Mathematics must have collected when taking observations in the two small observatories which he had succeeded in having built in the early 1900s in the College quadrangle (see the Plate facing p. 183 of the second part of my obituary article, *Biometrika*, 29 (1937)!)

As a result of this scepticism about the value of small samples in his lectures of 1921, K.P. gave Student's derivation of $z = (\bar{x} - \mu)/s$, but did not lay as much emphasis on its importance as on that of the distribution of s^2 and of Fisher's later multiple-space derivation of the distribution of the correlation coefficient r .

6.1.4 My apprenticeship at University College, 1921–6

After the First World War, the Department of Applied Statistics, which had come into existence in 1911, the year after Galton's death, was nominally divided into

- (1) a Biometric Laboratory,
- (2) a Galton (or Eugenics) Laboratory.

There was no division geographically within the building nor, as far as research went, between the two laboratories, but, for some years after the First World War, funds for the Eugenics Laboratory were administered by a Galton Committee of the *University*, not by the College. There are a number of documents now lodged in the Galton and Pearson Archives in University College which set out the history of the division. All that concerns me here is to say that, for many years after 1921, practically no numerical data came my way, except as standard class examples for which ample data could be drawn from past volumes of *Biometrika*. The nine papers which I published in *Biometrika* during the years 1922–7 are given below:

- (1) July 1922, 14, 23–102. On the variations in personal equation and correlation of successive judgments.
- (2) July 1922, 14, 127–56 (with K.P.). On polychoric coefficients of correlation.
- (3) March 1923, 14, 261–80. The probable error of a class index correlation.
- (4) August 1923, 15, 89–108. Natural selection and the age and area theory of Dr J. C. Willis.
- (5) May 1924, 16, 196–8. Note on the approximation to the probable error of a coefficient of correlation.
- (6) December 1925, 17, 388–442. Bayes' theorem examined in the light of experimental sampling.
- (7) July 1926, 18, 173–94. A further note on the distribution of range in samples taken from a normal population.
- (8) July 1927, 19, 216–22. The application of the theory of differential equations to the solution of problems connected with interdependence of species.
- (9) July 1927, 19, 223–4. Further note on the linear correlation ratio.
- (10) In addition, I was responsible for the compilation of *Tracts for Computers* No. VIII (1922), 'Table of the logarithm of the complete Γ -function for arguments 2 to 1200, i.e. beyond Legendre's range'.

Of these nine papers, six, namely (1), (2), (3), (5), (8), and (9), and also No. (10), were on subjects suggested to me, by K.P., as was most of the research work carried out during those years in the Biometric Laboratory. Papers (4) and (8) were probably initiated by books given to me by K.P. to review; in No. (4), I perhaps showed some originality, but I was, as it were, leaping to the defence of my laboratory's hero, Charles Darwin.

Paper (6) on Bayes' Theorem owed its theme to K.P.'s 1920 paper, (*Biometrika*, 13, 1–6) on 'The fundamental problem of practical statistics'. In this paper, K.P. followed Bayes' billiard table approach. I refer to the doubts about the soundness of this paper below. It must have been in 1922 or 1923 that I started on the long piece of observational 'counting', based on the rather ingenuous idea of exploring Edgeworth's statement that his justification for assuming a uniform distribution of prior probabilities between 0 and 1 was his own rough personal experience. Of course, as W. F. Sheppard pointed out in the oral examination for my London D.Sc. degree held in 1925 or 1926, my collection of several hundred results, following a U-shaped distribution in fact 'proved nothing', because a subjective element had inevitably entered into the characters which I chose to count.

In so far as there was theory in these papers, it fell within the category of 'modern mathematical statistics Mark I'. However, a number of distracting thoughts, largely resulting from the study of R. A. Fisher's flow of publications, luckily began to stimulate my thinking on problems of statistical inference, coming under the heading of 'Mark II'.

6.1.5 Doubts as to the adequacy of K.P.'s 'large-sample' theory

I knew how much I owed to these years of apprenticeship, when I realized the breadth of K.P.'s vision and received the stimulus of his lectures, both in 1921–2, in personal discussion on my research problems and in the drill of table-making, and, perhaps a little indirectly, from his ten or so annual lectures on the 'History of statistics in the 17th and 18th centuries'.* But the time had come when it was necessary for me to go through the painful process of experiencing growing doubts in my earlier belief in parental infallibility! A number of events contributed to this, the chief of which I recall were the following.

(1) The criticism applied to K.P.'s paper 'The fundamental problem of practical statistics' (*Biometrika*, 13, 1–16 (October 1920)). Of this, he had sent me an offprint while I was still at Cambridge and at the time I saw no flaw. However, when, in the first year course in the second term of 1921–2 session, K.P. lectured on this topic, it is clear that there had already been criticism of the paper, to which he refers without being specific in the Miscellanea Note, pp. 300–1, of a later part of the same volume issued in July 1921. It was not till the May issue of *Biometrika*, 16, 189 (1924), when W. Burnside set the criticism on paper, that K.P. (*ibid.*, pp. 190–3) came out in print in his own defence. However, I remember that, after the lectures of 1922, Oscar Irwin and possibly others had discussed this point and concluded

*Published by Chas. Griffin & Co. (1978).

that K.P. had slipped up in his 1920 paper in equating two functions $\phi(\xi)$ and $f(\xi)$, which would in general not be identical.

(2) In 1922, Fisher's paper on the 'Mathematical foundations of theoretical statistics' had been published in the *Philosophical Transactions of the Royal Society, Series A*, 222, 309–68. The most disturbing thing in this for me was the claim that the fitting of frequency curves by *moments* was 'inefficient' compared with fitting by the method of *maximum likelihood*. Because Fisher's proof depended upon asymptotic theory, it was impossible to tell at what sample size the loss in efficiency mattered in practice. The problem was made no simpler because in those days it was only possible to derive the maximum-likelihood estimates of parameters by applying a long series of hopefully convergent approximations to the estimates derived from moments! Some fifty years later, it has transpired that there was after all ample justification for doubts about the validity of this particular asymptotic theory. This has been shown by a random sampling experiment from a Pearson type III population and theoretically by Bowman and Shenton (1970, Union Carbide Corporation Report CTC 28). Thus, with samples of $n=200$, Fisher's asymptotic variances are completely inadequate, although these variances are still less than the corresponding variances derived from a moment solution.

(3) In 1922 and 1924, Fisher published papers criticizing K.P.'s use of the chi-squared test for goodness of fit. Since, when there are k frequency groups, the continuous density function $e^{-\chi^2/2}$ is only an asymptotic approximation to discontinuous multinomial frequencies, the degrees of freedom rule could not be exactly verified.

(4) In 1924, E. C. Rhodes read a paper to the Society of Statisticians and Biometricians entitled 'On the problem whether two given samples can be supposed to have been drawn from the same population' (*Biometrika*, 16, 239–48). In this, owing to some confusion in his reference sets, he appeared to have shown that there were tests based on degrees of freedom $\nu = 1, 2,$ and 3 , all to be applied to an identical expression for χ^2 . K.P. followed this in *Biometrika*, 16, 249–52, 'On the difference and doublet tests for ascertaining whether two samples have been drawn from the same population'. In this, he put forward the suggestion that, 'if there exist two or more tests which may be applied with equal logical validity', e.g. using means or the standard deviations or third moment coefficients, the statistician 'will, I should say, always be guided in rejecting or accepting common origin by the most stringent of those tests'. 'Stringency' might be interpreted in various ways, but it appeared that K.P. meant that the statistician should take the verdict of the test, which when applied to the given data, gave what could be termed the smallest 'P-value'.

It struck me at the time that this guiding rule was not acceptable. Indeed, though it was two years before I began to look into the matter, I am clear

from notes which I wrote later, that K.P.'s suggestion played a considerable part in making me look for a more logical principle to follow in choice among alternative statistical tests. This led to the introduction of the likelihood ratio principle a few years later.

(5) In 1925, Fisher published his *Statistical methods for research workers*. It was not written in a form which made the new approach readily acceptable to a mathematician. At the International Mathematical Congress held at Toronto in 1924, he had read a paper establishing mathematically the relation between the normal, chi-squared t -, and z - (or variance ratio) distributions and the tables of percentage points given in this book. But this paper was not available in print until 1926 and I may not have seen an offprint until a year or two later. My first reference to the Toronto paper is given in a footnote to my paper on 'Some notes on sampling tests with two variables' in *Biometrika* 21, 338 (December 1929).

That I was not alone in finding *Statistical methods for research workers* difficult was recently illustrated by G.A. Barnard (*Lecture Notes in Biostatistics* 18 (Springer, 1977)). In his last year at school, early in 1933, Barnard tells us he was seeking advice on how to interpret a small statistical survey; he was put on to Fisher. When he remarked that he was interested in pursuing the subject further but could find no suitable literature, Fisher picked up a copy of *Statistical methods* and told him that, if he read it, he 'would find many statements which called for proof', but that being a mathematician he ought to be able to work out the proofs by himself: if he did so, he would have learned mathematical statistics. 'The next time,' Barnard remarks, 'that I met Fisher was nearly twenty years later . . . I was able to tell him that I had just the week before, more or less completed the task he had set me nearly twenty years earlier!'

For all these reasons, it is hardly surprising that in 1925-6 I was in a state of puzzlement, and realized that, if I was to continue an academic career as a mathematical statistician, I must construct for myself what might be termed a statistical philosophy, which would have to combine what I accepted from K.P.'s large-sample tradition with the newer ideas of Fisher. As K.P. had given the statistics courses at University College up till 1926, I had not been faced with the serious job of putting my conflicting thoughts into order. It was only in the autumn of that year, as a result of an operation for cataract undergone that summer, that K.P. wanted me to undertake some of the lecturing work on statistical theory.

Luckily for me I was not unduly worried by the prospect of mastering my difficulties. At this period, the study of statistics was not the be all and end all of my activities: I was young enough to throw myself eagerly into an April visit to Italy between penning my letter to Gosset of 8 March 1926 and receiving his reply of 28 May. Also, I had much in my mind the prospect of

a five weeks' sail in my cousin's schooner yacht among the lochs and islands of the north-west coast of Scotland!

It was a lucky thought which caused me to write my first letter to Gosset, whom of course I knew from his visits to K.P. at University College, latterly often on his way to Rothamsted or to his father's house at Watlington.

6.2 DIFFICULTIES ABOUT z AND χ^2

This section concerns the letters in Group I, written between 7 April and 27 May 1926.

6.2.1 Editorial introduction

Gosset often called at University College when passing through London either on business for Guinness or to stay with his father at Watlington. He was the natural person to approach about Pearson's difficulties regarding both t and χ^2 . The earliest letter from Pearson to Gosset which has been preserved (No. I.1) is dated 7 April 1926 but was not actually posted until 5 May, after a holiday in Italy. Pearson began by referring to a recent visit to the Fruit Station at East Malling in Kent. While wandering among the apple plots, he was suddenly struck with a doubt as to exactly what interpretation can be laid on $z = (\bar{x} - \mu)/s$. Here, \bar{x} is the mean and s the standard deviation of a sample of size n from a normal distribution with mean μ and standard deviation σ . His confusion arose because in two samples he might have a sample point A_1 with coordinates (\bar{x}_1, s_1) and another A_2 with coordinates (\bar{x}_2, s_2) such that A_1 was less likely to occur than A_2 , while z_1 was smaller than z_2 and so more likely to occur. He presented the problem using a diagram of the joint density function $f(\bar{x}, s | \mu, \sigma)$ which was necessarily drawn on the scale of the unknown σ .

What is the z distribution telling us? We can hardly use it as a comparative criterion it seems to me; given two samples as at A_1 and A_2 , we cannot say that A_1 is more likely to have come from a population with mean μ than is A_2 , because $z_1 < z_2$, unless we build up some hypothesis as to the a priori possible values of σ_1 and σ_2 . This one naturally shuns attempting.

Of course I am very likely trying to get something out of the z distribution that I should not, and you can put me right. But what I am beginning to feel is that you cannot really apply any test of goodness of fit or probability of random sampling unless you actually *know* your population constants [i.e. parameters], or are prepared to take the risk of their differing significantly from certain definite assumed ones. In small samples this risk is very large.

This was the period of the General Strike. Gosset wrote at 12.15 a.m. on 10/11 May (letter No. I.3) to say that 'between a pusillanimous Government and a punctilious trades union' Pearson's letter had only just arrived, and he replied in detail on the following day (No. I.4).

Now in point of fact all you are supposed to know comes under two heads:

1. that the population is normal,
2. that a given unique sample has an S.D. and a mean at a known point *in the scale of s*.

What we are asked is 'What is the chance that the mean of the population lies at any given distance z from this mean measured in this scale?'

When we come to draw the correlation surface (corresponding to the population) that you have drawn my attention to in the scale of σ we have no possible means of connecting up σ and s at all or even the point $\bar{x} = \mu$ *accurately* with the mean of the sample. All we can do is to find out what happens if the point $\bar{x} = \mu$ is at a given distance *in the scale of s* from the mean of the sample.

He continued by noting that the positions of the points A_1 and A_2 in Pearson's diagram were irrelevant because the positions of the actual samples were unknown, but that the volume under the surface $y = f(\bar{x}, s | \mu, \sigma)$ falling beyond the section along which z was constant could be calculated and used for a valid test. Gosset then proceeded to 'put the thing round the other way', and in doing so he introduced the concept of an alternative hypothesis and raised the question of sampling non-normal distributions, both of which were influential in fixing the direction of Pearson's future work. An extract from this part of Gosset's letter was quoted by Pearson in two publications (1939: Letter I; 1966).

Difficulties concerning the chi-squared test of goodness of fit form the other theme of letters in Group I. The test was established by K.P. in 1900 for examining the agreement between observed and expected frequencies in situations where parameters used to calculate the expected frequencies are specified by hypothesis. When parameters are estimated from the sample, he argued that conclusions of acceptance or rejection would be the same as when the distribution is known a priori. This view was challenged by Fisher in a series of papers published between 1922 and 1924, which modified the original test using the concept of 'degrees of freedom', a number which is reduced by one for each parameter efficiently estimated. Bartlett (1981) quoted from a letter of Egon Pearson dated 30 March 1979:

I knew long ago that K.P. used the 'correct' degrees of freedom for (a) difference between two samples and (b) multiple contingency tables. But he could not see that χ^2 in curve fitting should be got asymptotically into the same category ...

Egon Pearson's letter of 7 April 1926, in which his doubts about the

interpretation of z were expressed, concluded with an outline of similar difficulties in regard to χ^2 .

The same problem seems to me to occur in Fisher's method of attacking the χ^2 distribution. He considers the distribution that you would find if you were to take repeated samples and in each case put into χ^2 the values of constants calculated from the sample giving, say χ_2^2 , instead of the true population values, which would give, say, χ_1^2 . If his algebra and assumptions are correct he gets a definite distribution of χ_2^2 , different from that of χ_1^2 which we should find if using the true population constants. But what value has this distribution? I feel very uncertain; just as in your z distribution it seems to me that if the unknown population had actually a certain range of values for its constants, then there well may be many samples which while giving a greater χ_2^2 than other samples, will yet be samples of more frequent occurrence—as the true values χ_1^2 would tell us if we could get them. What the old method of approach does is to assume that our constants calculated from the sample do not differ far from the population values, so that the constants based on these, will not differ seriously from the true χ_1^2 distribution. I think this may lead to a systematic error, since in general $\chi_2^2 < \chi_1^2$ which is not generally at all serious, except when few groups are fitted by very elastic curves. But it seems to me that it would be more sensible to correct χ_2^2 by adding the mean difference $\chi_1^2 - \chi_2^2$ for samples with n categories, than to use χ_2^2 in the distribution for $n - p$ categories, a distribution which is to be interpreted—well I don't know how.

Gosset replied to these queries on 25 May (letter No. I.5), the delay being caused by his need to gain some understanding of K.P.'s basic paper.

I have now read the χ^2 paper in *Phil. Mag.* (1900) Vol. 50, 157. It may be divided into three parts, one that I can follow as a man who could cut a block of wood into the rough shape of a boat with his pen knife might appreciate a model yacht cut and rigged to scale, the second I can only compare to a conjuring trick of which I haven't got the key (such for example as the transformation to polar coordinates on p. 158) and lastly quite a small part which I think I can understand.

Fortunately, or rather unfortunately, there is a sentence at the bottom of p. 160 which I include in the third part and which seems to me to justify Fisher; admittedly there may be something in the whole bag of tricks on the next two pages which has deceived me but I will put it up to you (and suggest that you get hold of the classic itself):

'Further if $e = m' - m$ give the error we have

$$e_1 + e_2 + \dots + e_{n+1} = 0$$

Hence only n of the $n + 1$ errors are variables; the $(n + 1)$ th is determined when the first n are known and in using formula (ii) we treat only of n variables.'

Now as I understand it, what Fisher says amounts to this. I will take first the simplest possible case in which you have fitted, let us say, a Poisson to a set of observations using, naturally enough, the mean, for that purpose.

In this case you have not only

$$e_1 + e_2 + \dots + e_{n+1} = 0 \quad (a)$$

but also $e_2 + 2e_3 + \dots + ne_{n+1} = 0 \quad (b)$

since you have made the mean value of m equal to m' .

By an exactly analogous reasoning only $n-1$ of the $n+1$ errors are variables (should we not say nowadays independent?) the $(n+1)$ th being determined by equation (a) and n th from equation (b), when the remaining $n-1$ are known.

Further if the second, third or fourth moments are used in fitting the curve you get a third, fourth or fifth equation connecting the e 's and each time restricting the number that can vary independently by one.

If you allow this the rest of the proof seems to me to follow unaltered (though it is to me of the conjuring trick variety) and to produce Fisher's result.

Pearson's response on 28 May (letter No. I.6) was to describe in great detail the problem and his difficulties. He distinguished between χ_1^2 , calculated from s specified probabilities and associated with the density function

$$f(\chi_1) \propto \chi_1^{s-2} \exp(-\frac{1}{2}\chi_1^2),$$

and χ_2^2 , calculated after the estimation of c parameters and—under carefully prescribed conditions—associated with the density function

$$f(\chi_2) \propto \chi_2^{s-2-c} \exp(-\frac{1}{2}\chi_2^2).$$

His discussion of the relationship between χ_1^2 and χ_2^2 is complex, but two points seem clear:

- (1) he was unconvinced by Fisher's proofs;
- (2) he continued to accept K.P. on the matter of estimation.

In the circumstances, his conclusion was hardly surprising.

I have written this out at length, but it helps to clear my own head and I hope you will follow. In your letter, you show you have got to the first stage which I reached some time ago on first casually studying Fisher, when I thought he seemed to be probably justified. I think a lot of people come to that conclusion at first study, for certainly the Americans who came over here have got hold of the idea. That's the danger of it; I can't say whether those who proceed to the second stage and read all through Fisher and reason about logic and probability come to my conclusion or not, for I have not yet come across anyone else who has done so. I wish I could.

