devoid of malice. He rarely spoke about personal matters but when he did his opinion was well worth listening to and not in the least superficial.

In the summer of 1934 he had a motor accident and broke the neck of his femur. He had to lie up for three months, of course working at statistics, and was a semi-cripple for a year. This was particularly irksome for such an active man, as was the sheer unnecessariness of the accident, for he ran into a lamp-post on a straight road, through looking down to adjust some stuff he was carrying; but with great hard work and persistence he eventually reduced the disability to a slight limp.

At the end of 1935 he left Ireland to take charge of the new Guinness brewery in London, and I saw comparatively little of him after that. The departure from Ireland of "Student" and his family was a great loss to many who had experienced their hospitality.

His work in London was necessarily very hard and accompanied by all the vexations inevitably associated with a big undertaking in its first stages, before any settled routine has been established; nevertheless, he still found time to continue his statistical work and wrote several papers.

His death at the comparatively early age of 61 was not only a heavy blow to his family and friends, but a great loss to statistics, as his mind retained its full vigour, and he would undoubtedly have continued to work for many more years.

I am painfully conscious of the inadequacy of this sketch, which cannot hope to convey more than a faint impression of his unique personal quality to those who did not know him, but it will have served its purpose if it helps any readers to grasp the essential unity and directness of the personality which lay behind such widely varied manifestations.

(2) "STUDENT" AS STATISTICIAN

By E. S. PEARSON

For many years after the publication of his first paper in *Biometrika*, in 1907, the name of "Student" was associated in statistical circles with an atmosphere of romance. Those who knew him only through his written contributions must often have wondered who was this unassuming man, content to remain anonymous, who wrote so clearly and simply on so wide a range of fundamental topics. To those of us who came into touch with him personally, the knowledge that "Student" was W. S. Gosset did not altogether dispel that romantic impression. Here, in London, he would pay us visits from time to time at the old Biometric Laboratory on his way to Euston station to catch the Irish mail;

he would be wearing the grey flannel trousers that were a tradition of his Wykehamist schooldays and carrying a rucksack on his back. And then after a short hour's talk, perhaps on statistical subjects, perhaps on his garden experiments in cross-breeding, he would be off again to that wild Ireland where, in the "bad times", we had heard that gunmen were to be found hiding behind his hedges or even searching his house for arms. We had heard too of great exploits by members of his family of an entirely non-statistical character, of their boat-building and of their construction of a pair of water-skis which they used for walking over Kingstown Harbour.

My one short winter visit to Gosset's house at Blackrock, a few miles outside Dublin, would hardly by itself have cleared away this element of myth or made me appreciate fully the sterling values that lay beneath that friendly and unassuming exterior. We talked very little about statistics during my stay, and the strongest impressions remaining are of a morning spent among the immense vats and varied smells of the brewery; of drives out of town on misty evenings through the badly lit Dublin suburbs in that old, high two-seater Model-T Ford of his, christened "The Flying Bedstead"; of the warm hospitality of his fellow-brewers; and of a Saturday in the snow-covered Wicklow Mountains when, letting his folk go off to test the more exciting slopes, he patiently tried to teach me to ski on a short stretch of mountain road.

My real understanding of Gosset as a statistician began, as no doubt for many others, when I joined that wide circle of his scientific correspondents. Perhaps to the majority of these he has stood as the friend who, with a greater mathematical knowledge, helped them to understand the statistical approach to experimental problems. In my own case the position was a little different, as his endeavour was always to temper my mathematical reasoning with sane common sense. I can think of no other statistician who would have shown that interest and forbearance over many years to a young man who was continually posting to him the results of half-finished investigations for comment and criticism. In looking back through this correspondence I realize more clearly now than I could ever have done at the time what its value to me has been, and I can see how many of his ideas scattered through these letters have since almost unconsciously become part of my own outlook. I think this must be true also in the case of other persons with whom he corresponded, so that one can say that the last thirty years' progress in the theory and practice of mathematical statistics owes far more to "Student" than could be realized by a mere study of his published papers.

One of the striking characteristics of these papers, also of course evident in correspondence, was the simplicity of the statistical technique he used. The mean, the standard deviation and the correlation coefficient were his chief tools; hardly adequate for treating specialized problems it might be thought; yet how extremely effective in fact in his skilled hands! There is one very simple and illuminating theme which will be found to run as a keynote through much of his work, and may be expressed in the two formulae:*

$$\sigma^2_{x+y} = \sigma^2_x + \sigma^2_y + 2\rho\sigma_x\sigma_y, \qquad \ldots\ldots(1)$$

$$\sigma^2_{x-y} = \sigma^2_x + \sigma^2_y - 2\rho\sigma_x\sigma_y. \qquad \ldots\ldots(2)$$

Perhaps we may count as one of his big achievements the demonstration in many fields of the meaning of that short equation (2); as he wrote in 1923 (**11**, p. 273, but with modified notation):

> The art of designing all experiments lies even more in arranging matters so that ρ is as large as possible than in reducing σ^2_x and σ^2_y.

It is a simple idea, certainly, but I cannot doubt that its emphasis and amplification helped to open the way to all the modern developments of analysis of variance, and there may be some who have felt that where this technique runs a risk of defeating its ends by over-elaboration is just where that simple maxim has been set on one side. Recently I came across a short passage in a letter to a friend in Australia which refers to this theme and illustrates Gosset's own humorously modest outlook on his own contributions. He had just received a good deal of criticism of a paper he read in March 1936 before the Industrial and Agricultural Research Section of the Royal Statistical Society (**21**), particularly because of his advocacy of the half-drill strip method of agricultural experiment. This is essentially a method of comparison whose efficiency depends on maximizing correlation by taking the difference between the yields of neighbouring strips of the two varieties or treatments compared. He wrote:

> Meanwhile I...enclose the rough proof of what I said at the Statistical. You will gather from that that I am not in the fashion....Some years ago an American referred to difference treatment as "Student's" method and, though at the time I referred it to Noah, I am beginning to think that there is something in the name.†

Another point which must be borne in mind in gaining a real understanding of Gosset's character and outlook is that all his most important statistical work was undertaken in order to throw light on problems which arose in the analysis of data connected in some way with the brewery. The subject of statistics was in no sense a whole-time job for him, nor, on the other hand, was it his hobby as it might perhaps be described in the case of W. F. Sheppard; he undertook theoretical investigations only when he or his colleagues were faced with difficulties which needed solution along statistical lines. Rapid if less accurate methods appealed to him because in much heavy routine work it was a question of finding such methods or of making no attempt at statistical treatment. He was in no hurry to see his results in print, and several of his papers in *Biometrika* were written in response to an editorial request rather than on his own initiative. In two cases at least, which I shall refer to below, he was using methods in the brewery ten years before publication was undertaken. He was indeed the ideal

* σ_x, σ_y, σ_{x+y} and σ_{x-y} are the standard deviations of x, of y, of $x+y$ and of $x-y$ respectively, and ρ is the coefficient of correlation between x and y.

† See (**14**, p. 709).

servant of his firm, and part of the value of his life's work would need to be recorded in a history of progress gained by scientific research in industry rather than in the pages of *Biometrika*.

Yet in spite of the fact that only a small part of his time was taken up with statistics, Gosset had a wonderful power of "getting there first" before the more professional statisticians. Perhaps it was because his greater detachment meant a continual freshness of mind. It is this characteristic, as well as those others I have mentioned, that I shall try to bring out in my description of his work in the following pages.

EARLY STATISTICAL INVESTIGATIONS

Gosset became one of the brewers of Messrs Arthur Guinness Son and Co., Ltd., in 1899. The firm had shortly before initiated the policy of appointing to their staff scientists trained either at Oxford or Cambridge, and these young men found before them an almost unexplored field lying open to investigation. A great mass of data was available or could easily be collected which would throw light on the relations, hitherto undetermined or only guessed at in an empirical way, between the quality of the raw materials of beer, such as barley and hops, the conditions of production and the quality of the finished article. With keen minds playing round the interpretation of these data it was almost inevitable that before long the need was realized of some understanding of the theory of errors. No doubt during the first few years of his appointment Gosset was mainly occupied with learning the routine work of his job, but once this knowledge had been gained it was natural that he, as the most mathematical of the younger brewers, should give his attention to the question of error theory. He seems to have made use of the following books: G. B. Airy, *Theory of Errors of Observations*; Lupton, *Notes on Observations*; M. Merriman, *The Method of Least Squares*.