6.2.2 Comments by E. S. Pearson

I cannot recall now what was the form of the doubt which struck me at East Malling, but it would naturally have arisen when discussing there the interpretation of results derived from small experimental plots. I seem to visualize myself sitting alone on a gate thinking over the basis of 'small-sample' theory and 'mathematical statistics Mark II'. When, nearly thirty years later (*Journal of the Royal Statistical Society Series B*, 17, 204 (1955)), I wrote refuting the suggestion of R.A.F. that the Neyman-Pearson approach to testing statistical hypotheses had arisen in industrial acceptance

procedures, the plot which the gate was overlooking had through the passage of time become a blackcurrant one!

It is clear from letter No. I.1 that my approach to the z - (or t -) test was still under the influence of K.P.'s view that significance should basically be judged by referring a statistic to its distribution in sampling from a completely specified population, although in large samples the values of the parameters of the population, if unknown, might be replaced without much risk of error by their estimates from the sample. For this reason, in my state of confused thought, I *first* turned to the bivariate sampling distribution $f(\bar{x}, s | \mu, \sigma)$ represented in two dimensions and *then* to the distribution $f(z)$ of $z = (\bar{x} - \mu)/s$. I was puzzled because the two tests might rank the significance of two independent samples differently.

It is relevant to note that a similar criticism of Student's z -test was still confusing K.P. when in 1931 he criticized the use by Gosset of this test in connection with the Lanarkshire milk experiment (*Biometrika*, 23, 409–15 (1931)).

Gosset's answering letter, No. I.4, at once made the obvious and convincing point that my introduction of the distribution $f(\bar{x}, s | \mu, \sigma)$ was irrelevant since the essential nature of the problem was that σ was unknown. I was almost certainly not satisfied with the particular presentation which he gave supporting the use of z , but his letter left me with two fundamental ideas:

- (a) The rational human mind did not discard a hypothesis unless it could conceive at least one plausible *alternative* hypothesis.
- (b) It was desirable to explore the sensitivity of his z -test to departures from normality in the population, i.e. the question which was later to be termed by G. E. P. Box that of *robustness*.

There was one remark in his reply which I might usefully have taken up but did not: he wrote '... but we can bet on the probability that the mean of the population shall lie within any given distance of the known mean of the sample ...'. He was clearly using an inverse probability approach, which, as already mentioned, had appeared in his 1908 papers on 'The probable error of a mean' and 'Probable error of a correlation coefficient'.

Discussion between us on this matter in the summer of 1926 might have brought out the fact that the probability statements:

$$\Pr \{z = (\bar{x} - \mu) / s < -z_{\alpha/2}\} = \frac{1}{2}\alpha \text{ and } \Pr \{z = (\bar{x} - \mu) / s > z_{\alpha/2}\} = \frac{1}{2}\alpha,$$

where $z_{\alpha/2}$ is the upper $100 \times \frac{1}{2}\alpha\%$ point of the z -distribution for $v = n - 1$ degrees of freedom, can be inverted into

$$\Pr \{\bar{x} + s z_{\alpha/2} = \mu_1(\bar{x}, s) > \mu\} = \frac{1}{2}\alpha \text{ and } \Pr \{\bar{x} - s z_{\alpha/2} = \mu_2(\bar{x}, s) < \mu\} = \frac{1}{2}\alpha.$$

From these, we might have reached the statement

$$\Pr \{\mu_1(\bar{x}, s) < \mu < \mu_2(\bar{x}, s)\} = 1 - \alpha. \quad (\text{A})$$

If the calculation of the variable limits $\mu_1(\bar{x}, s)$ and $\mu_2(\bar{x}, s)$ could be regarded as a rule of behaviour to be followed when drawing a sample of size n from any normal population, then (A) would have provided Neyman's *confidence interval* for the unknown μ . Of course, redefining s^2 as $\Sigma_i(x_i - \bar{x})^2 / (n - 1) = \Sigma_i(x_i - \bar{x})^2 / \nu$ and substituting $t/\nu^{1/2}$ for z , the limits $t_{\alpha/2}$ could, in 1926, have been obtained from Table IV of Fisher's 1925 *Statistical methods for research workers*, for $1 - \alpha = 0.90, 0.95, \text{ and } 0.99$. But the underlying philosophy had not yet surfaced: that the consequences of a 'rule of behaviour' applied in the long run could influence judgement when applied to a single sample of observations. It is possible that it was because this philosophy was not found acceptable by Fisher that he introduced the idea of *fiducial probability*.

A similar line of reasoning could of course have been applied in 1926 to derive confidence limits for σ^2 based on s^2 , using Fisher's Table III of the percentage points of χ^2 . It is interesting to note that, although the foremost reason for computing and publishing percentage point tables of t and χ^2 may have been the reluctance of K.P. to allow Fisher to reproduce the *Biometrika* copyright tables of the probability integrals of these statistics,* the existence of such a new form of tables made it easier to illustrate and follow out in practice the rule of behaviour concept in deriving confidence limits. Thus, when in 1932, Waclaw Pitkowski, on the basis of Neyman's lectures in Warsaw gave a confidence interval for a mean based on the t -distribution, he was able to take $t_{.025}$ for $\nu = 4$ from Fisher's table of 1925 (or perhaps from the 1928 edition). Without this table he would have had to interpolate backwards in one of Student's tables of the probability integral of z (*Biometrika*, 1908 and 1917) or of t (*Metron*, 1925). Whether the availability of Fisher's type of tables in any way influenced Neyman in his putting forward his confidence interval ideas in his Warsaw lectures, I do not know.

No comments of mine replying to Gosset's letter No. I.4 have survived, but during the summer and autumn of 1926 I must have been turning over in my mind his suggestion regarding 'alternative hypotheses'.

He gave me his views on the χ^2 degrees of freedom controversy in his letter No. I.5, of 25 May, to which I replied three days later in No. I.6, a letter which gave me the opportunity of putting my difficulties on paper. Clearly, I was in a state of some puzzlement over this problem also. If, using some hindsight, I try to summarize these difficulties, the position appears as follows.

In the z -test problem my geometrical approach led to my presentation of the situation in the two-dimensional (\bar{x}, s) space. As I was familiar with K.P.'s 1900 paper in which he speaks of χ as being constant on hyperellipsoidal contours in a space of k dimensions, where k is the number of frequency

*Because of the critical financial position of *Biometrika*, K.P. was afraid that if some of its tables were published elsewhere, this would affect sales of *Tables for statisticians and biometricians*.

groups, it was inevitable that I should think of the grouped sample as represented by a point (n_1, n_2, \dots, n_k) in this space, subject to the single linear restriction, $\Sigma n_i = N$, the sample size. When the k population expectations \bar{n}_i were known, the P used in the test was therefore the integral of the probability density falling beyond the hyperellipsoid on which the sample point fell. But when the expectations m_i were obtained by fitting a curve to the sample frequencies, what logical reason was there, I asked, for using the integral

$$P' = \int_{\chi_2}^{\infty} f(\chi_2) d\chi_2 \quad / \quad \int_0^{\infty} f(\chi_2) d\chi_2$$

as the criterion of goodness of fit? Here the χ_2 and $f(\chi_2)$ are defined by (ii) and (v) of my letter No. I.6.

Assuming that asymptotically the distribution of χ_2 with its reduced degrees of freedom were correct if the m_i were obtained by a method of *maximum-likelihood* fit, there would also be a distribution of χ_2 when the fit was by *moments*, though this might be hard to derive, giving say an integral P'' . Except for convenience in reaching the simple integral of $f(\chi_2)$ with its reduced degrees of freedom, *why* I wondered instinctively would P' be a better guide to judgement than P'' ? This was the 'second stage', with its question to which I had not been able to find an answer.

It was only later after my introduction of the likelihood ratio principle that, in the geometrical terms which appealed to me, I realized that, if the fit was carried out by minimizing χ^2 , the P'' could be shown asymptotically to be an integral in the k -dimensional space outside the envelope of hyperspheres whose centres were constrained to move on a $k - c$ dimensional prime. It was in this way, with Neyman's help, in 1927-8 that I obtained satisfaction!

Again, I had no further correspondence with Gosset on this problem, and after my letters of May 1926 there is a six months' gap when the development of my thoughts went, as it were, into a tunnel—no written record having survived.

6.3 ORDER STATISTICS AND RANGE

This section concerns the letters in Group II, written between 16 April 1926 and 23 March 1927.

6.3.1 Editorial introduction

Galton's work on heredity led him to ideas of order statistics and percentiles. The first volume of *Biometrika* contains his memoir on 'The most suitable proportion between the values of first and second prizes', and, in a note which immediately follows, K.P. considers the problem of finding an expression for

the mean difference between the p th and $(p + 1)$ th order statistics in a sample of size n .

A further investigation of the problem, so as to study the distributions involved, was made in papers published in *Biometrika* for 1925 by J. Oscar Irwin and Leonard H. C. Tippett, who were both working in the Biometric Laboratory at the time. On 16 April 1926 (letter No. II.1), Gosset asked Egon Pearson for advice.

When I was over the other day I heard that you were looking into the distributions of the range for small samples and, as this seems to afford one solution of the problem of rejection or repetition of observations, I am writing in case you could help with my particular trouble.

I am concerned with the question of when it is advisable to repeat routine chemical analyses which have already been done in duplicate, either because it is the routine to do so or because the first result appeared to be remarkable. Further than this we have the same problem with triplets and quadruplets, the latter especially being a regularly occurring case of certain contract samples which are always analysed four times, and when repetitions are made we get quintets, hexets, etc. We keep in pretty close touch with the errors of our analysis and for this purpose we may consider that the Standard Error is known. There is, therefore, no great difficulty about the pairs (assuming normality of the Error distribution) as their difference belongs to a normal population of known Standard Deviation ($\sigma\sqrt{2}$).

In the case of triplets Irwin's (*Biom.* XVII 241) table gives a means of estimating the improbability of one analysis lying wide of the other two and if necessary I think I can calculate the constants of the range of samples of 3 since, if I am not mistaken,* the three differences $x_1 - x_2$, $x_2 - x_3$ and $x_3 - x_1$, when taken for an infinite number of samples of 3 give in the aggregate a normal population of differences and between that and Irwin's table on page 107 the thing can be got out by simple algebra. (Incidentally Irwin's mean value for the difference between 1st and 2nd individual, which he apparently took from your father's early paper as .8458 should, I think, be .8463 ($= 3/2\pi^{1/2}$), but the difference is not material.)

But when we come to samples of 4, or, when further repetitions have been made, of 5 or 6, there are no published tables giving either the range distribution or that of the outside interval and if you have calculated any such I should be very glad to have them.

*As I may be; I should like your opinion as to this.

He wrote again the following day (letter No. I.2) to say that he had 'stumbled on a theorem', connecting $\chi_{n,p}$, the mean difference between the p th and $(p + 1)$ th order statistics, and r_n , the mean range, both for samples of size n . The remainder of the correspondence in Group II established that the theorem was useless for his purpose, and concluded with numerical aspects of fitting Pearson curves.

6.3.2 Comments by E. S. Pearson

Gosset always studied papers published in *Biometrika* with a view to seeing whether any of the theory or tables which they contained would be of help to him in the analysis of data which had to be interpreted in the Dublin brewery. In the three papers (a), (b), and (c) listed on p. 87 below, the theory assumed that (i) the population sampled was normal and (ii) that its standard deviation was known. As explained in his letter II.1 of 16 April, he was concerned with a problem of routine chemical analysis where it was desirable to discard what might be termed outlying observations, before estimating a mean, and also where there was sufficient past evidence to provide what could be regarded as a sufficiently accurate value of the standard error σ of well-carried-out analyses. The number of observations in a sample might be only two or even as large as six or more.

The procedure which he would like to carry out was evidently that which he was to advocate later on pp. 161–2 of his paper on 'Errors in routine analysis', comparing the range in the complete sample with the assumed known standard deviation and discarding successive observations as outliers until the ratio of the range of the remaining $n-1$, $n-2$, ... to σ fell below the prescribed upper limit of the distribution of w/σ .

Irwin in papers (a) and (b) had given approximate sampling distributions of the distance between the first two order statistics, e.g. of $\chi_{n,1}$ and $\chi_{n,2}$, in terms of σ , but had gone no further. Tippett, in paper (c), had computed the mean values of w/σ to five decimal place accuracy for $n=2(1)1000$ but—and only to less accuracy—the standard deviation of w/σ and $\beta_1(w)$, $\beta_2(w)$ for $n=2, 10, 20, 60, 100, 200, 500, 1000$.

If his proposed outliers test was to be applied, Gosset needed to know the distribution of w/σ for $n=3, 4, 5$, at least. In his letter II.2 of 17 April, he thought that he had found a solution by expressing the mean intervals $\chi_{n,p}$ between the p th and $(p+1)$ th individuals in terms of the mean ranges r_n, r_{n-1}, \dots which Tippett had computed (in a later notation r_n was written as $E(w_n)$, etc.). He determined the relations for $\chi_{n,1}, \chi_{n,2}$ and in a typical 'Student manner' guessed that these could be generalized to

$$\begin{aligned}\chi_{n,p} &= \frac{n!(-1)^{p-1}}{2p!(n-p)!} \left(r_n - n r_{n-1} + \dots + \frac{(-1)^q}{(n-q)!} r_q + \dots + (-1)^p r_{n-p} \right) \\ &= \frac{1}{2} \frac{n!}{p!(m-p)!} (-1)^p \Delta^p r_{n-p}.\end{aligned}$$

I established this 'guess' and at Gosset's suggestion put it in a note added on p. 193–4 of my paper (d), referred to below, using the w of Tippett in place of the r of Gosset's letter.

As stated in his next letter to me, II.3 of 28 April, Gosset quickly realized

that the difference equation led him nowhere, as he could not determine from it the higher moments of the $\chi_{n,p}$ which he required to estimate distributions of $\chi_{n,p}$ needed to give the critical limits wanted for his 'discard' procedure. However, by this time, my paper (d) published in the July 1926 issue of *Biometrika* was at press, so that Gosset was able to fill the gap left by Tippett at $n = 3, 4, 5$, and 6 , to fit Pearson curves using mean w , and the approximate $\sigma(w)$, $\beta_1(w)$, and $\beta_2(w)$; from these, he obtained the significance levels γ for w/σ which he used in the proposed outlier technique of his paper on 'Errors in routine analysis'. The attempt described in letters II.1-6 therefore became irrelevant, but the incident illustrates Gosset's way of attempting to adapt published theory and computation of others to a practical brewery problem.

The letters Nos. II.8, 9 of March 1927 show that I was already wanting to use his empirical curves to get approximations to the distribution of range in samples of $n=10$ (see E. S. Pearson and N. K. Adanthāya, *Biometrika* (December 1928), last line of p. 357 and Table I for my first account of experimental sampling work in hand in the Department of Statistics during 1927-8).

A more extended investigation into the distribution of range was published in a later paper (*Biometrika*, 24, 404-17 (1932)).

Papers published in Biometrika with which letters II.1-8 are concerned

- (a) J. O. Irwin (1925). The further theory of Galton's Individual Difference Problem. *Biometrika*, 17, 100-128. He here derived the moments of the differences between the p th and q th 'order statistics' in samples from a normal population. The unit of measurement was the population standard deviation σ .
- (b) J. O. Irwin (1925). On a criterion for the rejection of outlying observations. *Biometrika*, 17, 238-50. Again σ must be known, and he gave a warning that in substituting the sample s for σ , error would be involved if the sample was small. Three numerical examples were provided, the first being 15 astronomical observations taken from Chauvenet's *Astronomy*; the other two examples contained 17 and 424 observations, respectively.
- (c) L. H. C. Tippett (1925). On extreme individuals and the range of samples taken from a normal population. *Biometrika*, 17, 364-87. Although this problem had received some previous consideration, Tippett's treatment both in theory and in the use of computational procedures and in giving checks by random sampling was more thorough than any previous work. Besides providing a table of the mean range to five decimal places for $n = 2(1)1000$ in units of the population standard deviation σ , he computed values of $\sigma(w)$, $\beta_1(w)$, and $\beta_2(w)$ for $n = 2, 10, 20, 60, 100, 200, 500, 1000$ to three decimal places, remarking that in the case of the β -values little reliance could be placed on the final figure. His table of mean ranges

opened the way for the use of a single range or the mean range of several equal-sized samples, in industrial quality control problems.

A further feature of his work was that he prepared the 'Table of random numbers' published in 1927 as the Department's *Tracts for Computers* No. XV, which gave a great spur to research involving simulation sampling.

- (d) E. S. Pearson (1926). Further note on the distribution of range in samples taken from a normal population. *Biometrika*, 18, 173–94. The main purpose of this paper was to supplement Tippett's values of $\sigma(w)$ at $n = 3, 4, 5, 6$, and to compute estimates of $\beta_1(w)$ and $\beta_2(w)$ at these four values of n and also to provide greater accuracy for these three moment values at sample sizes for which Tippett had provided results.
- (e) Student (1927). Errors in routine analysis. *Biometrika*, 19, 151–64. On p. 162 a table of approximate 10, 5, and 2% points of w/σ for $n = 2(1) 10$ is given. The points for $n = 2(1) 6$ and 10 were derived by quadrature applied to Pearson curves having the moments given in paper (d), while those of $n = 7, 8, 9$ were obtained by interpolation.

6.4 RANDOM NUMBERS AND SAMPLING EXPERIMENTS

This section concerns the letters in Group III, written between 13 May 1927 and 7 December 1928.

6.4.1 Introduction by E. S. Pearson

6.4.1.1 *Computation and random numbers*

It will be useful to preface a discussion of these Group III letters by giving some fuller details on the work of the Biometric Laboratory during the period 1920–5. The southern half of the Bartlett Building along Gower Street, completed in 1914, had been assigned for the use of the newly created Department of Applied Statistics, but when war broke out it was handed over to the University College Hospital for the use of wounded soldiers. It was not until 1920 that it became available to meet its original purpose. Still maintaining the division between a Galton Eugenics Laboratory and a Biometric Laboratory, it is of interest to record some of the research output falling under the second heading.

Starting from mathematical tables undertaken by K.P. in connection with work on bomb and anti-aircraft shell trajectories in connection with work for the Admiralty and Ministry of Munitions, he had the idea of issuing a series known as *Tracts for Computers*. Included under this title were some of

the preliminary sections of A.J. Thompson's immense project of *20-figure logarithm tables*. In addition, some ten to twelve *Tracts* on a great variety of subjects had been completed by 1925. The *Tables of the incomplete gamma function*, the computation of which had long been in hand, was finally completed and published by H.M. Stationery Office in 1922.*

Work on the computation of the 'sister' *Tables of the incomplete beta function* may be said to have started seriously with H.E. Soper's exploration of computational methods in *Tracts for Computers* No. VII of 1921, although the final table, involving much cooperative effort was not published until 1934.