By 1904 he had made himself sufficiently familiar with the subject to draw up a *Report* on "The Application of the 'Law of Error' to the work of the Brewery". This document, dated 3 November 1904,* opens with some paragraphs which set out in simple terms a case for the introduction of statistical method in large-scale industry. They are worth quoting since they might be put before many a board of directors to-day with just as much cogency as they were put 34 years ago in Dublin:

> The following report has been made in response to an increasing necessity to set an exact value on the results of our experiments, many of which lead to results which are probable but not certain. It is hoped that what follows may do something to help us in estimating the Degree of Probability of many of our results, and enable us to form a judgment of the number and nature of the fresh experiments necessary to establish or disprove various hypotheses which we are now entertaining.

* I am extremely grateful to the firm for giving me permission to see and quote from this and other records available in their Dublin brewery.

When a quantity is measured with all possible precision many times in succession, the figures expressing the results do not absolutely agree, and even when the average of results, which differ but little, is taken, we have no means of knowing that we have obtained an actually true result, and the limits of our powers are that we can place greater odds in our favour that the results obtained do not differ more than a certain amount from the truth.

Results are only valuable when the amount by which they probably differ from the truth is so small as to be insignificant for the purposes of the experiment. What the odds should be depends:

(1) On the degree of accuracy which the nature of the experiment allows, and

(2) On the importance of the issues at stake.

It may seem strange that reasoning of this nature has not been more widely made use of, but this is due:

(1) To the popular dread of mathematical reasoning.

(2) To the fact that most methods employed in a Laboratory are capable of such refinement that the results are well within the accuracy required.

Unfortunately, when working on the large scale, the interests are so great that more accuracy is required, and, in our particular case, the methods are not always capable of refinement. Hence the necessity of taking a number of inexact determinations and of calculating probabilities.

The *Report* then introduces the error curve and discusses some of its properties. The curve is written in Airy's form

$$y = \frac{1}{c\sqrt{\pi}} e^{-x^2/c^2}, \qquad \ldots\ldots(3)$$

where c is the modulus. The method is given for estimating c from a sample of n observations, by calculating (a) the mean deviation, (b) the mean square deviation (dividing by $n-1$), and using the appropriate correcting factors. It is stated that (b) gives a better value "in proportion 114/100".* A numerical example is given and it is suggested that both methods (a) and (b) should be used to check one another. There is next some discussion given to what was then clearly a most important practical problem in the brewery: the size of sample needed to make the odds that the mean lay within desired limits sufficiently large. Chauvenet's criterion for the rejection of extreme observations is quoted, as well as the modulus of the estimate of c (obtained by the mean square process), namely $c/\sqrt{(2n)}$.

All this is simply Airy or Merriman put by Gosset into the form most useful for his fellow brewers. What, however, shows a flash of his own insight is the use which he makes of Airy's theorems on the "Error of the result of the addition (or subtraction) of fallible measures". Thus if

$$W = X \pm Y \pm Z \pm \text{ etc.}, \qquad \ldots\ldots(4)$$

* This is the ratio of the sampling variances of (a) the mean deviation, and (b) root mean square deviation estimates of c, in large samples. I do not know from what source Gosset obtained these figures. The full value of the standard error of the mean deviation for samples of any size from a normal population was first derived, I believe, by Helmert (1876), but Gosset could not have known of this paper.

and E, e, f, g, etc., are the probable errors (or alternatively the moduli or the mean errors) of W, X, Y, Z, ... respectively, then Airy gives the law

$$E^2 = e^2 + f^2 + g^2 + \ldots. \qquad \ldots\ldots(5)$$

Gosset had noticed in certain cases he had met with that the result $E^2 = e^2 + f^2$ did not hold, as it should according to this law, for both $W = X + Y$ and $W = X - Y$. In other words he found that if W, X and Y are measured from their means there was very considerable difference between *Sum* $(X+Y)^2$ and *Sum* $(X-Y)^2$. He concluded that when this was the case it was a sign of the existence of a correlation between the variables. Thus he was feeling his way towards the fundamental relations (1) and (2) of p. 212 above, but he had not yet been introduced to the correlation coefficient.

The concluding remarks of the *Report* are interesting:

We may point out that, although the proof of the law (of Error) rests on higher mathematics, the application of it only demands quite simple algebra. We have been met with the difficulty that none of our books mention 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.

This last difficulty was repeated in the summary which contains the sentence:

Explains that we have no information of the degree of probability to be accepted as proving various propositions, and suggests referring this question to a mathematician.

It is curious perhaps that Gosset should have felt at first that a mathematician was needed to solve this particular problem, which is just the point which the mathematician would now consider that the practical man must answer.* As we shall see in a moment he changed his view, but it seems to have been uncertainty on this question which led almost at once to that important contact between Gosset and Karl Pearson. A minute of March 1905 added to the printed *Report* indicates that arrangements for this meeting are to be made.

The interview was arranged through Vernon Harcourt, a chemistry don at Oxford whose pupil Gosset may have been and who perhaps got into touch with Pearson through Weldon, who was then Professor of Comparative Anatomy at Oxford. The opportunity for a meeting came about 12 July 1905 when Pearson was spending his long vacation at East Ilsley in Berkshire and Gosset bicycled over from his father's house at Watlington, preceded by a list of questions from which the following paragraphs are taken:

(1) *My original question and its modified form.* When I first reported on the subject, I thought that perhaps there might be some degree of probability which is conventionally treated as sufficient in such work as ours and I advised that some outside authority should be consulted as to what certainty is required to aim at in large scale work. However it would appear that in such work as ours the degree of certainty to be aimed at must depend

* I have, however, heard of another very recent case where an industrialist considered that it was the mathematical statistician's job to suggest the appropriate odds to use.

on the pecuniary advantage to be gained by following the result of the experiment, compared with the increased cost of the new method, if any, and the cost of each experiment. This is one of the points on which I should like advice.

(2) *Another problem.* I find out the P.E. of a certain laboratory analysis from n analyses of the same sample. This gives me a value of the P.E. which itself has a P.E. of P.E./$\sqrt{2n}$. I now have another sample analysed and wish to assign limits within which it is a given probability that the truth must lie. E.g. if n were infinite, I could say "it is 10 : 1 that the truth lies within 2·6 of the result of the analysis". As however n is finite and in some cases not very large, it is clear that I must enlarge my limits, but I do not know by how much.

(3) *What is the right way to establish a relationship between sets of observations?* I use the following method when endeavouring to establish a relationship between sets of observations, but I have reason to suppose that it is not a good way and would like criticism on my method and advice as to the proper way. Suppose observations A and B taken daily of two phenomena which are supposed to be connected. Let A_1, A_2, A_3, etc. be the daily A observations and let B_1, B_2, B_3, etc. be the daily B observations. (I reduce the B observations if necessary or increase them by multiplying by a constant so that the P.E. of the A and B is about the same.) Then I form two series A_1+B_1, A_2+B_2, etc. and A_1-B_1, A_2-B_2, etc. and find the P.E. of each of the new series. If they are markedly different, it is clear (sufficient observations being taken) that the original series A and B are connected and proceed to attempt to find it quantitively. I cannot however at present find the P.E. of my results, nor can I be quite sure how great a difference between the P.E.'s of the sum and difference series is necessary to shew the connection.

(4) *What books would be useful?* When you talk with me you will doubtless find out many other points on which I require enlightenment and could perhaps recommend me some books on the subject.