Other work undertaken during these active years, 1920–5, by the Biometric Laboratory, its staff, and research students can be traced in the pages of *Biometrika*; it included Irwin's two papers on the distribution of intervals between 'order statistics' already referred to on p. 86 above; several papers attempting to develop frequency surfaces which were not bivariate normal; and a good deal of attention to the sampling moments of moments, reaching down to results for small samples. In the latter connection, there were tables of mathematical functions, but no attempts to apply these results.

It was in the middle of the 1920s that a new line of research was opened out. This is illustrated by the publication of two papers:

- (a) A. E. R. Church's papers in *Biometrika*, 17, 79–83 (1925) and 18, 321–94 (1926); in the latter were carried out random sampling experiments to test whether Pearson curves with the correct first four moments would fit distributions of variance in samples of size 10, obtained experimentally from normal and non-normal populations;
- (b) L. H. C. Tippett's paper in *Biometrika* 17, 364–87 (1925). In this, the distributions were tackled of the extreme individuals and of the range ($w = x_{(n)} - x_{(1)}$) in samples of sizes $n = 2, \dots, 1000$ from a normal population.

This work on the distribution of the range was to be of great value in connection with industrial quality control and for various uses in the application of what J. W. Tukey called 'quick and dirty' methods. What was common to both these lines of research, (a) and (b), was the need for some relatively speedy method of drawing artificial random samples from a specified population. Drawing small cardboard tickets had been tried by Student in 1906, but had not been found altogether successful because of the difficulty of adequate shuffling between successive draws (with replacement). Both Church and Tippett found the same thing; the former had then tried using coloured beads, each colour representing observations from a particular

*For a historical account of the development of this project, planned in 1903 or 1904, see pp. vii–ix of K. P.'s Introduction to the Tables.

population frequency group, but bias was introduced because beads of different colours were not all quite of the same size.

It was then that K.P. suggested to Tippett the idea of preparing a list of 'random numbers'. These numbers were subsequently published as *Tracts for Computers* No. XV in 1927, but had been used by both Tippett and Church while still in manuscript form. In his foreword to this *Tract*, K.P. wrote:

A very large amount of labour has been spent in recent years in testing various statistical theories by aid of artificial random samples; in many cases the theory itself may only be approximate or it may be mathematically correct when $N \rightarrow \infty$, where N is the size of the sample. In the former cases it is desirable to have some practical experience of the degree of approximation and in the latter case to ascertain what value of N for statistical practice may be considered to approach infinity. Occasionally the two desirabilities are combined as when we assume a given correlation table to be practically a [bivariate] normal distribution and determine the probable error of its coefficient as if that coefficient in small- or even moderate-sized samples followed a normal law of distribution.

The second paragraph of the foreword starts:

In order to get over the difficulty of random sampling for experimental purposes in the Biometric Laboratory, its Director suggested to Mr L. H. C. Tippett, when he was struggling with 'ticket' sampling, that he should replace the whole system of tickets by a single random system of numbers ranging from 0000 to 9999. These numbers, if truly random, could be used in a great variety of ways for artificial sampling. In order to form this table of random numbers 40,000 digits were taken at random from census reports and combined by fours to give 10,000 numbers [of four digits].

To illustrate how the numbers might be used, K.P. gave three examples:

- (a) drawing samples of $n = 10$ from a grouped normal distribution;
- (b) drawing samples of 100 (with replacement) from a 2×2 table;
- (c) sampling from a 5×6 contingency table.

6.4.1.2 *Sampling experiments and studies of robustness: the contribution by Guinness's brewers*

The existence of these random numbers opened out the possibility scarcely dreamed of before, of carrying out a great variety of experimental programmes, particularly of answering in considerable depth and breadth the kind of questions about the robustness of the 'normal theory' tests based on z (or t), s^2 , r and χ^2 raised by Gosset in his letter to me of 11 May 1926, (No. I.4). This programme I started on in 1927 and results began to appear, as they became available, in *Biometrika* papers published between 1928 and 1931. In the sampling and computation process, I had help from a variety of workers, but in particular from N. K. Adyanthāya, a postgraduate student

who combined a two-year statistics course (1927–9) with the sampling work with me.

The letters between Gosset and myself put together in Group III make many references to this programme; of particular interest to Gosset was the piece of research sampling undertaken by G. F. E. Story, to be published at the end of 1928 under the pseudonym of 'Sophister'.

Following the example of sending Gosset to London for part of the University Session 1906–7 to study statistics under K.P., Arthur Guinness Son & Co. Ltd., no doubt at Gosset's suggestion, had sent Edward Somerfield in 1922 to work with R. A. Fisher at Rothamsted. Then, in 1927, it was suggested that George F. E. Story be sent to work for a good part of the session 1927–8 under K.P. at University College. For Story, as in the case of Gosset, it was arranged that attendance at lectures and gaining experience from working out associated numerical examples should be combined. Probably because the firm did not wish it to be known by rival brewers that they were training some of their scientific staff in statistical theory and its application, these men were only allowed to publish under pseudonyms. Just as Gosset had been 'Student', Somerfield was 'Mathetes' and Story was 'Sophister'. Two other members of Guinness's staff, E. L. Kidd and Launce McMullen attended a statistical course at University College and A. L. Murray worked under Fisher. In the 1920s, very few scientists in industry made use of mathematical statistical methods: there was this large group at Guinness's, L. H. C. Tippett at the Cotton Industry Research Laboratories, and Bernard Dudding in the Research Laboratory of the General Electric Co. But no doubt there were others whose existence I have overlooked.

Gosset's letters to me contain a number of references to the help he received from Somerfield in producing certain data for me and in checking numerical calculations. But, in the case of Story, while the arrangements for his visit and general supervision of his work were made between K.P. and Guinness's, the detailed planning of the research project he was to undertake was left for discussion between Gosset and myself.

After first considering whether he should draw random samples from a symmetric, leptokurtic distribution, probably represented by a Pearson type VII (or Student) curve, we ultimately agreed on the choice of a skew type III or gamma distribution, with $\beta_1 = 0.50$ and $\beta_2 = 3.75$. Since, to introduce the random number sampling process, it was necessary to break up the population into a large number of equally spaced frequency groups, the beta coefficients of the grouped population actually sampled could not be made to agree exactly with those of the continuous type III distribution—in fact, Story's population histogram had values of $\beta_1 = 0.49$ and $\beta_2 = 3.72$.

For one thousand samples of $n = 5$ and a thousand of $n = 20$, Story derived the empirical distributions of means \bar{x} , variances s^2 (and s), range w , and Student's $z = (\bar{x} - \mu)/s$.

6.4.2 Edited version of E. S. Pearson's comments

In 1927, Pearson had been put on a small committee of the British Association for the Advancement of Science. He then made his first contact with Fisher, who was the secretary. A report was produced entitled 'Biological measurements; recommendations for the taking and presentation of biological measurements, and to bring such before persons or bodies concerned'. In drafting the report, the question arose of whether the sum of squares about the mean should be divided by n or $n-1$. Pearson asked Gosset for his opinion, and on 13 May 1927 (letter No. III.1) Gosset suggested the following statement for the text, not in a footnote.

The Standard Deviation has been defined both as

$$\left(\frac{S(x-\bar{x})^2}{N-1} \right)^{1/2} \quad \text{and as} \quad \left(\frac{S(x-\bar{x})^2}{N} \right)^{1/2}$$

and both formulae are to be found in different textbooks. The former is calculated to give a mean value which is independent of the size of the sample and so that of the population. On the other hand with large samples the difference between the formulae is quite negligible while the arithmetical work in the calculation of further statistics is much simplified. On the whole it is probably better to use $N-1$ with small samples and N with large. The important thing however is to state clearly which formula is being used in each case.

Pearson's presence on the committee had been suggested to Fisher by Gosset, who referred to the question when writing to Fisher on 1 June 1927 (Guinness Collection, No. 83).

I hope you and E.S.P. have managed to 'find a formula'. He wrote me quite a nice letter in which, as it seemed to me, he appreciated your point of view quite as well as his own, but I have not heard what you thought of (his version of) my suggestion.

However the relations between Fisher and Pearson in 1927 might have developed, harmony was rudely shattered by Pearson's 1929 review of *Statistical methods for research workers*, discussed in § 6.5.

The frequency of the exchange in Group III arose from the fact that a number of the letters were concerned with the research programme which Story had in hand. Pearson wrote on 28 July 1927 (letter No. III.2) about the problem on which Story should be put when he came to University College in October, and suggested a study of the distribution of Student's z in samples from a skew population, later specified by a type III curve. Gossett used the results coming out of Story's research to emphasize in several letters (9 November 1927 and 4 January, 14 April, and 21 May 1928; Nos. III.7, 9, 13, 15) that z (or Fisher's t) should be used as a two-tail test. This view was in accordance with the table of percentage points of t , published first by Fisher as Table IV on p. 137 of the 1925 edition of *Statistical methods for research*

workers. Gosset's explanation of his view was based on a folded-over version of the z -distribution presented as follows on 4 January 1928 (letter No. III.9) and again on 14 April and 21 May 1928 (letters Nos. III.13, 15).

As the population is unknown, skewness can have no definite direction with me and it is reasonable to suppose that in the long run the skewness will go one way as often as the other so that I must consider Story's population in conjunction with the corresponding population with the opposite skewness. To do this I merely have to reverse the sign of the z 's and add the new frequency distribution to the old and it is this combined frequency which I want to compare with Student's integral.

Pearson doubted at first (12 April and 19 April 1928; letters Nos. III.12, 14) whether this explanation gave an adequate answer to many problems but eventually agreed (6 October 1928; letter No. III.19) that he had misunderstood the point.

Story presented an interim report to Gosset on 19 January 1928 (letter No. III.10) and described on 4 July 1928 (letter No. III.16) how his research was being concluded. The results were published in December 1928 (*Biometrika*, 20A) under the pseudonym of 'Sophister' and put forward Student's view in the following terms.

So if we use 'Student's' Tables to decide whether a given sample has been drawn from a population of which we only know the mean, these results suggest that we shall be right *in the long run* provided that the skewness of the population does not exceed that used in this paper.

Karl Pearson added a footnote to this sentence.

Supposing 50% of prisoners tried for murder were acquitted and the remainder found guilty, should we be right *in the long run* to drop the trial and toss up for judgement? Ed.

Gosset's letter of 18 May 1929 (No. III.24) was partially reproduced as Letter II in Pearson (1939). The concluding paragraph refers to K.P.'s footnote, and Gosset enclosed a suggested note for *Biometrika*, reproduced with Letter II, written in reply to his mistaken criticism. There is no record of whether Pearson showed the draft to K.P. but, as Gosset anticipated, the matter was taken up in a joint paper by Pearson and Adyanthāya published in December 1929 (*Biometrika*, 21), which refers to Student's views on pp. 262 and 274.

During the period of Story's investigation, work on the robustness of Student's z in samples from rectangular and triangular populations was published by Shewhart and Winters of the Bell Telephone Laboratories in *Journal of the American Statistical Association*. Gosset's letter of 4 July 1928 (No. III.17) drew Pearson's attention to their results and made brief comments. This is the first mention of Walter A. Shewhart, who exerted a friendly influence on Pearson and was invited to lecture in London after Pearson's American visit of 1931.

The correspondence in Group III ran parallel to Pearson's joint work with Neyman and so was naturally directed towards questions of statistical inference. Pearson's letter of 28 July 1927 (No. III.2) referred not only to the robustness of Student's z but also to the distributions of two criteria provided by the likelihood ratio approach. They were z' , appropriate for shifts in the midpoint of a rectangular population, and z'' , appropriate for shifts in the start of an exponential population, in both cases when the parameters measuring the population variability were unknown. Inspired by Gosset's seminal letter of 11 May 1926 (No. I.4), Pearson had in mind an ambitious programme of sampling from skew distributions, but his next letter (1 August 1927, No. III.3) shows that Gosset had pointed out that the brewers were concerned in their routine analyses with long-tailed symmetric distributions. A consequence was that Adyanthāya was put on to sampling from type VII (i.e. symmetrical IV) curves, and Pearson asked (1 November 1927, letter No. III.4) how large β_2 should be. Gosset replied on 6 November (letter No. III.5) with information from Somerfield and suggestions about the sampling procedure. On 12 April 1928 (letter No. III.12), Pearson discussed the distribution of z' and described Adyanthāya's sampling results for symmetrical populations other than the rectangular. After Neyman and Pearson published their paper in *Biometrika* for July 1928, Pearson wrote on 6 October (letter No. III.19) to acknowledge that Gosset had suggested the idea of an alternative hypothesis and took the opportunity of setting out more fully the Neyman and Pearson approach, as it had been developed at that date.

The derivation of the sampling distribution of z' provided Pearson with a lead to the comparison of the sensitivity of alternative tests of a statistical hypothesis by a study of what Neyman and he afterwards termed their power functions. He worked on this line of approach with Adyanthāya's help using experimental sampling, and the results were published in their joint paper mentioned above. Here, Tables V and VII record the first, largely empirical, power functions for z and z' , respectively. Such experimental and visual approaches came to Pearson naturally and in fact it was only through them that he could see what were the mathematical problems to solve. The mathematician Neyman considered that real progress had to be made in another way, and, in retrospective articles concerned with the difficulty of mathematical research, he seems to have overlooked the fact that their joint work during the years 1927–30 led to ideas that were fundamental to their progress, for example

- (i) the concept of the likelihood ratio principle led to the derivation of 'good tests' in situations where no uniformly most powerful test existed, i.e. it led to answers where the results of the 1933 *Philosophical Transactions* paper could not be applied;
- (ii) the essential study of the robustness of 'normal theory' and other tests;

(iii) experimentation which alone perhaps made it possible to grasp the concept of power.

In this way, the contributions of Neyman and Pearson were complementary.

Pearson's letter of 7 November 1927 (No. III.6) introduced in a crude visual way a method by which the problem of interval estimation might be tackled. He made a vague and undefined suggestion of some form of inverse distribution for the shape coefficient β_2 based on the sample value b_2 . The method resembles the standard procedure of finding a 'confidence distribution' but was applied to a problem which remains unsolved. Pearson made a cardboard model representing the visual concept which was used in lectures at University College and in 1931 at the University of Iowa.

Gosset's letter of 18 May 1929 (No. III.24) has perhaps the widest cover of any which he sent to Pearson. It marks the end of Group III, in the sense that the discussions regarding Story's work and Pearson's experimental sampling exploration are wound up. Gosset emerges as a practical man who combined probability measures derived from application of what were the theoretically correct methods for testing a statistical hypothesis or for estimation, with other considerations—vague prior knowledge, economic limitations, and approximateness of the mathematical model.

6.5 REVIEW OF 'STATISTICAL METHODS FOR RESEARCH WORKERS'

This section concerns the letters in Group IV, written between 8 June and 30 September 1929.

6.5.1 Editorial introduction

Statistical methods for research workers was first published in 1925. Twenty-five years later, Yates (1951) described the reception of the book.

As is only to be expected with a book that marks such a fundamental break with tradition, its full significance was not immediately recognized. Nevertheless the reviewers of the first edition did perceive that the book was an important one, and they confined their criticisms mainly to lack of due deference to authority and to questions on intelligibility and presentation.

One of the reviewers was Egon Pearson, in *Science Progress*, and when the second edition appeared in 1928 he reviewed the book again, this time in *Nature*. There followed a chain of twenty-eight letters extending over four months of 1929.

The review (8 June) acknowledged the need for small-sample methods and pointed to Fisher's considerable extension of statistical theory in that direction. However, the book was criticized for not sufficiently emphasizing the

assumption of normality on which many of the methods were based, and Pearson questioned the stress on the 'exactness' of tests when they became more or less inexact for populations which diverged from normality.

That the tests, for example, connected with the analysis of variance are far more dependent on normality than those involving 'Student's' z (or t) distribution is almost certain, but no clear indication of the need for caution in their application is given to the worker. It would seem wiser in the long run, even in a text-book, to admit the incompleteness of theory in this direction, rather than risk giving the reader the impression that the solution of all his problems has been achieved.

Fisher sent a letter of complaint to the Editor of *Nature* and on 17 June this was forwarded to Pearson for comment. The letter has not survived, but Fisher seems to have regarded the review as offensive, and to have taken particular exception to the remarks about 'exactness'. Pearson drafted a reply (18 June), in the course of which he quoted from an article by Tolley (1929), who had warned that standard analyses of economic data were limited by the fact that most of the frequency distributions were not normal, but that new methods were now available.

Recently, however, the English school of statisticians has developed formulas and probability tables to accompany them which, they state, are applicable regardless of the form of the frequency distribution. These formulas are given in R. A. Fisher's book, *Statistical methods for research workers*, published in 1925.

At this stage, Gosset visited University College, and Pearson showed him Fisher's letter. Gosset volunteered to write to Fisher, and much of what followed consisted of exchanges between these two, with Pearson on the sidelines but kept in touch by Gosset.

When he opened the exchange (19 June), Gosset expressed profound disappointment that Fisher's letter to *Nature* was rather intemperate, drew attention to Tolley's article, and took the line that Pearson was regularly asked about how far Fisher's methods applied in non-normal samples, had a real appreciation for Fisher, and was genuinely trying to work with him in spite of the delicate position *vis-à-vis* Karl Pearson. Fisher replied (20 June) to say that the reason he was annoyed was not because the review expressed doubts concerning the accuracy of his methods when applied to non-normal data, but because of the implication that this accuracy was what he claimed, and he asked whether there was any basis for Tolley's statement in Gosset's writings. After expressing annoyance at some length, he concluded more calmly.

I do not think an agreed statement should be at all impossible, but I am not writing to E.S.P. until I hear from you again, as I want to be sure that you have a glimmer of what I am driving at.

In his second letter (24 June), Gosset brought another character into play, his assistant Edward Somerfield, who had been a pupil of Fisher at Rothamsted and had corrected the proofs of his book. Asked to express an opinion, Somerfield thought the review was good, and was shocked to find how much he ignored the question of normality. Gosset's letter transmitted this information, rebutted Fisher's charge that Pearson had imputed dishonesty, noted that Fisher had forgotten to say 'one word' about how 'exactness' depended on normality, advised Fisher that his business was to solve the practical problem of how much non-normality mattered, explained that Gosset was innocent of misleading Tolley, and ended by suggesting how Fisher's letter to *Nature* could be modified so as to put his point of view from Gosset's angle.

Fisher's reply (27 June) opened and closed peacefully.

Why do you not write to *Nature* and let me withdraw my letter? I imagine you could say what would be better than nothing from my point of view, and harmless to all others, ... Anyhow see what you can do, if you will. I do not think I like unpleasantness more than you do.

However, much of the letter was far from peaceful, and the reference to Somerfield was evidently unwelcome.