The solution of "*another problem*" was to be given $2\frac{1}{2}$ years later in Gosset's paper on "The probable error of a mean" (2). The method described in paragraph (3) is interesting. I do not know exactly how Gosset attempted to measure the relationship quantitively, but if, as would seem natural, he compared the difference between $\Sigma(A+B)^2$ and $\Sigma(A-B)^2$ with their average, then by adjusting the scale so as to make the P.E.'s of A and B approximately the same, he had secured a maximum value for this ratio, and therefore presumably minimized the risk of overlooking a relationship. For

$$\frac{\Sigma(A+B)^2-\Sigma(A-B)^2}{\frac{1}{2}\{\Sigma(A+B)^2+\Sigma(A-B)^2\}}=\frac{4r_{AB}\,\sigma_A\,\sigma_B}{\sigma_A^2+\sigma_B^2},$$

which, for a given value of r_{AB}, has a maximum value of $2r_{AB}$ when $\sigma_A=\sigma_B$. One feels that, given a little more time, with his unerring instinct for reaching the best solution, Gosset would have found for himself Galton's correlation coefficient, just as he was later to rediscover Poisson's limit to the binomial and Helmert's distribution of a squared standard deviation.

Among Pearson's rough jottings written down for Gosset at the interview is the basic formula that he needed,

$$\sigma_{A\pm B}^2=\sigma_A^2+\sigma_B^2\pm 2r\sigma_A\sigma_B$$

(with the letter r doubly underlined), the probable error formula for r and also references to a number of papers on the theory of statistics.

Gosset was a quick learner; the immediate results of this visit include a *Supplement* to the brewery *Report* of 1904, from which I have quoted, and a second *Report* on correlation dated 30 August 1905. In both of these the influence of new ideas received from Pearson is evident. The *Supplement* contains a warning that distributions may not always be normal, although in small sample problems "it is practically convenient to use a curve... which has been thoroughly investigated, of which the values have been tabled, and which in the majority of cases describes them 'within the error of random sampling'". His colleagues are also advised to use the standard deviation and not the mean error. The *Report* is headed "The Pearson Co-efficient of Correlation", and describes, with a numerical example, the method of calculating this coefficient, r, as well as the use of the regression straight line for prediction.

This idea of correlation, which in origin is of course Galton's rather than Pearson's, has more than once during the past fifty years brought with it a stimulus leading to fresh discovery. The conception, presented with all its novelty to minds which had hitherto only considered the perfect relationship of the physicist as a relationship which could be scientifically handled, has seemed to provide a key to the solution of a host of problems. The inspiration which Galton's discussion of correlation in his *Natural Inheritance* gave to Weldon and Pearson in the early nineties has often been referred to and, now, the introduction of the new ideas opened out fresh avenues of research to both Gosset and his colleagues. The crude method which Gosset had invented of examining the difference between $\Sigma(A+B)^2$ and $\Sigma(A-B)^2$ could be abandoned. It became possible to assess with precision the relative importance of the many factors influencing quality at different stages in the complicated process of brewing, and before long the methods of partial and multiple correlation were mastered and applied.* The *Reports* circulated within the brewery constantly quote correlation coefficients and their probable errors, while Gosset's rough notebooks of this date contain numerous correlation tables. Apart from the actual calculation of r, the idea of arranging data in a two-way table was possibly novel and certainly illuminating to the brewers.

It seems, however, to have been at once obvious to Gosset that the methods developed by Pearson and his co-workers for handling the large samples met with in biometric inquiries would probably need modification when applied to the problems of the brewery. In his *Report* on correlation of August 1905 he notes that "correlation coefficients are usually calculated from large numbers of cases, in fact I have only found one paper in *Biometrika* of which the cases are as few in number as those at which I have been working lately". He was dealing at this time with all the possible correlations between a number of characters for which 31 observations were available; in another problem only 10 observations

* A *Report* of Gosset's of June 1907 applies multiple correlation to prediction. The mathematical Appendix dated 27 September 1906 is stated to have been read through by Karl Pearson.

could be used. He gives a reason which, though faulty, is extremely interesting, for doubting the validity of the probable error formula for r in small samples. Thus, if r is an observed correlation from a sample of n individuals, he takes the ratio

$$\frac{\text{Deviation of } r \text{ from zero}}{\text{Probable error of } r} = \frac{r}{0{\cdot}6745\,(1-r^2)/\sqrt{n}} \qquad \text{......(6)}$$

as a measure of the significance of the correlation, remarking that if the ratio is greater than $2\frac{1}{2}$ the odds are about 20 : 1 on the existence of a real relationship. He then says that if n be very small "I expect a larger ratio is required", and illustrates this by supposing that $r = 0{\cdot}9$, $n = 4$, when the probable error calculated as in (6) becomes 0·064 and the ratio is 14. "Yet", he remarks, "no one would claim any certainty from four experiments."