What has Somerfield to do with it? As I understand it there is a α -Somerfield who would not give too hoots for normality, and β -Somerfield who is shocked at his ignorance and indifference. You think β -Somerfield is the wiser man, so does he; in the absence of evidence I cannot see any ground for judging between them. But you are not content with condemning α as the villain; it appears that I am responsible for his villainy, which I could not be even if I had nursed him through his teething.

Fisher repeated his charge that the review indicated he had made a false claim. He expressed the view that normality was the least important difficulty when choosing good examples, and that a change of variate would do all that was wanted.

I have fairly often applied a z-test to crude values, and to log values, even when the translation is a severe strain, as in an early paper on potatoes with Miss Mackenzie, but have never found it to make any important difference.

Gosset agreed on 28 June to write to *Nature*. As to Fisher's remarks on normality,

... I think you must for the moment consent to be analysed into α -Fisher the eminent mathematician and β -Fisher the humble applier of his formulae.

Now it's α -Fisher's business, or I think it is, to supply the world's needs in statistical formulae: true β -Fisher doesn't think the particular ones that I want are necessary but between ourselves that's just his cussedness. In any case I quite agree that what we are doing with non-normal distributions is no business of either of them; it is merely empirical whereas α -Fisher is interested in the theoretical side and β -Fisher in whatever seems good to him. But when β -Fisher says that the detailed examination

of the data is his business and proceeds to examine them by means of tables which are strictly true only for normally distributed variables I think I'm entitled to ask him what difference it makes if in fact the samples are not taken from this class of variables. And moreover he is so far tolerant of the question as to instance cases in which a violent transformation made but little difference. That's what I believe but what I don't feel entitled to assume.

There was something for everyone in the letter which Gosset drafted (28 June) from Student to *Nature*. Tolley's statement was described as the impression that a careless reader may get; Pearson's review was moderated in tone by changing 'admit' to 'stress' in the passage quoted; Fisher's belief that normality made little difference received qualified support; and Gosset hoped that Fisher would 'indicate to us' what changes in the tables were necessary when non-normality obtained. At the same time, Pearson wrote a friendly letter to Fisher, blaming himself for failing to appreciate how his review would be seen from the other side.

The Editor of *Nature* was somewhat taken aback by the intervention of a third party, but he decided (4 July) to publish Student's letter in the correspondence columns, because both Fisher and Pearson agreed to publication and the letter 'does not appear to be likely to open a discussion on the review itself'. He made the last point clear by changing 'indicate to us' to 'show us elsewhere' in the version published on 20 July, but this exercise of his editorial powers only succeeded in irritating Fisher, who told Gosset that 'I ignore your "elsewhere", if it is yours' when enclosing his draft reply on 26 July. Another source of discord was that Student's address was given as the Galton Laboratory, University College, London, and, although Pearson 'wondered a little at the time', as he told Gosset on 24 July, he took no action. This was a mistake, because on 3 August *Nature* printed a letter from K.P., who had written to object while away on holiday.

I feel sure that he [Student] will recognize, on fuller consideration, that the task of a director of a laboratory would become impossible if anyone could use its address without first obtaining the permission of the director.

Egon Pearson went off to Poland at the end of July, and when he returned after an absence of two months he found that *Nature* had published Fisher's reply on 17 August. Fisher disagreed that modification of his tables for the analysis of variance, so as to adapt for non-normality, would be a legitimate extension of his methods. Even if 'an army of computers had extended the existing tables some two hundred fold', so as to provide tests for the Pearsonian system, there would still remain the distributions outside that system, problems would arise from the sampling errors when estimating parameters, and other statistics would become more appropriate. He assured the readers of *Nature* that biologists need not worry.

I have never known difficulty to arise in biological work from imperfect normality of the variation, often though I have examined data for this particular cause of difficulty; nor is there, I believe, any case to the contrary in the literature.

Pearson drafted a rejoinder, and on 25 September Gosset encouraged him to send it to *Nature*.

Fisher is only talking through his hat when he talks of his experience; it isn't so very extensive and I bet he hasn't often put the matter to the test; how could he?

Pearson's response (26 September) acknowledged the influence of their earlier discussions.

... after all it is a statement of the case largely built up on comments you have made to me from time to time.

He also gave some results on testing $\sigma_1^2 = \sigma_2^2$ for independent samples from leptokurtic distributions, where he found that agreement with the procedure based on normal theory was 'absolutely rotten'. His letter to *Nature*, written on 30 September, was published on 19 October and was the last link in this long chain. He asserted that the problem of non-normality was real, and that, although the population would seldom be known, nevertheless the study of practical examples could provide useful information about the adequacy of the standard tests.

6.5.2 Comments by E. S. Pearson

The light which the letters throw on the outlook of the three writers brings out what in their letters of 27 and 28 June Fisher and Gosset called the ' α ' and ' β ' characteristics in all mathematical statisticians. The first, α , is concerned with the theoretical approach which in many situations had, and still has only been developed by assuming either that random variables are normally distributed or by using asymptotic results. The second, β , concerns the behaviour of the statistician when he comes to the analysis and interpretation of real observational data.

The α -Fisher had derived brilliantly conceived and mathematically based techniques. He was aware of the difference between his α and β selves, but by various investigations had satisfied himself that departure from normality did not invalidate the conclusions drawn from the interpretation of data with which he, and most other likely users of his *Statistical methods for research workers*, were concerned. The reference in his letter to Gosset of 27 June is to his 1923 paper with Miss W. A. Mackenzie (*Journal of Agricultural Science*, 13, 311–20) on the manurial response of different potato varieties. Here, the authors gave analysis of variance tables assuming that the yield from a given variety grown on a different manurial treatment is (a) the sum and (b) the product of two factors. The results were almost the same. But, as far as the

outside reader was concerned, such an investigation by itself hardly provided an adequate study of the general problem and, as far as I know, no other full details of such investigations were discussed elsewhere.

Gosset, on the other hand, may have come into contact with a wider variety of univariate data distributions and was anxious for a more thorough exploration to be carried out on what we should now term the 'robustness' of Fisher's 'normal theory' techniques. My own training in K.P.'s tradition had recognized the existence of real non-normal frequency distributions, often quite well represented by one of his system of frequency curves. As a result, when Gosset, in his letter to me of 11 May 1926, raised the question of 'robustness', I had very readily seized on the idea of developing a systematic attack on the problem, using what could be termed experimental sampling. Even with the help of Tippett's *Tables of random sampling numbers* (published in 1927) this exploration was inevitably slow and patchy. When I wrote my review in *Nature*, the sampling programme was still in progress. A first report on it had appeared in *Biometrika*, 20, 356-60 (1928), and subsequently details were reported up till my paper in *Biometrika*, 23, 114-33 (1931), where I established in numerical terms that certain tests involving the ratio of two estimates of variance (using what Fisher called the test statistic z) were much more sensitive to non-normality than others.

By June 1929, I had already accumulated enough evidence to convince me that my warning about the normality assumption given in the review was justified. I was undoubtedly riled by claims that Fisher's tests were 'exact', with the implication that long-accepted procedures were incorrect and should be discarded. However, it was unwise of me to have inserted in the review the rather vague reference to 'the analysis of variance and the z - (and t -) distributions'. The basis of my statement was derived from experimental sampling, but not published until 1931.

If $z = \frac{1}{2} \log_e (s_1^2/s_2^2)$, then, when the variation is not normal, two situations must be taken into account.

- (i) The two estimators may be essentially independent, e.g. derived from separate samples; then the variance ratio test is not robust.
- (ii) In problems of the analysis of variance, e.g. when s_1^2 is a 'between group' and s_2^2 a 'within group' estimator of an error variance σ^2 ; in this case for non-normality, there is a correlation between s_1^2 and s_2^2 , tending to keep the variance ratio test robust.

Writing in 1929, I had possibly only analysed my sampling results under the former situation.

It was in consonance with Fisher's temperament that he should take offence at the implication that he had concealed the importance of 'normality', but the long drawn-out discussion shows that Gosset sympathized with my view that more attention should have been given to this point in Fisher's book,

and it also shows how anxious he was to prevent controversy between Fisher, E.S.P., and K.P., all of whom were in different ways his good friends.

Of course, in most of my more theoretical work with Jerzy Neyman, I was an ' α -person' ready to admire much of the ' α -Fisher'. However, when I came into contact with Walter Shewhart and with the British Standards Institution, what I wrote and published shows the ' β -side' of my approach. It will be found that, apart from a few references, I never discussed Neyman's and my α -approach with Gosset, because he was not a sufficient mathematician to appreciate it.

A fuller and more systematic exploration of this problem of robustness of 'normal theory' tests only became possible many years later with the introduction of the electronic computer (see e.g. the paper by Pearson & Please, *Biometrika*, 62, 223–41 (1975)).

6.6 EFFECTS OF NON-NORMALITY

This section concerns the letters in Group V, written between 19 July 1929 and 14 April 1931.

6.6.1 Editorial introduction

Pearson wrote on 19 July 1929 (letter No. V.1) to put forward, for comment by Gosset and Somerfield, a scheme for deriving from his univariate samples a sample containing k arrays drawn from leptokurtic type VII curves. Exactly one year later, his conclusions were beginning to emerge, and he summarized them on 6 September 1930 (letter No. V.6).

Some rather interesting things do come out though, and I believe that for the simplest case of analysis of variance—testing for the presence of a single factor as well as the random variation—(using η^2 in other words) for this, the 'normal theory' test will hold remarkably well for very wide variations in populations. It comes down to the fact that the criteria used in the test is a ratio of two estimates, and when the variation ceases to be normal there is +ve correlation between numerator and denominator, hence distribution of ratio doesn't change.

The same point is presented on pp. 129–31 of his paper published in the November 1931 issue of *Biometrika*, 23. This was the point which he had in mind in his 1929 *Nature* review of Fisher's book, but failed to make evident. Gosset had himself made the point about correlation between numerator and denominator in 1908. He may have forgotten, but it would be characteristic of him not to remind Pearson.

Studies on correlation were also envisaged in the letter of 19 July 1929, and Pearson referred on 1 November 1930 (letter No. V.7) to his results for the distribution of the sample correlation coefficient when the population

value is zero and the variables follow the same leptokurtic curve. The details were published in his paper 'The test of significance for the correlation coefficient' in *Journal of the American Statistical Association* for June 1931. A further paper written jointly with Leone Chesire and Elena Oldis was printed a year later in the same journal.

At the end of this period, Pearson turned to the robustness of the analysis of variance technique when applied to the Latin square. He asked for advice on 27 March 1931 (letter No. V.11) in regard to a suggested sampling experiment, and Gosset gave a concise reply on 1 April (letter No. V.12).

The Latin Square problem in practice departs, I think, from the theoretical conditions in two ways: (i) The populations tend to be leptokurtic and (ii) The variances are not really identical. Generally speaking the samples are too small to tell you so but I'm pretty sure of all the facts.

You propose to deal with (i) but I think that (ii) is equally important, in fact some of Fisher's disciples are, I believe, beginning to query the thing from that point of view.

Personally I expect that the approximations need not be very close for the method to work pretty well and that is Fisher's view but I want to see it tested.

After another letter from Pearson on 10 April (No. V.14), describing his proposed sampling scheme in more detail, Gosset wrote again on 14 April (letter No. V.15) to explain that his problem was not quite the same.

But my problem is this: (a) Allowing that, with normally distributed variates, Fisher's z test will allow us to estimate the significance of any difference in variance there may be between the 'random error' and the varieties, does this hold with non-normal variates? or (b) allowing that, with normal variates, 'Student's' tables can be used to judge the significance of any difference between variates, does this hold with non-normal variates?

To test this, I should take *one* quite abnormal population with β_1 and β_2 high and proceed to take a largish number of samples of 25, arrange them in Latin Squares and draw for places, i.e. where the 'letters' are to go, in the approved fashion, and then see what is the distribution of the four variances (total, row, column, variety) and also the distribution of 'Student's z ', dividing means of rows, columns and varieties* by their appropriate 'random S.D.'. Presumably, if the means work well the difference between means will work better.

* As there will be no real differences between row, column and variety, one may as well use all three.

Another topic under discussion was the distribution of range. Pearson described on 26 November 1930 (letter No. V.8) the difficulties which had arisen in fitting curves with the correct theoretical moments to form a table of percentage points, and Gosset sent him on 28 November (letter No. V.9) a book of calculations which Somerfield had carried out in preparing the crude table in Gosset's 1927 paper. The result of Pearson's computations appeared in his paper 'The percentage limits for the distribution of range in samples

from a normal population ($n \leq 100$), not published until the November 1932 issue of *Biometrika*, 24.

With his letter of 19 July 1929, Pearson enclosed an offprint of a long Bell Telephone Report by Shewhart, and pointed out that, in putting forward his four criteria by which to judge lack of statistical control in industrial production, Shewhart was clearly not aware of, or had not understood, Fisher's methods of analysis. The criteria were afterwards reproduced in Shewhart's book *Economic control of quality in manufactured product*. Shewhart had invited Pearson to comment and these matters were later discussed between them on Pearson's visit to New York in 1931.

6.6.2 Comments by E. S. Pearson

By the summer of 1929, I had decided that it was time to extend my robustness research beyond the study of univariate tests. At that point of time, few readers of *Biometrika* would have shifted their approach to that of Fisher's Mark II. In so far as they had a mathematical background and wanted to understand how the Mark II statistical procedure had been developed, they would have needed to devote much time to deriving for themselves the theory underlying *Statistical methods for research workers*. As I have said, the mathematical paper which Fisher had read in 1924 at the Toronto Congress of Mathematicians was not printed until 1926; in so far as Fisher would not know to whom to send offprints, a wide understanding would have been difficult to come by.

No doubt when I wrote my papers published in *Biometrika* in 1929 and 1931, it seemed necessary for me to start with the derivation of the distribution of what was in fact Fisher's $z = \frac{1}{2} \log_e (s_1^2/s_2^2)$ distribution, for the ratio of two independent estimates of a common variance. Also, as the users of Mark I were still taking the squared correlation ratio η^2 to test for linearity of regression, I expressed Fisher's test in terms of this statistic, familiar to those whose theoretical basis was that of K.P.'s Mark I.

In reading the letters of 1, 9, 10, and 14 April 1931 (Nos. V.12-15) to and from Gosset after a long interval of time, it is clear that I had not yet grasped Fisher's practice of randomization, the basis for which has been very fully described in Joan Fisher Box's *R. A. Fisher: The life of a scientist* (Wiley, 1978). In 1931, I was wanting to investigate the robustness of a model, in which a random normal 'error term' was added to 'row' and 'column' means in a Latin square. At Joan Box points out (p. 147), by allocating varieties (say) to plots at random, Fisher realized that, under the null hypothesis of 'no treatment differences', the distributions of the test criteria $F = S_B^2/S_W^2$ or of $z = \frac{1}{2} \log F$ would not exactly follow the normal theory distributions. As she remarks, 'this conclusion was difficult to justify theoretically and was for

years to cause trouble among statisticians'. It was not till 1933 that Eden and Yates (*Journal of Agricultural Science*, 23, 6–17) demonstrated on a large-scale uniformity trial that this assumption was approximately justified in practice. B.L. Welch, working from my Department of Statistics in 1937 published a paper (*Biometrika*, 29, 21–52) entitled 'On the z-test in randomized blocks and Latin squares', again studying the randomization distributions, derived from uniformity trials. His summary and conclusions are given on his pp. 47–8. It may appropriately be noted that Fisher and myself were Welch's Ph.D. examiners; I remember that in the discussion at the oral examination Welch was able to keep his end up. I think that Fisher regarded the investigation of value, and incidentally lending support, for practical purposes, to the randomization assumption; but he never, that I know of, referred to this in print. My own views on randomization were given in the paper which followed that of Welch (pp. 53–64).

It was only later in the 1930s that I began to realize the meaning and purpose of randomization. That Student was really warning me off my projected experimental sampling project seems to follow from his remark in the third paragraph of his letter of 14 April (No. V.15), where he indicates that my experiment should involve 'drawing for places' to show to which of the Latin square plots my five letters A, B, C, D, and E should be allotted. As I left for America a few days later, I did not take the investigation further, nor did I come back to it on my return to England in September.

My long four-year period of testing for robustness was practically completed, when I sailed to New York. On my return, five months later my energies were largely concerned with theoretical work with Neyman, and also with the introduction of statistical techniques into industrial problems.

6.7 DIFFERENCES ON z-TEST AND DESIGN OF EXPERIMENTS

This section concerns the letters in Group VI, written between 29 January 1932 and 15 October 1937.

6.7.1 Editorial introduction

Gosset opened his letter of 29 January 1932 (No. VI.1) by suggesting the calculation of tables of the 'studentized' range, and the relevant paragraphs were reproduced as Letter III in Pearson (1939). Such tables were eventually published in 1943 with the assistance of H. O. Hartley.

Many of the other letters were written because Gosset came under attack, first from his old teacher, and then from his former pupil. The issue of *Biometrika* for May 1931 carries a paper by K.P. 'On the nature of the relationship between two of 'Student's' variates (z_1 and z_2) when samples are

taken from a normal bivariate population'. His problem was to test the hypothesis that the population means are equal. But, since matters are complicated by the existence of correlation, he warned against the use of Student's z .

For example, if we test for the relative effectiveness of two drugs or two methods of factory production on the *same* groups of individuals and find a significant difference, we have not obtained evidence that there would be a significant difference had the drugs or same methods of production been tested on different groups of individuals.

Gosset wrote on 14 July 1931 enclosing a note on the use of z in testing the significance of the average difference between correlated variables.

I hope you will see your way to put it in as I have always attached considerable importance to arranging matters so that the correlation should be as high as possible. In the case of agricultural experiments, it has been my chief criticism of Fisher that he does not take all possible steps in this direction.

Further letters from Gosset on 23 July and 18 August show that K.P. had difficulty in grasping Gosset's comments, but these were published in the issue of *Biometrika* for December 1931. Here, Gosset's two-page response to K.P.'s criticisms is immediately followed by a seven-page rejoinder from K.P., who remarked that "'Student" seems to me to misinterpret the outcome of his own test'. The situation was becoming impossible, and Gosset wrote at length to Egon Pearson in the continuation of his letter of 29 January 1932.

As to K.P.'s attack on z I feel that his style is somewhat cramped by his wish not to make me appear too ridiculous, and indeed my own is to some extent.

It's rather a pity for naturally his opinion carries the greatest weight and yet, in this case, he is definitely wrong.

Is it any use making a further reply? It wouldn't convince him and would perhaps only convince others if one let oneself go which I certainly won't do. May I write an answer to you which I could send him if you thought it advisable—even in an altered form?

The points I would like to make are ...

He received a very sympathetic reply, which advised him to make a case to K.P. in a private communication. Gosset's letter, sent on 29 March 1932, begins with his reaction to the personal remarks published in *Biometrika*.