If we are asking whether an observed r is consistent with sampling from a population in which the correlation, say ρ, is zero, then the appropriate probable error is approximately $0{\cdot}6745/\sqrt{n}$ and not the value used in (6). Thus in Gosset's example the ratio is really 2·7 and not 14; as he was afterwards to show, it was not the standard error that was seriously at fault in testing significance when dealing with small samples, but the assumption of normality. For $n = 4, \rho = 0$, the distribution of r is rectangular. The faulty reasoning involved in the interpretation of equation (6) has been used again and again in statistical literature; the reason that in 1905 the difficulty had not caught the attention of the workers at the Biometric Laboratory was that they were dealing with large samples and, for these, the error involved is of relatively small consequence. It was Gosset, "naughtily" playing about with absurdly small numbers,* who stumbled on the inconsistency, although not at first understanding its reason. Here perhaps we may see the first illustration of the tremendous gain in clear thinking that has followed in statistics from an approach to the subject from the small-sample end. Also this is one of the many occasions on which Gosset was first on the spot.

There were other difficulties in application that he was already turning over in his mind. For instance, he wished to obtain a combined measure of the correlation between two characters measured on several varieties of the barley used for malting and he considered the possibility of taking deviations from variety means. "I hope to find out the limitations of this device at some later date", he reported. "I am using it and similar devices pretty freely.... "

A point which may be of interest to industrial statisticians to-day is that the practical brewer of thirty years ago, as the practical engineer to-day, was objecting to the introduction into his reports of the statistician's term *population*, yet was unable to suggest an appropriate substitute. A footnote to the word population ran as follows: "This appears to be a general statistical term to

* Writing to Gosset on 17 September 1912 on the subject of the standard deviation, not correlation, Pearson remarked that it made little difference whether the sum of squares was divided by n or $(n-1)$, "because only naughty brewers take n so small that the difference is not of the order of the probable error!"

express a number of things or people of the same kind. We have tried to find a word in common use to express this, but have failed."

The *Report* closes with a characteristic piece of sound advice:

It must be borne in mind, however, that the better the instrument the greater the danger of using it unintelligently: it is more important than ever to think carefully in what way any connection may have arisen accidentally, and, more especially, any semi-constant variation must be treated with particular care.

Statistical examination in each case may help much, but no statistical methods will ever replace thought as a way of avoiding pitfalls, though they may help us to bridge them.

THE YEAR IN LONDON, 1906–7, AND THE WORK ON SMALL SAMPLES

Following a general practice of the brewery, Gosset was sent away from Dublin for a year's specialized study. He spent the greater part of this time either working at or in close contact with the Biometric Laboratory, where he arrived at the end of September 1906. During the year which had elapsed since he first met Karl Pearson he must have given a great deal of time and thought to the application of current statistical methods to the type of experimental and routine data analysed in the brewery. He was now anxious to obtain Pearson's opinion on the work he had already done and to ask his advice on a number of unsolved questions. Probably he had already realized that the most important problem on which he required further information was the behaviour of frequency constants in small samples. In a letter written to a friend at the brewery on 30 September 1906, just after his arrival, he outlines, however, only a modest programme:

Then he [K. P.] proposes to give me a room to work in, that I should attend his lectures, and become as far as possible accustomed to the calculations, etc., of his department. I had a long talk with him, and told him the lines I had been going on in the Hops..., and he seemed to consider that I had been over most of the ground, but points soon cropped up which showed him the necessity for going deeper. I think that from what he said I am more or less on the right lines so far; perhaps when the reports have been considered you might let me have a copy of each of them, to ask about anything which may have occurred to me by then about them. I think he would be very willing to give us advice on any points which crop up.