While it would not be reasonable for me to expect you to see eye to eye with me in the matter of Student's z , it does seem to me that you do not quite realise just how we do use it in practice.

He enclosed a five-page analysis of questions which arise when working with correlated material 'for your own perusal & not for publication', and there the episode ended.

Fisher had advocated randomization in field experiments since 1926, but Gosset was always in favour of systematic designs. The last of Gosset's letters

to survive from his correspondence with K.P. is dated 19 February 1935, and shows that he was working on a paper for *Biometrika* about experimental design while recovering from his car accident in the summer of 1934.

I am sorry to say that the work on the Half Drill Strip paper is only about half done owing to the fact that my leg got sufficiently well to allow me to get back to work—in irons—by the beginning of October.

A meeting to discuss 'Co-operation in large-scale experiments' was arranged by the Industrial and Agricultural Research Station of the Royal Statistical Society, and held on 26 March 1936. Gosset opened the discussion, and urged the merits of the half-drill strip, whereas Fisher, who spoke next, voiced strongly opposing views. Beaven was there to support his old friend, and Fisher's criticism of the half-drill strip system left him unmoved.

He was not sure whether Professor Fisher had damned it with faint praise, or unqualified censure, but in either case his withers would be unwrung.

The meeting was soon followed by Fisher's paper with Barbacki on 'A test of the supposed precision of systematic arrangements' [CP 139], and by a further exchange with Gosset in the correspondence columns of *Nature*. Meanwhile, Gosset continued to prepare his last paper, which was almost complete at his death and required only minimal editing by Pearson and Neyman before being published. He sent an appendix on 30 March 1937 (letter No. VI.8) and explained that he was finding it rather hard to describe Fisher's paper adequately and simply, more especially the latter.

He has taken advantage of the fact that Beaven proposed to determine the error of plots by finding the error of sections of them as if they were independent whereas they are of course correlated to discredit Balanced arrangements although (i) I at once pointed out the fallacy and (ii) it could equally fallaciously be used to determine the error of the most randomised arrangements. And three out of the four conclusions depend on this: anyone could have foreseen the sort of results that would follow though not, I admit, their extent which depends on faulty technique on the part of the cultivator of the uniformity experiment which results in certain of the drills having periodically systematically higher yields than the others: every eighth* giving the highest yield of its eight in twelve out of the 15 eights and second in the other three. As he took six together and then missed one his 'half drill strips' were high or low throughout their length according as they contained more or less of the high yielding drills and the correlation between sections of them was much higher than would be usual. But enough of this.

*Doubtless his seed drill had eight tines and they were badly spaced so that one drill had more room.

Another letter on the same topic, dated 19 April 1937 (No. VI.10), was reproduced as letter VI in Pearson (1939).

Karl Pearson died on 29 April 1936, and in preparation for his biographical

essay, Egon Pearson asked Gosset whether any lecture notes from 1906–7 had survived. None was enclosed with Gosset's letter of 11 May 1936 (No. VI.5), but he supplied a variety of recollections which continue those in § 3.2.

I also learned from him how to eat seed cake for at about 5 o.c. he would always come round with a cup of tea and either a slice of seed cake or a petit beurre biscuit and expect us to carry on till about half past six. . . . During the war I ran in one day and was put on to bomb dropping tables which were unfortunately inaccurate owing to faulty interpolation and had to be done again. I remember going out to 7 Well Road . . .

Gosset's letter of 13 October 1937 (No. VI.13) mentions thrombosis and he died three days later.

6.8 FINAL COMMENTS BY E. S. PEARSON

In the five years from April 1926 to April 1931, Gosset and I had exchanged 83 letters; in contrast, during the six years from January 1932 to October 1937, only fifteen letters, all from Gosset to E.S.P., have survived. Eight of these were written in the last year of his life when he was putting together the draft of his paper 'A comparison between balanced and random arrangements in field plots', published posthumously in 1938 (*Biometrika*, 29, 363–79). The explanation of this change in the tempo of our correspondence is not hard to find. I shall take this opportunity to summarize what I owed to Student and how our relationship changed over the five years preceding my visit to the USA in 1931.

In the two months covered by the six letters of Group I, he had given me criticism and enlightenment which enabled me to embark on a statistical philosophy which bridged the gap between modern mathematical statistics Marks I and II. I came upon ideas and problems justifying my seeking cooperation with Jerzy Neyman. In the nine letters of Group II, just overlapping those of Group I, Gosset and I were jointly discussing the use of sample range, to which both made some contributions. This discussion led to his proposal to use range as a practical tool, illustrated in his 1927 paper 'Errors in routine analysis'. The 25 letters of Group III covered a period of two years, May 1927 to May 1929, when I discussed with him how to carry out his suggestion of exploring the robustness of univariate, normal theory tests. In this period, our discussion of G.F.E. Story's research left me with many valuable ideas, greatly broadening my knowledge of how small-sample statistical analysis was carried out in the Dublin brewery and how it might be applied elsewhere in industry. This was a great help to me in the years which followed, when I was in contact with W.A. Shewhart of the Bell Laboratories, with industry in England, with the British Standards Institution,

and with the formation of the Industrial and Agricultural Research Section of the Royal Statistical Society. Whereas Neyman had a good deal of practical experience in handling of data in Warsaw, in an advisory capacity, my claim to be more than an academic statistician was largely due to Student's guidance.

The 28 letters of Group IV, all written between June and September of 1929, arose directly from my *Nature* review of the 1928 second edition of *Statistical methods for research workers*. Again, by his remarks, Gosset gave me confidence to stand by my views in so far as they differed from Fisher's. Also, I obtained a warning as to the way in which Fisher could not compromise.

There were 15 letters in Group V written between July 1929 and April 1931 when I was preparing to go by sea to the USA. While rounding off for publication my experimental study of robustness of univariate normal theory tests, on which Student made a few encouraging comments, I had begun to consider tests involving two or more variables. I was also trying to master the analysis of variance and the concept of randomization; he sent me to read his article on 'Field trials' in Hunter's *Encyclopedia of agriculture*. The majority of these letters are from E.S.P. to Gosset. I was reporting the progress of my work; at this time, perhaps the most useful thing which he did for me was to warn me of the difficulties that would be involved if I tried to devise a sampling experiment based on a model representing the different factors influencing the result of a Latin square arrangement. As I have said, in this he was successful.

It seems that neither Student nor Mathetes thought that I should gain much from my American visit, but in this they were wrong. The visit widened my outlook and gave me confidence and experience in teaching. Through W. A. Shewhart, it put me into touch with quality control problems in industry and because Sam Wilks attended my seminars at the University of Iowa it introduced into America an outlook on modern mathematical statistics Mark II which was not solely subservient to the diktats of the Fisher School.

Gosset's letters of 1937 (Nos. VI.8-15) had as their main subject his advocacy of the use of balanced rather than randomized designs in plot experimentation. The conclusions were based on his own practical experience. His views were printed in the posthumous paper published in the February 1938 issue of *Biometrika*. As Jerzy and I had been discussing his ideas with him shortly before his death—indeed visited him in the nursing home where he was for some weeks—we added a note after the paper (*ibid.*, pp. 381-8). Of course, these views were in direct contradiction to the teaching and practice of the 'Fisher School'; a correction from that quarter was inevitable. After this, Gosset's case seems to have gone by default, as far as the printed word is concerned, but it is impossible to say how far the experimenters among his many correspondents continued to believe and practise balance in their

designs. Just as Gosset had five years before, in 1932, been most unwilling to enter into controversy with K.P. in regard to the *t*-test, so now when preparing his last paper he had written that the paper was some months in preparation 'not so much owing to lack of time as to lack of inclination to controversy'.

Yates's paper in reply, 'The comparative advantages of systematic and random arrangements in the design of agricultural and biological experiments', was published in January 1939 (*Biometrika*, 30, 440-66). It is a carefully written paper with much informative tabled evidence. If Student had not died in 1937, there is little doubt, knowing his character, that the views of these doughty opponents would have gradually got closer; but, as it was, there was no reply given to Yates and war came in September 1939. The whole controversy was dropped; I never discovered what were the views of field experimenters like C. E. Lane Poole, the Inspector General of Forests in Australia. Having myself no practical experience in design of experiments, I could not attempt an answer, even if in 1939-45 I had had time to examine the matter.

Speaking from outside I should judge that neither Fisher nor Yates would have attempted to find fault with the advice which Student gave to his brewery or to the Irish Department of Agriculture. In his capacity as an adviser, he had to take into account many other considerations besides the probability measures resulting from the analysis of a single experiment by so-called 'valid' techniques. As he wrote to me in one of his illuminating letters (No. VI.10 of 19 April 1937),

But, in fact experiments at a single station *are* almost always valueless; you can say 'In heavy soils like Rabbitsbury potatoes cannot utilise potash manures'. But when you are asked 'What are heavy soils like Rabbitsbury?' you have to admit until you have tried elsewhere - that what you mean is 'At Rabbitsbury etc.'. And that, according to X may mean 'In the old cowfield at Rabbitsbury'. What you really want to find out is 'In what soils and under what conditions of weather do potatoes utilise the addition of potash manures?'

To do that, you have to try it out at a representative sample of farms of the country and correlate with the character of the soil and weather. . . .

Few practising statisticians have the ability and experience of a Gosset, Fisher, or Yates, and therefore to safeguard against blunders it is no doubt important to teach would-be applied statisticians the value of randomization where it is possible in design of experiments. But this does not prove that Fisher was 'right' and Gosset 'wrong'. It is perhaps too often forgotten that mathematical models using probability theory are there to provide an aid to human judgement. Their derivation gives intriguing exercises to mathematically trained minds; when used, they help to clarify the ideas of the applied statistician in reaching his conclusions leading to action.

Long ago, I was much struck by a remark made by Tippett, another

distinguished applied statistician; he pointed out that the summarizing arrangement of data in an analysis of variance table provides illumination on the sources of component variation, even without the introduction of 'valid' probability measures. In the recent history of mathematical statistics, there are many instances of the development of these mathematical models, which put probability theory into gear with our process of thought. They are helpful, even when only approximately representing that illusive thing—reality. For example:

- (a) There are asymptotic expressions for the variance of maximum-likelihood estimators, based on sufficient statistics; these are suggestive but in some cases manifestly inadequate even in moderate-sized samples.
- (b) There is the theory of fiducial inference, which seems now to be universally suffering replacement by Neyman's simpler confidence interval theory.
- (c) What practical statistician nowadays often makes use of the mathematical concept of 'amount of information'?
- (d) The whole of Neyman's and my approach to the testing of statistical hypotheses was, as Tippett recently remarked, illuminating in the way in which it sorted out the shaky intellectual basis for many sampling tests current in the late 1920s. But it was just one approach—neither wrong nor right.
- (e) Today in many circles the current vogue is a neo-Bayesian one, which is of value because it calls attention to the fact that, in decision making, prior information must not be neglected.

And so things go on. Suddenly there came into my head a canto from Dante's *Purgatorio*, which, on reflection, strikes me as having a certain relevance:

In the circle of the Proud: the fickleness of fame

Oh vana gloria dell'umane posse!
 com' poco verde in su la cima dura,
 se non è giunta dall'etati grosse!
 Credette Cimabue nella pittura
 tener lo campo, e ora ha Giotto il grido,
 sì che la fama di colui è scura.

(Canto XI, lines 91–6)

7. General commentary

Gosset went to Dublin in the autumn of 1899 to take up the post of a Brewer with Arthur Guinness, Son & Co. Ltd. He was twenty-three years old, and his ability had been shown by the scholarships which took him to Winchester College and Oxford University, where he had been trained in mathematics and science. Guinness was a long-established and important business, the management of which had decided on the greater use of scientific methods in the brewing of stout. This policy was implemented by a series of appointments of men from Oxford or Cambridge with science degrees, so that Gosset joined a group of colleagues with the same social background, among whom he found those of a similar age who shared his interest in the outdoor activities offered by the Irish coast and countryside. Within a few years, he was married, and his home near Dublin soon became the centre of his family life and a widening circle of friends.

During his early years with Guinness, Gosset was presented with problems in the analysis of experimental data from brewery work. He had no knowledge of statistical methods, and turned for assistance to standard textbooks on the combination of observations. There he discovered the law of error, later known as the normal distribution, the concept of probable error, the distribution of the mean of a sample from a normal population, the difficulties arising from 'entangled measures', Airy's treatment of the question of whether or not means are discordant when judged by probable errors, and Merriman's account—following Gauss—of the precision of estimates of error in large samples. However, Airy's and Merriman's treatments were of limited use because they were concerned with observations made under stable conditions in astronomy and geodesy, respectively, whereas brewery data arose from short runs and were greatly affected by variability of materials and changes in the laboratory environment. Gosset's attention was thus directed towards the need for methods which could be applied with small samples to assess the discordance between means and the relationship between variables. Statistical inference as presented by Merriman was firmly based on the principles of inverse probability, so that, when Gosset reported to the Board of Guinness in 1904 on the desirability of professional advice, his recommendation was couched in that form.

We have met with the difficulty that none of our books mentions the odds, which are conveniently accepted as being sufficient to establish any conclusion, and it might be of assistance to us to consult some mathematical physicist on the matter.

Contact with Karl Pearson was effected through intermediaries in Oxford, with the result that Gosset and Pearson met for the first time about eight months after the appearance of the report. Pearson was forty-eight years old, a man with tremendous energy, great determination, and clear objectives. He was in charge of the only university department in Great Britain and Ireland where statistics was taught to any depth and where graduate students worked on an integrated programme of research. The publications of the Biometric School, which he and Weldon created, appeared in the journal *Biometrika*, which they founded and he edited, and in memoirs collected in special series, for all of which he obtained the necessary funds. He had the overpowering enthusiasm required to drive forward the study of statistics in relation to evolution and heredity, and the strength of character to accomplish what the solitary and reserved Edgeworth, despite his outstanding ability, could never have achieved. But Pearson's dominating personality meant that mistakes could be acknowledged only with reluctance, and differences of opinion could lead to open hostility, so there was a cost in respect of damaged feelings and broken friendships.

Through his association with Pearson, Gosset encountered topics which had not then reached textbook level: the idea of correlation and large-sample properties of the correlation coefficient, the Pearson system of frequency curves, and the chi-squared test for goodness of fit; and, in the Biometric Laboratory, he found a practical outlook with which he deeply sympathized. Thus Gosset's early work marks the point at which the long-established methods used to combine observations in astronomy and geodesy, concerned with the normal distribution and the estimation of parameters, joined the new stream of techniques devised in the Biometric School, concerned with association, non-normal distributions, and testing for agreement. The union is best illustrated by his investigation of the probable error of a mean. This called for the distribution of the variance in normal samples, which he found by calculating moments and fitting a Pearson curve, unaware that the geodesist Helmert had in 1876 used mathematical analysis to obtain the distribution, as K.P. made known in 1931. Gosset's paper presented the first table of Student's z -distribution and is a good example of an important discovery made on the boundary between two fields of enquiry with quite different objectives. When account is also taken of the background to his work, namely the pioneering use of statistical methods in an industrial research laboratory, the combination of circumstances is so exceptional as to go some way to answer the question which Fisher posed in his obituary notice.

How did it come about that a man of 'Student's' interests and training should have made an advance of fundamental mathematical importance, the possibility of which

had been overlooked by the very brilliant mathematicians who have studied the Theory of Errors?

Gosset's stay in London was the beginning of a lifelong friendship with Karl and Egon Pearson, and he always retained a strong interest in their statistical activities, whether in the Biometric Laboratory or in the successor Department of Applied Statistics, usually calling at University College when passing through London. His employment meant that he was located in Dublin for almost the whole of his career, during which time the journey to London by boat and train typically began late in the evening and ended early in the following afternoon. He was therefore somewhat isolated within a small English community which endured a period of violent civil disturbances between the Easter Rebellion of 1916 and the Civil War of 1922. Gosset was clearly keen to see more of his friends on the mainland, as he explained to Fisher on 27 June 1923.

The worst of living in this beastly country is that one hasn't been able to ask Christian people to stay with one for years and years and that consequently one must spend one's Easter holidays at home.

However, the postal service then was much superior to what is available now, and like all his contemporaries he was an enthusiastic correspondent—although not in the same class as K.P., where the archives contain over sixteen thousand letters.

Gosset's statistical correspondence was not only a means of keeping in touch with developments in methodology at a time when there were few professional meetings to attend but also a way in which he could participate by adding comments, suggesting topics for research, and making offers of help. He was always courteous and considerate, never solemn or offended. His dislike of controversy did not imply a willingness to compromise. He tended to describe his own achievements with extreme modesty, which sometimes took the form of self-disparagement, and he made light-hearted analyses of his supposed shortcomings. Nearly all the letters were written by hand at home, and nearly all those addressed from St James's Gate dealt with the business of the Guinness brewery. Enquiries were usually answered on the day of receipt, sometimes late in the evening, and often with an appendix of further thoughts on the following day. The correspondence reflects Gosset's integrity, and amply confirms the personal characteristics recalled by those who knew him well.

Most of the statistical interest in the letters to K.P. arises from the period between 1907 and 1919. Gosset's work on the probable error of a mean was followed up with a larger table of Student's z -distribution. His discoveries on correlation were extended from the first test procedure for small samples to the estimation of the coefficient from simultaneous time series and the treatment of ties in ranking, while his rediscovery of the exponential limit of a

binomial led him to examine the assumptions for a Poisson distribution, study the concept of a mixture, and almost complete the circuit back to the negative binomial distribution. Pearson recognized Gosset's originality, accepted his advice on estimation, and sought his opinion on experimental design, but they soon came to differ in respect of the application of Bayes' theorem. Pearson was willing to apply a non-uniform prior to the likelihood for a parameter such as a correlation coefficient, on the basis of past experience in cognate fields. Gosset perceived that such a practice could easily result in failure to notice specific features of the data set giving rise to the estimate, and urged that uniform priors alone should be used in the first place. The First World War imposed a huge burden of extra work on K.P., which Gosset was very willing to lighten. This shift of emphasis coincided with a gradual change in the focus of Gosset's attention away from laboratory and pilot-plant experiments towards agricultural field trials, and most of the early themes were left for others to explore, so that his interests drifted away from those of K.P. The correspondence surviving from the 1920s is largely personal, with family photographs and school results but little further comment on the work of the Department of Applied Statistics. By this time, K.P. was 'perhaps a little intolerant of criticism', and Gosset had more than enough to occupy his evenings at home. Notwithstanding his eventual realization that K.P. had become crusty and remote, the devotion of Student to his old professor was lifelong, and the tone of the letters is respectful and affectionate even when Gosset must have been sorely tried.