The first problem which he took up was of considerable practical importance in one department of the brewery activities: the question of the sampling error involved in counting yeast cells with a haemacytometer. In his paper (1) published early in 1907 he derived afresh Poisson's limit to the binomial distribution, namely,

$$e^{-m}\left\{1+m+\frac{m^2}{2!}+\ldots+\frac{m^r}{r!}+\ldots\right\}, \qquad \ldots\ldots(7)$$

and showed by a comparison of the series with four sets of experimental results that it did represent well the observed distribution of cell counts in an investigation carried out under carefully controlled conditions. The paper should be read in conjunction with another that he wrote on the same subject twelve years later (9).

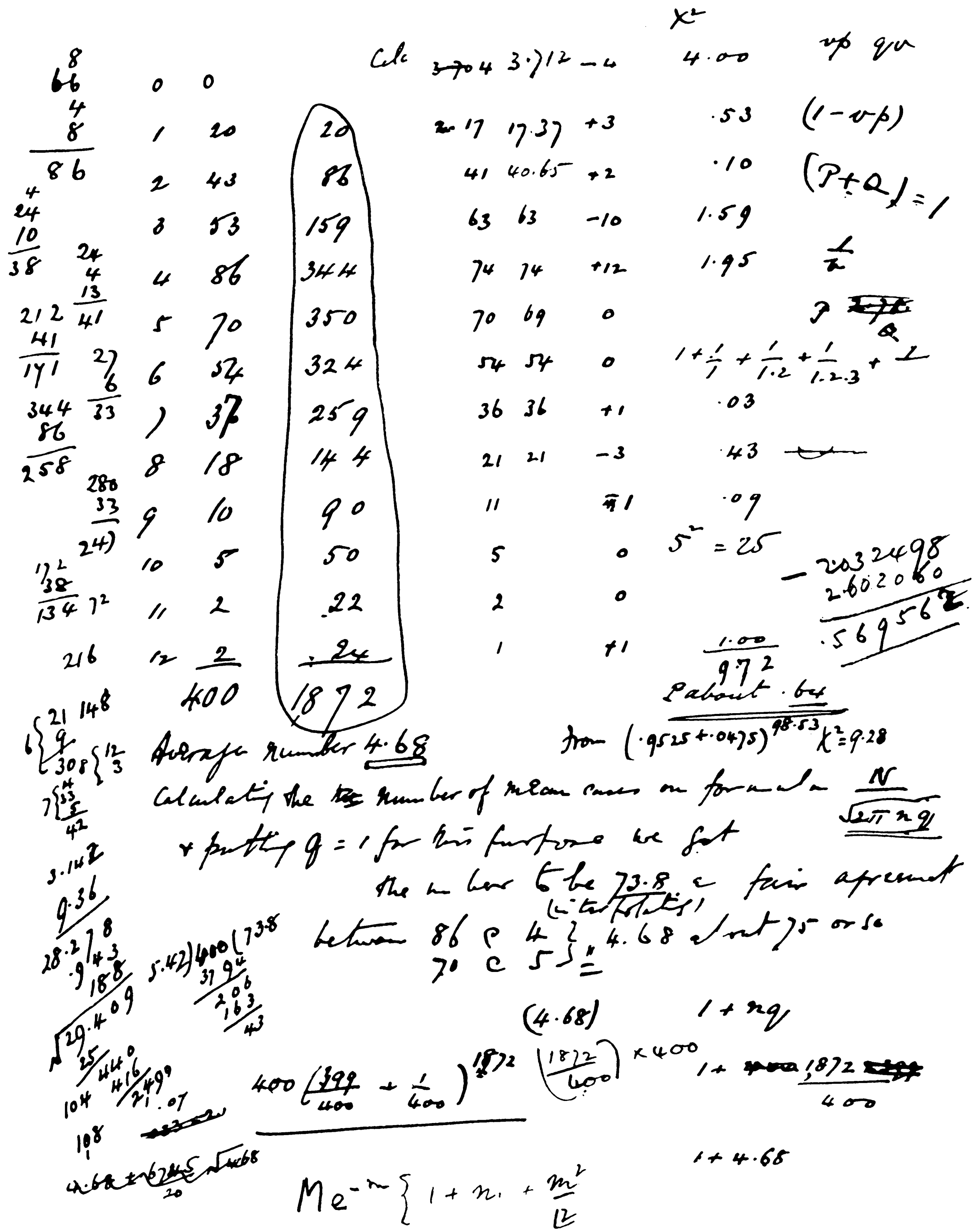
The derivation of the limiting form of the binomial was not in itself an achievement of any special difficulty; the series has been obtained independently from time to time by a number of investigators. But it was characteristic of "Student's" flair or, as he himself would have said, luck that when he had a practical problem to solve he should go straight to the correct solution; and that because it was a fundamental type of biological problem his research should have been of much greater value in the field of applied statistics than von Bortkiewicz's work, illustrated by fitting the Poisson series to suicides of German women and deaths of Prussian soldiers from the kicks of a horse.