When Fisher attended a course on the combination of observations given by his tutor Stratton, he met a tradition which had also contributed to Gosset's paper on the probable error of a mean. Thus, the background which Gosset and Fisher shared helps to explain how they first became acquainted. Gosset's paper profoundly influenced Fisher, and led to his remarkable development of normal sampling theory, culminating in 1928 with the general sampling distribution of the multiple correlation coefficients in the multivariate case. When mentioning Gosset's work at intervals between 1915 and 1938, Fisher nearly always referred to the 1908 paper on Student's z -distribution, which also forms a major theme of his obituary notice. These comments are uniformly favourable, and expressed using phrases such as 'brilliant researches' and 'revolutionary step'. While the main thrust of the paper is towards direct probabilities, Gosset accepted contemporary ideas of statistical inference in using inverse probability to express his conclusions. By 1916, K.P. had decided that, when using Bayes' theorem, the choice of a uniform prior distribution is arbitrary. Fisher's statement that his absolute criterion, known later as the method of maximum likelihood, is derived from the principle of inverse probability, was thus the source of much confusion. K.P. criticized the supposed use by Fisher of inverse probability in estimating a correlation coefficient, while Gosset welcomed his supposed adherence to classical tenets.

However, Gosset's work actually strengthened the reaction against inverse probability and Fisher explicitly rejected prior distributions after 1916.

From the point of view of the foundations of statistical inference, it is a great pity that Gosset's modesty led him to avoid leaving behind more than a series of highly insightful remarks bearing on the subject. In discussing examples in his 1908 paper, Gosset translated his P -value into a multiple of the probable error of a normal distribution, thus relating his inference to a form which had become standard in astronomy and surveying in the preceding century. In those sciences, the 'probable error' (p.e.) of a given measurement was defined as that deviation from the true value such that the error was as likely to exceed as to fall short of it. Deviations up to twice the p.e. were not unreasonable, therefore, and only when deviations exceeded three times the p.e. were there good grounds for suspicion, while the rapidity with which the normal distribution fell away in the tails meant that four times the p.e. was regarded as conclusive evidence that something was wrong. This usage of the p.e. could be given an interpretation related to a Bayesian posterior relative to a uniform prior distribution, but could equally well sustain a non-Bayesian interpretation, based on the transference of the improbability of large deviations from a normal mean to the improbability of the assumptions implying such a deviation. K.P. was wont to interpret the P -values derived from his chi-squared test as odds for or against the adequacy of his fitted curves; but it would seem that neither he nor Gosset felt the need to attempt much more in the way of logical precision. Indeed, when Gosset learned from Fisher that he would have to enter the χ^2 tables with fewer degrees of freedom than Pearson's rules had suggested, his comment was that he would have to get used to higher P -values—as if a given P -value had no absolute meaning. In the light of more recent studies of the difficulties people have in assessing the precise import of probabilities, Gosset's view has more to be said for it than might appear. At the same time, it is amusing to reflect that the stated reason for Gosset's first contact with K.P. was to learn what odds 'are conveniently accepted as being sufficient to establish any conclusion'.

A concept of the 'rigorous specification of uncertainty' was introduced when Fisher tabulated the t -, z -, and chi-squared distributions in terms of their percentage points. His emphasis on the exactness of these distributions, and his disparagement of the probable error (whose 'common use ... is its only recommendation') all served to tighten up the logic of statistical procedures—a process which, in a sense, reached a peak with Neyman and Egon Pearson's classic paper of 1933. Combined with Fisher's insistence on the distinction between the two forms of measurable uncertainty which he called 'mathematical probability' and 'likelihood', reference to prior distribution of parameters almost disappeared from the literature of mathematical statistics for some twenty years after 1930. For his part, Gosset, as

late as 1922, appeared willing to regard the likelihood as an 'inverse probability' arising from a uniform prior for the parameter being estimated. As has been noted in Chapter 5, he perceived the important fact—not always attended to by latter day users of Bayesian methods—that inferences about a given parameter θ from different sources can be directly compared or combined only if each inference is based on a uniform prior for θ .

Gosset was evidently much less impressed than was Fisher by the argument that, if θ is unknown, then so also is θ^3 , and a uniform prior for θ is incompatible with a uniform prior for θ^3 . This may be related to the fact that the parameters with which Gosset was concerned were usually 'dimensional' quantities such as the yield of a crop, whereas in his genetical investigations Fisher had to deal with quantities such as recombination fractions, where dimensionality was much less well defined. If Gosset had been asked why he was prepared to assume that an unknown gain in crop yield, rather than the cube of such a gain, should be taken as uniformly distributed a priori, he would probably have enquired what meaning could possibly be given to the cube of a yield per acre.

After Fisher went to Rothamsted, his interests were inevitably enlarged to include agricultural field experiments, and, since Gosset by then had much experience to offer, his advice was naturally sought. But Gosset had acquired from Beaven a deep knowledge of the practical advantages of systematic designs and from Pearson the view that much could be explained in terms of correlation. These were the twin pillars on which his ideas about agricultural experiments were founded, whereas the ideas which Fisher presently introduced, notably the use of randomized designs, were bound to conflict in correspondence or in public sooner or later. Meanwhile, Gosset assisted the preparation of *Statistical methods for research workers* with much tedious calculation performed on a hand-operated machine, helpful suggestions which modified the text, and proof-reading which expedited publication. As one edition succeeded another, the most important of his early discoveries was accorded a recognition more general than the pages of *Biometrika* could possibly give. The presentation of the tables of Karl Pearson's χ^2 and Student's t was a consequence of copyright problems, but marked a change from the previous system of calculating P -values and interpreting them as inverse probabilities. Instead, fixed significance levels were recommended as 'more convenient', and they formed the basis of the Neyman–Pearson theory which soon followed. However, thirty years later, Fisher wrote in *Statistical methods and scientific inference* that 'no scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses'. At present, fixed levels are slowly being abandoned in favour of a system where P -values broadly assess the weight of evidence against the null hypothesis, indicating a return to the view that K.P. held in 1900.

The correspondence between Gosset and Fisher was a well-balanced and

generally good-tempered exchange, supported by mutual help and esteem, which began with Gosset encouraging the promising young man and ended with Fisher advising the busy senior manager. Common interests led to differences of opinion, but Gosset's dislike of controversy, gift for diplomacy, and good humour, were invaluable assets. The lines of communication between K.P. and Fisher eventually passed through Gosset, but he never compromised in the event of disagreement with either of these mighty opposites. Although Fisher's gift for words was used mainly in the service of science, he was a master of polemics in which half-truths were skilfully deployed. Gosset was bound to respond when at last he became a public target for Fisher's animadversion, and their clash on experimental design was more serious because the contestants were by this time set in their views. However, Fisher paid warm tribute to a loyal and generous friend in his obituary notice, although even here the account is peppered with polemical allusions to K.P., Egon Pearson, and Neyman.

Egon Pearson joined K.P.'s department in 1921 as an Assistant Lecturer, after contributing to the war effort and completing his mathematical studies. Fisher's early work on normal sampling theory and the analysis of variance, together with the concept of likelihood and the new theory of estimation, made a tremendous impact and led Egon to think about fundamental ideas in statistical inference. By 1926, he was convinced that there was a marked divergence between his statistical outlook and that of K.P., but this was not an easy matter to resolve when the son had been brought up to hold the father in great respect, while the father had a strong personality combined with a distaste for criticism. At this crisis in his professional career, Egon turned for guidance to Gosset and for collaboration to Neyman. Gosset was a familiar figure in K.P.'s department, twenty years Egon's senior, and could provide sound advice, practical experience, and warm encouragement, in effect filling the place which K.P. was unable to occupy. Neyman was much the same age as himself, a romantic figure whose fascinating life history captured Egon's imagination, and who had a mathematical expertise which was the ideal complement to Egon's good sense and visual appreciation. The associations with Gosset and Neyman were influential on Egon's statistical activities until the early 1930s, and thereafter his contributions assumed an individual shape, concerned notably with statistical methods in industry, statistical inference in practice, the history of statistics, and the editing of *Biometrika*.

During this latter period, Egon Pearson never forgot the debt which he owed to Gosset. His 1939 essay on Student as statistician is the major section of a long obituary notice, and represents a careful examination of Gosset's early activities in the Guinness brewery and subsequent papers in *Biometrika*. There followed, in the 1960s, a series of historical articles, some concerned to place Gosset's work in perspective as part of the British contribution to

modern statistical methods. After retirement, Egon Pearson's time was first occupied in producing an edited version of the huge amount of material accumulated by K.P. for a course of lectures on the history of statistics in the seventeenth and eighteenth centuries. Finally, he turned to reconsider with understanding and insight the part played by Student in his pre-war statistical career, and the earlier relationship between Gosset and K.P. This last and almost completed tribute, consisting of about two hundred pages of typescript, expresses his deep appreciation of the generosity of an older friend, and is where the present book began.

Aspects of the characters of K.P. and Gosset in maturity, and of Fisher and Egon Pearson at earlier stages in their careers, are well brought out by the episode arising from Egon's review of *Statistical methods for research workers*. Fair comment produced a Fisherian rage of a type which later came to mar Fisher's relations with many even of his relatives and friends. Egon's lack of self-confidence, combined with the awe which Fisher inspired among contemporaries, was a serious impediment to adequate response, and Gosset's recognition of these factors led him to intervene. As Fisher gradually became more reasonable, Gosset's search for a peaceful outcome drew within visible distance of success, but just at the crucial moment he addressed a letter from the Galton Laboratory. Egon was aware that the director might object, but shyness meant silence, and K.P.'s objection came swiftly, notwithstanding the fact that Gosset was a friend of twenty years standing. Matters were eventually resolved by Gosset's unrecorded peacemaking with K.P. and his robust estimate of Fisher's pretensions. Nobody could have done more than he with people who were so intransigent, and whose fires were merely banked before breaking out at other times in other places. However, even during these sharp exchanges, there were signs of mutual respect, if not of affection. In his letter, K.P. indicated some regret that Gosset had no right to the Galton Laboratory address, while Fisher suggested an approach to the problem which Student had proposed which, he said, a student of Student might naturally think of. Further aspects of relationships of that period were recorded by Egon Pearson in a postscript to the account of his correspondence with Gosset.

There occurred an incident some time in 1933 or 1934 which I have never before put on record. It is interesting because it illuminates the relationship then existing between K.P., R.A.F., E.S.P. and W.S.G. R.A.F. asked me whether I would approach K.P. suggesting that K.P. and he should combine in putting forward Gosset as a candidate for a Fellowship of the Royal Society. I fear that I took no action on this; I had no standing in the matter and was afraid of a rebuff which would make matters between us worse. I was aware how much K.P. had been hurt by R.A.F.'s scattering the contents of his treasured Museum; I remember that in 1931 K.P. had written three papers criticizing Student's *t*-test work; I was also at the moment in K.P.'s bad books for advocating many of R.A.F.'s methods. After all it was not my business to act in the matter; if he would not write himself to K.P. he could have got Yule or

Greenwood, long established Fellows, to join with him. However whether K.P. would have agreed or not, it is clear that if K.P. and R.A.F. had together sponsored Gosset he would have been elected. He might therefore be described as an F.R.S. *manqué!*

Gosset's research publications and his professional correspondence together show that his contribution to the progress of statistical methods in industry and agriculture far exceeds the *t*-test with which his pseudonym will always be associated. Although told in 1909 to consider himself 'a brewer first & only secondly a statistician', he continued to write statistical papers for the rest of his career, and those concerned with the relative advantages of balanced and random arrangements in experimental design still command attention. Gosset criticized the Lanarkshire milk experiment because a randomized allocation of children to treatments was modified, but in his last paper he gave the arguments in favour of systematic designs for field trials. Egon Pearson expected in 1939 that the debate on randomization would be resolved within ten or twenty years, but in 1979 he pointed out that the controversy was dropped when the Second World War came. Research on randomization has continued steadily since about 1950 and is now focused on two major aspects, namely the construction of optimal designs when the errors are assumed to be correlated, and restricted randomization, where the choice of design is made from a subset of all possible arrangements. A Fisher-Gosset compromise seems thus to have been achieved by a dispassionate examination, from a strictly mathematical viewpoint, of the ideas which both men so strongly advocated.

Since about 1970, interest in the economic and social aspects of science has greatly increased, the use of collective biography has been introduced, and science has been described as a joint enterprise directed towards specified goals. In his book *Statistics in Britain 1865-1930*, MacKenzie has explained Gosset's work on the probable error of a mean in terms of enhanced profits for Guinness, and sees his motive as a desire to achieve promotion. The development of the theory of small samples may well have brought financial rewards for Guinness, but Gosset's extension of existing theory most likely arose from the sheer intellectual attraction of the problem. As for motivation, this is always a hazardous field in which to speculate. Gosset was typical of his age in being a loyal servant of the firm. His alleged desire for promotion is unsupported by factual evidence, although there can be few employees in any organization who object to advancement. A brief reflection on the life and work of this remarkable man is enough to make clear his commitment to scientific advance often quite distant from the work of the brewery, and his invariable wish to help friends in circumstances where the consequences for his career were non-existent.

Notes on the text

CHAPTER 1

The biographical memoir of Egon Sharpe Pearson by Bartlett (1981) contains a personal history and reviews the statistical publications. There are also appreciations by Bartlett and Tippett (1981), obituary notices by David (1981) and Johnson (1981), and a tribute by Moore (1975). Personal recollections are gathered by Reid (1982).

CHAPTER 2

The development of statistical methods during the nineteenth century is studied in several books (Cullen 1975; MacKenzie 1981; Porter 1986; Stigler 1986) and reviews (Mairesse 1989), and is the subject of a systematic examination by Sheynin (1980, 1982, 1984a, b, 1985, 1986). Two articles by E. S. Pearson entitled 'Karl Pearson: An appreciation of some aspects of his life and work' were published in the volumes of *Biometrika* for 1936 and 1937 and reissued in book form in 1938. The obituary notice of K. P. by Yule and Filon (1936) and the scientific biography by Eisenhart (1974) are also important sources. Specific aspects of K. P.'s career and outlook are considered by Norton (1978) and Kevles (1985). A bibliography of his statistical and other writings was compiled by G. M. Morant with the assistance of B. L. Welch, and issued in 1939. E. S. Pearson made a selection of K. P.'s early statistical papers which was published in 1948. Appendices to E. S. Pearson (1938) contain the syllabuses of lectures delivered by K. P. at Gresham College and University College London, together with Yule's summary of the lectures he attended (also in Yule 1938). The lecture of 1 November 1892 has been printed (K. Pearson 1941). Papers and correspondence of K. P. held in the manuscripts room of the library of University College London are listed by Merrington *et al.* (1983), and contain all the surviving letters from Gosset to K. P.

1. Various aspects of De Morgan's career are reviewed by Neumann (1984), who appeals for a modern biographer.
2. Ellis was an able mathematician with wide interests, which extended to the best way of constructing a Chinese dictionary; but his health was never

good, and after years of suffering from rheumatism he died at the age of forty-two. His papers were edited by William Walton and published in 1863.

3. Herschel's anonymous essay in the *Edinburgh Review* was reprinted in 1857.

4. Glaisher wrote many papers on definite integrals; he enjoyed the calculation of mathematical tables; and he was interested in the history of mathematics (Forsyth 1929).

5. The term 'minimum squares' was used by Harvey (1822) and in the first Report of the British Association, but Ivory (1825-6), Ellis (1844), Galloway (1846), and Donkin (1857) all used 'least squares'.

6. Airy's book would have been reviewed in *Nature*, had that periodical been issued before 1870.

CHAPTER 3

The main sources for Gosset's life and work are the accounts by Launce McMullen and E. S. Pearson, published in the issue of *Biometrika* for January 1939, together with Pearson's typescript already described and Gosset's correspondence. A slightly modified version of McMullen's account constitutes the foreword to the collection of Student's papers edited by Pearson and John Wishart which appeared in 1942. Most of the obituary notice by Fisher (1939) [CP 165] is concerned with the *t*-distribution and Gosset's interest in the theory of evolution. Letters from Gosset to Fisher between 1915 and 1936, with some of the corresponding letters from Fisher to Gosset, were collected in four volumes, plus a volume of summaries by Fisher and a foreword by McMullen. They were reproduced by Arthur Guinness Son & Co. (Dublin) Ltd. in 1962, and circulated privately with the permission of Mrs Gosset. The originals of Gosset's letters, and Fisher's carbon copies of his own letters, are held in the library of University College London, classmark MS ADD 274. Personal reminiscences of Gosset's nephew George Philpotts are quoted by Cunliffe (1976), who writes about Guinness from the viewpoint of a former employee. Boland (1984) gives a biographical glimpse. The environment in which Gosset worked at the Guinness brewery is described by Joan Fisher Box (1987).

1. Gosset compared (a) the mean deviation and (b) the mean-square deviation as estimators of the modulus of a normal distribution, and he stated that (b) gives a better value 'in proportion 114/100'. This ratio is doubtless taken from Merriman's book on the method of least squares, and originates from Gauss (1816), as explained in §2.4. When Fisher wrote his 1920 paper in which the idea of sufficiency is introduced [CP 12], he was unaware that his comparison of estimates of precision derived from the mean square and the

mean deviation had been anticipated by Gauss over a century earlier, in the sense that posterior variances for uniform priors are asymptotically equal to sampling variances.

2. The quotations are from a letter written by Gosset to Egon Pearson on 11 May 1936.
3. This beginning to Gosset's interest in the exponential limit of the binomial distribution is identified by Joan Fisher Box.
4. Letters from H. G. Lane Poole in 1912 and G. Udny Yule in 1913 asking for Gosset's advice survive in the Pearson papers, list number 284. A long letter from C. E. Lane Poole to E. S. Pearson in 1938 expresses deep appreciation of how greatly Gosset helped him with his difficulties in laying out experiments for forest crops. Perhaps Gosset's most regular correspondent was Edwin S. Beaven, a friend for thirty years, whose papers are kept in the library of the Institute of Agricultural History at the University of Reading.
5. The data on average annual salaries are quoted from Halsey and Trow (1971).
6. Minutes, which include summaries of the papers read at the meetings, and correspondence relating to the dissolution of the society, are preserved in the Pearson papers, list number 254.
7. Maurice G. Kendall (1952) ends his biographical memoir of G. Udny Yule with random extracts from correspondence, one of which follows.

Gosset came in to see me the other day. He is a very pleasant chap. Not at all the autocrat of the *t*-table. . . .

The autocrat of the breakfast table by Oliver Wendell Holmes was famous for much of the nineteenth and early twentieth centuries.

CHAPTER 4

1. The life and work of William Fleetwood Sheppard are described by N. F. Sheppard *et al.* (1937). His correspondence with Galton is discussed by MacKenzie (1981).
2. Herbert Edward Soper was a statistician of ability and character, to both of which Greenwood (1931) does full justice. A contemporary view of Soper (1913) is expressed in a letter from Yule to Gosset dated 13 May 1913 (Pearson papers, list number 284). Gosset had written to say that Soper's article gave little help for the inverse problem, and Yule suggested that the prior distribution of R could be determined empirically in preference to the assumption of a uniform distribution.
3. Pearson's collaborators, often unpaid workers, were mostly women. Beatrice Mabel Cave and Frances Evelyn Cave-Browne-Cave were daughters

of Sir Thomas Cave-Brown-Cave (*Who was who, Burke's peerage, baronetage and knighthood*). Alice Lee played an important part in work for K.P., accompanied the Pearson family on holidays, and contributed ballads to E.S.P.'s youthful journal *Biochronicle*. Her scientific career is examined by Love (1979).

4. In 1896, Pearson derived his 'best value' of the correlation coefficient by maximizing the posterior distribution when all prior distributions are uniform. He later decided that this 'Gaussian rule' was 'logically at fault' because the choice of a *uniform* prior distribution was arbitrary. When Kirstine Smith compared methods of estimating parameters in her paper of May 1916, she favoured minimum χ^2 in preference to the Gaussian principle on the grounds that finite probabilities were used instead of probability densities. Fisher argued that the value of χ^2 depends on the way in which the data are grouped, and that his absolute criterion 'derived from the Principle of Inverse Probability' eliminated the need for arbitrary grouping. Pearson replied that the logic of the 'Gaussian rule' had to be demonstrated, and he disagreed with Fisher's arguments about the grouping. Although Fisher submitted a further justification of his criticism, K.P. rejected the paper, which he described as controversial. While the advantages of maximum likelihood have since 1922 been made abundantly clear, the theory of estimation in 1916 rested on a much less secure foundation, and was prone to confusion and misunderstanding. The details of this controversy between K.P. and Fisher are discussed by Egon Pearson (1968) and Edwards (1974).

5. Karl Pearson's lectures on Condorcet, delivered in the autumn of 1927 and published in 1978, include the following remarks on the subject of uniform prior distributions (pp. 499–500).

... $\varphi(x)$ represents our past experience of the distribution of x 's and is not a constant, i.e. we distribute our ignorance not equally but according to past experience of x .

I know I have been preaching this doctrine vainly in the wilderness for many years, and made a distinguished statistician a permanent enemy by suggesting it, but I believe it to be correct.

6. The history of the variate difference method is reviewed by Yule (1921) and Tintner (1940). Obituary notices of Reginald Hawthorn Hooker (Yule 1944) and Oskar Johann Victor von Anderson (Wold 1961) give details of their careers. Chuprov's papers at Moscow University contain a letter from Anderson (in Russian) dated 17 June 1914 about the submission of his paper to *Biometrika*, also a copy of K.P.'s reply. We thank Oscar B. Sheynin for this information.

7. Later, Sir Daniel Hall, Principal of Wye College, Director of Rothamsted Experimental Station (Russell 1966).

8. There is a brief biography of Edward Gordon Peake in *Alumni Cantabrigiensis* (Venn 1953).

9. Siméon Denis Poisson (1837) considered a sequence of independent trials with probability p for success and q for failure. He expressed the upper tail of a binomial distribution for the number of successes in μ trials as the lower tail of a negative binomial distribution for the number of trials to attain m successes, and from the latter derived the cumulative exponential limit when $q \rightarrow \theta$ and $\mu \rightarrow \alpha$ such that $q\mu = \omega$. Simon Newcomb (1860) suggested the limit distribution as a fit to data, and Ernst Abbe (1878) applied it to haemocytometer counts of blood corpuscles. The exponential limit of the binomial distribution was credited to Poisson by von Bortkiewicz (1898), who fitted the Poisson distribution to suicides of German women and deaths of Prussian soldiers from the kick of a horse. These results were unknown to either Gosset or Pearson in 1907, or to Bateman in 1910. An accusation of censoring in von Bortkiewicz's illustrations, implicitly made by Whitaker under K.P.'s influence, was refuted in 1915.

10. In 1898, von Bortkiewicz asserted that a set of observations from different Poisson distributions behaves like a sample from a single Poisson distribution, and he called this discovery the law of small numbers. Further details concerning Poisson's distribution and the law of small numbers are given by Haight (1967), Stigler (1982), Seneta (1983), and Quine and Seneta (1987).

11. Ethel Mary Elderton was sister to William P. Elderton, the actuary. She contributed much to the work of the Eugenics Laboratory. Her career is summarized in *Who was who* and assessed by Love (1979).

12. Kirstine Smith was born 1878 in a small town in Jutland, and graduated in mathematics from the University of Copenhagen in 1903. Private secretary to T.N. Thiele 1903–4. Assistant at Bureau internationale pour l'exploration de la mer 1904–15. Studied at University College under K.P. 1915–17, awarded degree January 1918. Employed during 1918–24 in Copenhagen by the Committee for Marine Investigations and by the Carlsberg Laboratory, a research organization financed by the breweries. Her career ends as a teacher: 1925–30 Master of Mathematics at the High School in Tønder, Southern Jutland; 1930–9 Senior Master at Aurehøj High School, in a suburb of Copenhagen. Died 1939. We thank Anders Hald for this information. Her letters to K.P. (Pearson papers, list number 857/6) show that (i) when she returned to England in the winter of 1916, she stayed mostly in the English Lake District, because the London climate had an adverse effect on her nose and throat, and (ii) he offered her a job in 1920. Kiefer (1959) acknowledged her contribution to optimal design.

13. The correspondence between Leonard Darwin and Fisher is printed with explanatory notes by Bennett (1983).

CHAPTER 5

A selection of what Fisher considered to be his most outstanding statistical papers was published in 1950, with an index prepared by John Tukey. All Fisher's papers are collected in five volumes edited by Bennett (1971–4), and each paper has a CP number. The account of Fisher's life and work by Joan Fisher Box (1978) is essential reading, and covers both Fisher's statistical work and his contributions to evolutionary biology. Among the reviews of this book, perhaps the most thorough is by William Kruskal (1980), although he confines himself to the statistical side of Fisher's work. Other general commentators include: Yates and Mather (1963), Savage (1976), Fienberg and Hinkley (1980), and MacKenzie (1981). Both Churchill Eisenhart (1979) and Joan Fisher Box (1981) study the development of the *t*-test from Student's *z*-test. Aspects of Fisher's early career are considered by E. S. Pearson (1968, 1974), and Joan Fisher Box (1980) traces his work on the design of experiments between 1922 and 1926. The history of agricultural science by Russell (1966) can be supplemented by the reviews of Cochran (1976) for comparative experimentation, and Gower (1988) for the relationship between statistics and agricultural research.

1. Fisher warmly acknowledged in the preface the notes and corrections Gosset made during proof-reading, but the wording varies from one edition to another.

[First edition, 1925]

I owe more than I can say to Mr W. S. Gosset, Mr E. Somerfield and Miss W. A. Mackenzie, who have read the proofs and made valuable suggestions. Many small but troublesome errors have been removed; . . .

[Ninth edition, 1944]

With the encouragement of my colleagues, and the valued help of the late W. S. Gosset ('Student'), his assistant Mr E. Somerfield, and Miss W. A. Mackenzie, the first edition was prepared and weathered the hostile criticisms inevitable to such a venture.

Section 5, originally a list of mathematical tables, later became a 'Historical Note', and ended with a review of Gosset's work on exact sampling distributions.

[Thirteenth edition, 1958]

'Student's work was not quickly appreciated (it had, in fact, been totally ignored in the journal in which it had appeared), and from the first edition it has been one of the chief purposes of this book to make better known the effect of his researches, . . .

The sentence in parentheses does not appear in the ninth edition, otherwise the wording is the same. In fact, the lead which Student had given was

acknowledged in *Biometrika*, both in the contents of Soper (1913), and in the subtitles of K. P. (1915) and the 'Cooperative Study' (1917). Egon Pearson made the following comment on 20 October 1979.

But as far as I can see nowhere says how much he, the great mathematician, owed to the experimenter 'Student', who in turn had learnt from E. S. Beaven.

2. The first edition of *Statistical methods for research workers* was reviewed by E. S. Pearson in *Science Progress*. Harold Hotelling reviewed the first seven editions with enthusiasm in *Journal of the American Statistical Association*.

3. By modern standards, the analysis of variance in this paper is open to criticism (Yates and Mather 1963; Cochran 1980). The nested design called for two estimates of error but they were combined. Dung and potash were not distinguished as separate factors. No randomization was used.

4. What constitutes analysis of variance is a question not yet satisfactorily resolved (Speed 1987), and so the detection of examples in the past is a hazardous occupation. There is no continuous line of development prior to Fisher's work, but some achievements are worth recording. Airy (1861) described a model similar to one with two components of variance. Lexis worked on the stability of statistical series from 1876, and the dispersion theory which he originated anticipates the analysis of one-way classifications (Heyde and Seneta 1977: §3.4; Stigler 1986). Edgeworth's results for two-way classifications were presented in 1885 (Stigler 1978, 1986). Thiele's work was published in Danish from 1889 onwards, and he analysed a general form of two-way classification. The poor English translation (1903) of his book on the theory of observations only discussed a simpler model (Hald 1981: §7). Agricultural field experiments provided a suitable medium for strong growth and transplantation elsewhere.

5. Balanced views of the controversy between Gosset and Fisher are presented by Yates and Mather (1963) and Cochran (1976). The debate remains inconclusive. Some experimenters continue to recommend randomization (Pearce 1983), while others still prefer systematic arrangements (Hurlbert 1984). As a compromise, methods of restricted randomization have been devised which exclude undesirable experimental layouts (Bailey 1987). Optimal experimental designs have been developed for situations where the errors are assumed to be correlated (Kunert 1985; Azzalini and Giovagnoli 1987).

Bibliography

- Abbe, E. (1878). Ueber Blutkörper-Zählung. *Gesammelte Abhandlungen*, 1, 173–80.
- Adrain, R. (1808). Research concerning the probabilities of the errors which happen in making observations, &c. *The Analyst; or, Mathematical Museum*, 1, 93–109. Reprinted in Stigler (1980) (q.v.), Vol. 1.
- Airy, G. B. (1861). *On the algebraical and numerical theory of errors of observations and the combination of observations*. Macmillan, London. (2nd edn, 1875; 3rd edn, 1879).
- Anderson, O. (1914). Nochmals über 'The elimination of spurious correlation due to position in time and space'. *Biometrika*, 10, 269–79.
- Azzalini, A. and Giovagnoli, A. (1987). Some optimal designs for repeated measurements with autoregressive errors. *Biometrika*, 74, 725–34.
- Bailey, R. A. (1987). Restricted randomization: A practical example. *Journal of the American Statistical Association*, 82, 712–19.
- Bartlett, M. S. (1981). Egon Sharpe Pearson, 1895–1980. *Biographical Memoirs of Fellows of the Royal Society*, 27, 425–43.
- Bartlett, M. S. and Tippett, L. H. C. (1981). Egon Sharpe Pearson, 1895–1980. *Biometrika*, 68, 1–12.
- Bartlett, S. (1931). Nutritional value of raw and pasteurized milk. *Journal of the Ministry of Agriculture*, 38, 60–4.
- Bateman, H. (1910). Note on the probability distribution of α -particles. *Philosophical Magazine*, [6] 20, 704–7.
- Bayes, T. (1764). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions*, 53, 370–418. Reproduced with biographical note by G. A. Barnard in *Biometrika*, 45, 293–315. Reprinted in Pearson and Kendall (1970) (q.v.), pp. 131–53.
- Bennett, J. H. (ed.) (1983). *Natural selection, heredity, and eugenics. Including selected correspondence of R. A. Fisher with Leonard Darwin and others*. Clarendon Press, Oxford.
- Blakeman, J. and Pearson K. (1906). On the probable error of mean-square contingency. *Biometrika*, 5, 191–7.
- Boland, P. J. (1984). A biographical glimpse of William Sealy Gosset. *American Statistician*, 38, 179–83.
- Boole, G. (1854). *An investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities*. Walton and Maberley, London. Reprinted by Dover Publications, New York (1958).
- Bortkiewicz, L. von (1898). *Das Gesetz der kleinen Zahlen*. Teubner, Leipzig.
- Bortkiewicz, L. von (1915). Realismus und Formalismus in der mathematischen Statistik. *Allgemeines Statistisches Archiv. (Deutschen Statistischen Gesellschaft)*, 9, 225–56.
- Bowman, K. O. and Shenton, L. R. (1987). *Properties of estimators of the gamma distribution*, Marcel Dekker, New York.
- Box, J. F. (1978). *R. A. Fisher: The life of a scientist*. Wiley, New York.

- Box, J.F. (1980). R. A. Fisher and the design of experiments, 1922–1926. *American Statistician*, 34, 1–7.
- Box J.F. (1981). Gosset, Fisher and the t distribution. *American Statistician*, 35, 61–6.
- Box, J.F. (1987). Guinness, Gosset, Fisher, and small samples. *Statistical Science*, 2, 45–52.
- Cave, B.M. and Pearson, K. (1914). Numerical illustrations of the variate difference correlation method. *Biometrika*, 10, 340–55.
- Cave-Browne-Cave, F.E. (1904). On the influence of the time factor on the correlation between the barometric heights at stations more than 1000 miles apart. *Proceedings of the Royal Society, Series A*, 74, 403–13.
- Cochran, W.G. (1976). Early development of techniques in comparative experimentation. *On the history of statistics and probability* (ed. D.B. Owen), pp. 1–25. Marcel Dekker, New York.
- Cochran, W. G. (1980). Fisher and the analysis of variance. *R. A. Fisher: An appreciation* (ed. S. E. Fienberg and D. V. Hinkley), pp. 17–34. Springer-Verlag, New York.
- Cullen, M. J. (1975). *The statistical movement in early Victorian Britain: The foundations of empirical social research*. Harvester, Hassocks, Sussex.
- Cunliffe, S. V. (1976). Interaction. *Journal of the Royal Statistical Society, A*, 139, 1–19.
- David, H. A. (1981). In memoriam: Egon S. Pearson, 1895–1980. *American Statistician*, 35, 94–5.
- De Morgan, A. (1837). Theory of probabilities. *Encyclopaedia metropolitana*, Vol. 2, pp. 359–458. (2nd edn, 1845, Vol. 2, pp. 393–490).
- De Morgan, A. (1838). *An essay on probabilities and on their application to life contingencies and insurance offices*. Lardner's Cabinet Cyclopaedia, London.
- De Morgan, A. (1847). *Formal logic: Or the calculus of inference, necessary and probable*. Taylor and Walton, London.
- Donkin, W. F. (1857). On an analogy relating to the theory of probabilities and on the principle of the method of least squares. *Quarterly Journal of Pure and Applied Mathematics*, 1, 152–62.
- Edgeworth, F. Y. (1884). The philosophy of chance. *Mind*, 9, 223–35.
- Edgeworth, F. Y. (1885). Methods of statistics. *Jubilee volume of the Statistical Society*, pp. 181–217.
- Edgeworth, F. Y. (1897). Miscellaneous applications of the calculus of probabilities. *Journal of the Royal Statistical Society*, 60, 681–98.
- Edgeworth, F. Y. (1907). Statistical observations on wasps and bees. *Biometrika*, 5, 365–86.
- Edwards, A. W. F. (1974). The history of likelihood. *International Statistical Review*, 42, 9–15.
- Eisenhart, C. (1974). Karl Pearson. *Dictionary of scientific biography*, Vol. 10, pp. 447–73.
- Eisenhart, C. (1979). On the transition from 'Student's' z to 'Student's' t . *American Statistician*, 33, 6–10.
- Elderton, E. M. (1933). The Lanarkshire milk experiment. *Annals of Eugenics, London*, 5, 326–36.

- Elderton, E. M. and Pearson, K. (1907). On the measure of the resemblance of first cousins. *Eugenics Laboratory Memoirs*, IV.
- Elderton, E. M. and Pearson, K. (1910). A first study of the influence of parental alcoholism on the physique and ability of the offspring. *Eugenics Laboratory Memoirs*, X.
- Ellis, R. L. (1844). On the method of least squares. *Transactions of the Cambridge Philosophical Society*, 8, 204–19. Reprinted in Ellis (1863) (q.v.).
- Ellis, R. L. (1863). *The mathematical and other writings of Robert Leslie Ellis* (ed. W. Walton). Deighton, Bell, Cambridge.
- Fienberg, S. E. and Hinkley, D. V. (eds.) (1980). *R. A. Fisher: An appreciation*. Springer-Verlag, New York.
- Fisher, R. A. (1939). 'Student'. *Annals of Eugenics, London*, 9, 1–9.
- Fisher, R. A. (1950). *Contributions to mathematical statistics*. Index prepared by John Tukey. Wiley, New York.
- Fisher, R. A. (1971–4). *Collected papers of R. A. Fisher* (ed. J. H. Bennett). University of Adelaide.
- Forsyth, A. R. (1929). James Whitbread Lee Glaisher. *Journal of the London Mathematical Society*, 4, 101–12.
- Galloway, T. (1846). On the application of the method of least squares to the determination of the probable errors of observations in a portion of the Ordnance Survey of England. *Memoirs of the Royal Astronomical Society*, 15, 23–69.
- Galton, F. (1889). *Natural inheritance*. Macmillan, London.
- Galton, F. (1902). The most suitable proportion between the values of first and second prizes. *Biometrika*, 1, 385–90.
- Gauss, C. F. (1809). *Theoria motus corporum coelestium &c.* Perthes and Besser, Hamburg.
- Gauss, C. F. (1816). Bestimmung der Genauigkeit der Beobachtungen, *Zeitschrift für Astronomie und verwandte Wissenschaften*, 1, 187–97.
- Gauss, C. F. (1823). *Theoria combinationis observationum erroribus minimis obnoxiae*. Göttingen.
- Glaisher, J. W. L. (1872). On the law of facility of errors of observations, and on the method of least squares. *Memoirs of the Royal Astronomical Society*, 39, 75–123.
- Gower, J. C. (1988). Statistics and agriculture. *Journal of the Royal Statistical Society A*, 151, 179–200.
- Greenwood, M. (1931). Herbert Edward Soper. *Journal of the Royal Statistical Society*, 94, 135–43.
- Greenwood, M. and White, J. D. C. (1909). A biometric study of phagocytosis with special reference to the 'opsonic index'. First memoir. On the frequency distributions of phagocytic counts. *Biometrika*, 6, 376–401.
- Greenwood, M. and White, J. D. C. (1910). A biometric study of phagocytosis with special reference to the 'opsonic index'. Second memoir. On the distribution of the means of samples. *Biometrika*, 7, 505–30.
- Greenwood, M. and Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*, 83, 255–79.

- Haight, F. A. (1967). *Handbook of the Poisson distribution*. Wiley, New York.
- Hald, A. (1981). T.N. Thiele's contributions to statistics. *International Statistical Review*, **49**, 1–20.
- Halsey, A. H. and Trow, M. A. (1971). *The British academics*. Faber and Faber, London.
- Harvey, G. (1822). On the method of minimum squares, &c. *Edinburgh Philosophical Journal*, **7**, 292–301.
- Helmert, F. R. (1876). Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers directer Beobachtungen gleicher Genauigkeit. *Astronomische Nachrichten*, **88**, 113–32.
- [Herschel, J. F. W.] (1850). Quetelet on probabilities. *Edinburgh Review*, **92**, 1–57.
- Herschel, J. F. W. (1857). *Essays from the Edinburgh and quarterly reviews, with addresses and other pieces*. Longman, Brown, and Co., London.
- Heyde, C. C. and Seneta, E. (1977). *I. J. Bienaymé: Statistical theory anticipated*. Springer-Verlag, New York.
- Hooker, R. H. (1901a). Correlation of the marriage rate with trade. *Journal of the Royal Statistical Society*, **64**, 485–92.
- Hooker, R. H. (1901b). The suspension of the Berlin Produce Exchange and its effect upon corn prices (with discussion). *Journal of the Royal Statistical Society*, **64**, 574–613.
- Hooker, R. H. (1905). On the correlation of successive observations; illustrated by corn prices. *Journal of the Royal Statistical Society*, **68**, 696–703.
- Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, **54**, 187–211.
- Ivory, J. (1825–6). On the method of the least squares. *Philosophical Magazine*, **65**, 3–10, 81–8, 161–8; **68**, 161–5.
- Jevons, W. S. (1874). *The principles of science. A treatise on logic and scientific method*. Macmillan, London. (2nd edn. 1877).
- Johnson, N. L. (1981). E. S. Pearson, 1895–1980. *Journal of the Royal Statistical Society A*, **144**, 270–1.
- Johnston, J. F. W. (1849). *Experimental agriculture, being the result of past and suggestions for future experiments in scientific and practical agriculture*, Blackwood, Edinburgh.
- Kempthorne, O. (1983). A review of R. A. Fisher: An appreciation. *Journal of the American Statistical Association*, **78**, 482–90.
- Kendall, M. G. (1952). George Udny Yule, 1871–1951. *Journal of the Royal Statistical Society A*, **115**, 156–161. Reprinted in Pearson and Kendall (1970) (q.v.), pp. 419–25.
- Kevles, D. J. (1985). *In the name of eugenics*. Alfred Knopf, New York.
- Kiefer, J. (1959). Optimum experimental designs (with discussion). *Journal of the Royal Statistical Society B*, **21**, 272–319.
- Kruskal, W. (1980). The significance of Fisher: A review of R. A. Fisher: *The life of a scientist*. *Journal of the American Statistical Association*, **75**, 1019–30.
- Kunert, J. (1985). Optimal experimental design when the errors are assumed to be correlated. *Statistics and Decisions*, Suppl. Issue No. 2, 287–98.
- Laplace, P. S. (1774). Mémoire sur la probabilité des causes par les événements. *Mémoires de l'Académie Royale des Sciences de Paris (Savants étrangers)*, **6**, 621–56.
- Laplace, P. S. (1810). Mémoire sur les approximations des formules qui sont fonctions

- de très-grand nombres, et leur application aux probabilités. *Mémoires de l'Académie des Sciences de Paris*, Année, 1809, 353–415, 559–65.
- Laplace, P. S. (1811). Mémoire sur les intégrales définies, et leur application aux probabilités, et spécialement à la recherche du milieu qu'il faut choisir entre les résultats des observations. *Mémoires de l'Institut Impériale de France*, Année 1810, 279–347.
- Laplace, P. S. (1812). *Théorie analytique des probabilités*. Courcier, Paris. (2nd edn, 1814, includes *Essai philosophique sur les probabilités*; 3rd edn, 1820.
- Levasseur, E. L. (1885). La statistique graphique. *Jubilee volume of the Statistical Society*, pp. 218–50.
- Love, R. (1979). 'Alice in eugenics-land': feminism and eugenics in the scientific careers of Alice Lee and Ethel Elderton. *Annals of Science*, 36, 145–58.
- [Lubbock, J. W. and Drinkwater, J. E.] (1830). *On probability*. Society for Diffusion of Useful Knowledge, London.
- Lupton, S. (1898). *Notes on observations &c*. Macmillan, London.
- MacKenzie, D. A. (1981). *Statistics in Britain 1865–1930. The social construction of scientific knowledge*. Edinburgh University Press.
- McMullen, L. (1939). William Sealy Gosset, 1876–1937. (1) 'Student' as a man. *Biometrika*, 30, 205–10. Reprinted in Pearson and Kendall (1970) (q.v.), 354–60.
- Mairesse, J. (ed.) (1989). Statistical methodology in Europe, 1789–1889. *Bulletin of the International Statistical Institute*, 53 (1), 119–76.
- Marey, E. J. (1878). *La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine*. Masson, Paris. (2nd edn, 1885).
- Mercer, W. B. and Hall, A. D. (1911). The experimental error of field trials. *Journal of Agricultural Science*, 4, 107–27.
- Merriman, M. (1884). *A textbook on the method of least squares*. Wiley, New York. (2nd edn, 1886; 8th edn, 1900).
- Merrington, M., Blundell, B., Burrough, S., Golden J., and Hogarth, J. (Comp., 1983). *A list of the papers and correspondence of Karl Pearson (1857–1936) held in The Manuscripts Room, University College London Library*. University College London.
- Moore, P. G. (1975). A tribute to Egon Sharpe Pearson (born August 11th, 1895). *Journal of the Royal Statistical Society A*, 138, 129–30.
- Morant, G. M. and Welch, B. L. (1939). *A bibliography of the statistical and other writings of Karl Pearson*. Cambridge University Press.
- Neumann, B. H. (1984). Augustus De Morgan. *Bulletin of the London Mathematical Society*, 16, 575–89.
- Newcomb, S. (1860). Notes on the theory of probabilities. *The Mathematical Monthly*, 2, 134–40. Reprinted in Stigler (1980) (q.v.), Vol. 2.
- Norton, B. J. (1978). Karl Pearson and statistics: The social origins of scientific innovation. *Social Studies of Science*, 8, 3–34.
- Pearce, S. C. (1983). *The agricultural field experiment: A statistical examination of theory and practice*. Wiley, New York.
- Pearson, E. S. (1926). R. A. Fisher: *Statistical methods for research workers*. *Science Progress*, 20, 733–4.
- Pearson, E. S. (1931a). The analysis of variance in cases of non-normal variation. *Biometrika*, 23, 114–33.

- Pearson, E. S. (1931*b*). The test of significance for the correlation coefficient. *Journal of the American Statistical Association*, 26, 128–34.
- Pearson, E. S. (1932). The percentage limits for the distribution of range in samples from a normal population ($n \leq 100$). *Biometrika*, 24, 404–17.
- Pearson, E. S. (1938). *Karl Pearson: An appreciation of some aspects of his life and work*. Cambridge University Press.
- Pearson, E. S. (1939). William Sealy Gosset. 1876–1937. (2) 'Student' as statistician. *Biometrika*, 30, 210–50. Reprinted in Pearson and Kendall (1970) (q.v.), pp. 360–403.
- Pearson, E. S. (1965). Some incidents in the early history of biometry and statistics, 1890–94. *Biometrika*, 52, 3–18. Reprinted in Pearson and Kendall (1970) (q.v.), pp. 323–38.
- Pearson, E. S. (1967). Some reflections on continuity in the development of mathematical statistics, 1885–1920. *Biometrika*, 54, 341–55. Reprinted in Pearson and Kendall (1970) (q.v.), pp. 339–53.
- Pearson, E. S. (1968). Some early correspondence between W. S. Gosset, R. A. Fisher and Karl Pearson, with notes and comments. *Biometrika*, 55, 445–57. Reprinted in Pearson and Kendall (1970) (q.v.), pp. 405–17.
- Pearson, E. S. (1974). Memories of the impact of Fisher's work in the 1920s. *International Statistical Review*, 42, 5–8.
- Pearson, E. S. and Adyanthāya, N. K. (1928). The distribution of frequency constants in small samples from symmetrical populations. *Biometrika*, 20A, 356–60.
- Pearson, E. S. and Adyanthāya, N. K. (1929). The distribution of frequency constants in small samples from non-normal symmetrical and skew populations. 2nd paper: The distribution of 'Student's' z . *Biometrika*, 21, 259–86.
- Pearson, E. S., Chesire, L. and Oldis, E. (1932). Further experiments on the sampling distribution of the correlation coefficient. *Journal of the American Statistical Association*, 27, 121–8.
- Pearson, E. S. and Hartley, H. O. (1943). Tables of the probability integral of the studentized range. *Biometrika*, 33, 89–99.
- Pearson, E. S. and Kendall, M. G. (eds.) (1970). *Studies in the history of statistics and probability*. Griffin, London.
- Pearson, K. (1902). Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society of London, A*, 200, 1–66.
- Pearson, K. (1907). Mathematical contributions to the theory of evolution. XVI. On further methods of determining correlation. *Drapers' Company Research Memoirs, Biometric Series*, IV.
- Pearson, K. (1914). On certain errors with regard to multiple correlation occasionally made by those who have not adequately studied this subject. *Biometrika*, 10, 181–7.
- Pearson, K. (1915*a*). On the distributions of the standard deviations of small samples: Appendix I to papers by 'Student' and R. A. Fisher. *Biometrika*, 10, 522–9.
- Pearson, K. (1915*b*). On certain types of compound frequency distributions in which the components can be individually described by binomial series. *Biometrika*, 11, 139–44.

- Pearson, K. (1931a). On the nature of the relationship between two of 'Student's' variates (z_1 and z_2) when samples are taken from a bivariate normal population. *Biometrika*, 22, 405–22.
- Pearson, K. (1931b). Further remarks on the 'z' test. *Biometrika*, 23, 408–15.
- Pearson, K. (1931c). Historical note on the distribution of the standard deviations of samples of any size drawn from an indefinitely large normal parent population. *Biometrika*, 23, 416–8.
- Pearson, K. (1941). The laws of chance, in relation to thought and conduct. *Biometrika*, 32, 89–100.
- Pearson, K. (1948). *Karl Pearson's early statistical papers*. Cambridge University Press. Reprinted 1956.
- Pearson, K. (1978). *The history of statistics in the 17th and 18th centuries against the changing background of intellectual, scientific and religious thought*. (ed. E. S. Pearson). Griffin, London.
- Pearson, K. and Filon, L. N. G. (1898). Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philosophical Transactions of the Royal Society of London A*, 191, 229–311. Reprinted in Pearson (1948) (q.v.), pp. 179–261.
- Pearce, B. (1852). Criterion for the rejection of doubtful observations. *Astronomical Journal*, 2, 161–3. Reprinted in Stigler (1980). (q.v.), Vol. 2.
- Persons, W. M. (1917). On the variate difference correlation method and curve-fitting. *Quarterly Publications of the American Statistical Association*, 15, 602.
- Peters, C. A. F. (1856). Über die Bestimmung des wahrscheinlichen Fehlers einer Beobachtung aus den Abweichungen der Beobachtungen von ihrem arithmetischen Mittel. *Astronomische Nachrichten*, 44, 1034, 29–32.
- Poisson, S. D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civiles, précédées des règles générales du calcul des probabilités*. Bachelier, Paris.
- Porter, T. M. (1986). *The rise of statistical thinking 1820–1900*. Princeton University Press.
- Quetelet, A. (1846). *Lettres à S. A. R. le Duc Régnant de Saxe-Cobourg et Gotha, sur la théorie des probabilités, appliquée aux sciences morales et politiques*. Hayez, Bruxelles. Translated into English by O. G. Downes, 1849.
- Quine, M. P. and Seneta, E. (1987). Bortkiewicz's data and the law of small numbers. *International Statistical Review*, 55, 173–181.
- Reid, C. (1982). *Neyman—from life*. Springer-Verlag, New York.
- Russell, Sir E. John (1966). *A history of agricultural science in Great Britain 1620–1954*. Allen and Unwin, London.
- Savage, L. J. (1976). On rereading R. A. Fisher (with discussion). *Annals of Statistics*, 4, 441–500.
- Seneta, E. (1983). Modern probabilistic concepts in the work of E. Abbe and A. de Moivre. *Mathematical Scientist*, 8, 75–80.
- Sheppard, N. F., Aitken, A. C. and Fisher, R. A. (1937). W. F. Sheppard, F. R. S. E., Sc.D., L.L.M. *Annals of Eugenics, London*, 8, 1–14.
- Sheppard, W. F. (1898). On the geometrical treatment of the 'normal curve' of

- statistics, with especial reference to correlation and to the theory of error. *Proceedings of the Royal Society*, 62, 170-3.
- Sheynin, O. B. (1980). On the history of the statistical method in biology. *Archive for History of Exact Sciences*, 22, 323-71.
- Sheynin, O. B. (1982). On the history of medical statistics. *Archive for History of Exact Sciences*, 26, 241-86.
- Sheynin, O. B. (1984a) On the history of the statistical method in astronomy. *Archive for History of Exact Sciences*, 29, 151-99.
- Sheynin, O. B. (1984b). On the history of the statistical method in meteorology. *Archive for History of Exact Sciences*, 31, 53-95.
- Sheynin, O. B. (1985) On the history of the statistical method in physics. *Archive for History of Exact Sciences*, 33, 351-82.
- Sheynin, O. B. (1986). A. Quetelet as a statistician. *Archive for History of Exact Sciences*, 36, 281-325.
- Smith, K. (1916). On the 'best' values of the constants in frequency distributions. *Biometrika*, 11, 262-76.
- Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, 12, 1-85.
- Soper, H. E. (1913). On the probable error of the correlation coefficient to a second approximation. *Biometrika*, 9, 91-115.
- Soper, H. E. (1914). Tables of Poisson's exponential binomial limit. *Biometrika*, 10, 25-35.
- Soper, H. E., Young, A. W., Cave, B. M., Lee, A., and Pearson, K. (1917). On the distribution of the correlation coefficient in small samples. Appendix II to the papers of 'Student' and R. A. Fisher. A Cooperative Study. *Biometrika*, 11, 328-413.
- 'Sophister' (1928). Discussion of small samples drawn from an infinite skew population. *Biometrika*, 20A, 389-423.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101.
- Speed, T. P. (1987). What is an analysis of variance? (with discussion). *Annals of Statistics*, 15, 885-941.
- Stigler, S. M. (1978). Francis Ysidro Edgeworth, statistician (with discussion). *Journal of the Royal Statistical Society A*, 141, 287-322.
- Stigler, S. M. (ed.) (1980). *American contributions to mathematical statistics in the nineteenth century*. Arno Press, New York.
- Stigler, S. M. (1982). Poisson on the Poisson distribution. *Statistics and Probability Letters*, 1, 33-5.
- Stigler, S. M. (1986). *The history of statistics. The measurement of uncertainty before 1900*. Harvard University Press, Cambridge, Massachusetts.
- 'Student' (1926). R. A. Fisher: *Statistical methods for research workers*. *Eugenics Review*, 18, 148-50.
- 'Student' (1929). Robert R. Kuczynski: *The balance of births and deaths*. Vol. 1. *Western and northern Europe*. *Eugenics Review*, 21, 130-1.
- 'Student' (1942). *'Student's' collected papers* (ed. E. S. Pearson and John Wishart, with a foreword by Launce McMullen). Issued by the Biometrika Office, University College

- London, and printed in Great Britain at the University Press, Cambridge. Reprinted 1958.
- Thiele, T. N. (1903). *Theory of observations*. Layton, London.
- Tintner, G. (1940). *The variate difference method*. Principia Press, Bloomington, Indiana.
- Tolley, H. R. (1929). Economic data from the sampling point of view. *Journal of the American Statistical Association*, 24, Suppl. 69–72.
- Venn, J. (1866). *The logic of chance &c.* Macmillan, London. (2nd edn, 1876; 3rd edn, 1888).
- Venn, J. A. (1953). *Alumni cantabrigienses*. Cambridge University Press.
- Welch, B. L. (1958). 'Student' and small sample theory. *Journal of the American Statistical Association*, 53, 777–88.
- Welch, B. L. (1970). Statistics—a vocational or a cultural study? (with discussion). *Journal of the Royal Statistical Society A*, 133, 531–54.
- Westergaard, H. (1890). *Die Grundzüge der Theorie der Statistik*. Gustav Fischer, Jena.
- Whitaker, L. (1914). On the Poisson law of small numbers. *Biometrika*, 10, 36–71.
- Whitworth, W. A. (1867). *Choice and chance. An elementary treatise on permutations, combinations, and probability*. Deighton, Bell, Cambridge. (2nd edn, 1870; 5th edn, 1901).
- Wold, H. (1961). Oskar Anderson, 1887–1960. *Annals of mathematical Statistics*, 32, 651–60.
- Wood, T. B. and Stratton, F. J. M. (1910). The interpretation of experimental results. *Journal of Agricultural Science*, 3, 417–40.
- Wright, A., Lee, A., and Pearson, K. (1907). A comparative study of queens, drones and workers in *Vespa vulgaris*. *Biometrika*, 5, 407–22.
- Yates, F. (1939). The comparative advantages of systematic and randomized arrangements in the design of agricultural and biological experiments. *Biometrika*, 30, 440–66.
- Yates, F. (1951). The influence of *Statistical methods for research workers* on the development of the science of statistics. *Journal of the American Statistical Association*, 46, 19–34.
- Yates, F. (1967). A fresh look at the basic principles of the design and analysis of experiments. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (ed. L. M. Le Cam and J. Neyman), Vol. 4, pp. 777–90. Reprinted in Yates (1970) (q.v.), pp. 267–80.
- Yates, F. (1970). *Experimental design: Selected papers of Frank Yates, C. B. E., F. R. S., Griffin*, London.
- Yates, F. and Mather K. (1963). Ronald Aylmer Fisher, 1890–1962. *Biographical Memoirs of Fellows of the Royal Society*, 9, 91–120.
- Young, A. W. (1916). Note on the standard deviations of samples of two or three. *Biometrika*, 11, 277–80.
- Young, A. (1771). *A course of experimental agriculture*. Exshaw and Co., Dublin.
- Yule, G. U. (1921). On the time-correlation problem with especial reference to the variate-difference method (with discussion). *Journal of the Royal Statistical Society*, 84, 497–537.
- Yule, G. U. (1938). Notes of Karl Pearson's lectures on the theory of statistics, 1894–96. *Biometrika*, 30, 198–203.

- Yule, G. U. (1944). Reginald Hawthorn Hooker, M. A. *Journal of the Royal Statistical Society*, **107**, 74–7.
- Yule, G. U. and Filon, L. N. G. (1936). Karl Pearson 1857–1936. *Obituary Notices of the Royal Society of London*, **2**, 73–110.
- Zeuner, G. (1869). *Abhandlungen aus der mathematischen Statistik*. Arthur Felix, Leipzig.