I have reproduced in facsimile in Plate II two pages from Gosset's notebook containing the rough working for this paper. The experimental data are those of the series IV (see his p. 357). They are quoted also as an example of a Poisson distribution by R. A. Fisher in *Statistical Methods for Research Workers* (1938, p. 58). The left-hand page contains the 400 individual yeast cell counts and the resulting frequency distribution and histogram; the right-hand page shows the calculation of the mean, m, as well as the theoretical series and the derivation of χ^2. The expression $N/\sqrt{(2\pi mq)}$ (or $N/\sqrt{(2\pi nq)}$), where q is put equal to unity, is an approximation to the frequency in the group containing the mean. In the notes, Gosset seems to have reached this result by a rather lengthy method, but it can be obtained by putting $r=m$ in the general term of series (7) and using the first order term in Stirling's approximation to $m!$ No reference to this comparison was made in the published paper. A few figures, which are in pencil in the note-book, appear to be in Pearson's hand; e.g. the theoretical frequencies 3·712, 17·37 and 40·65 as well as the three terms of the Poisson series at the bottom of the page. They were jottings made no doubt by K. P. on one of his daily "rounds" of the laboratory.

A good part of the work on Gosset's second paper on "The probable error of a mean" (2) was also carried out during his year in England; with it is closely associated his third paper on the "Probable error of a correlation coefficient" (3), as both were supported by the same piece of experimental sampling. I have already referred to Gosset's doubts regarding the distribution of r in small samples; since in the brewery work a mean value had often to be estimated from eight or ten determinations he also felt uneasy about the applicability to such work of accepted theory regarding the distribution of the mean and the standard deviation. A letter written on 12 May 1907 to a colleague in Dublin shows him to be in the middle of his investigation. After dealing with some points about the significance of differences* he adds:

> Herewith my answer to your questions. I hope it is quite clear, but I am afraid I rather increase the difficulties when I try to explain anything as a rule.

* There is a reference to that long-standing difference of opinion regarding n and $n-1$, in the following sentence: "When you only have quite small numbers, I think the formula we used to use for the P.E. $(\sqrt{\{\Sigma(x^2)/(n-1)\}}\times 0{\cdot}6745)$ is better, but if n be greater than 10 the difference is too small to be worth taking the extra trouble." Here K. P. and Airy were in disagreement.



						χ^2	
0	0		Calc ~~3.704~~	3.712	−4	4.00	vp qa
1	20	20	17	17.37	+3	.53	$(1-vp)$
2	43	86	41	40.65	+2	.10	$(P+Q)=1$
3	53	159	63	63	−10	1.59	
4	86	344	74	74	+12	1.95	
5	70	350	70	69	0		
6	54	324	54	54	0		$1+\frac{1}{1}+\frac{1}{1.2}+\frac{1}{1.2.3}+$
7	37	259	36	36	+1	.03	
8	18	144	21	21	−3	.43	
9	10	90	11		1	.09	
10	5	50	5		0	$5^2=25$	
11	2	22	2		0		
12	2	24	1		+1	1.00	
	400	1872				9.72	

P about ·64

from $(\cdot 9525+\cdot 0475)^{98.53}$ $\chi^2=9.28$

Average number 4.68

Calculating the number of mean cases on formula $\frac{N}{\sqrt{2\pi nq}}$ & putting $q=1$ for this purpose we get the number to be 73.8 (interpolated!) a fair agreement between 86 @ 4 } 70 @ 5 } & 4.68 about 75 or so

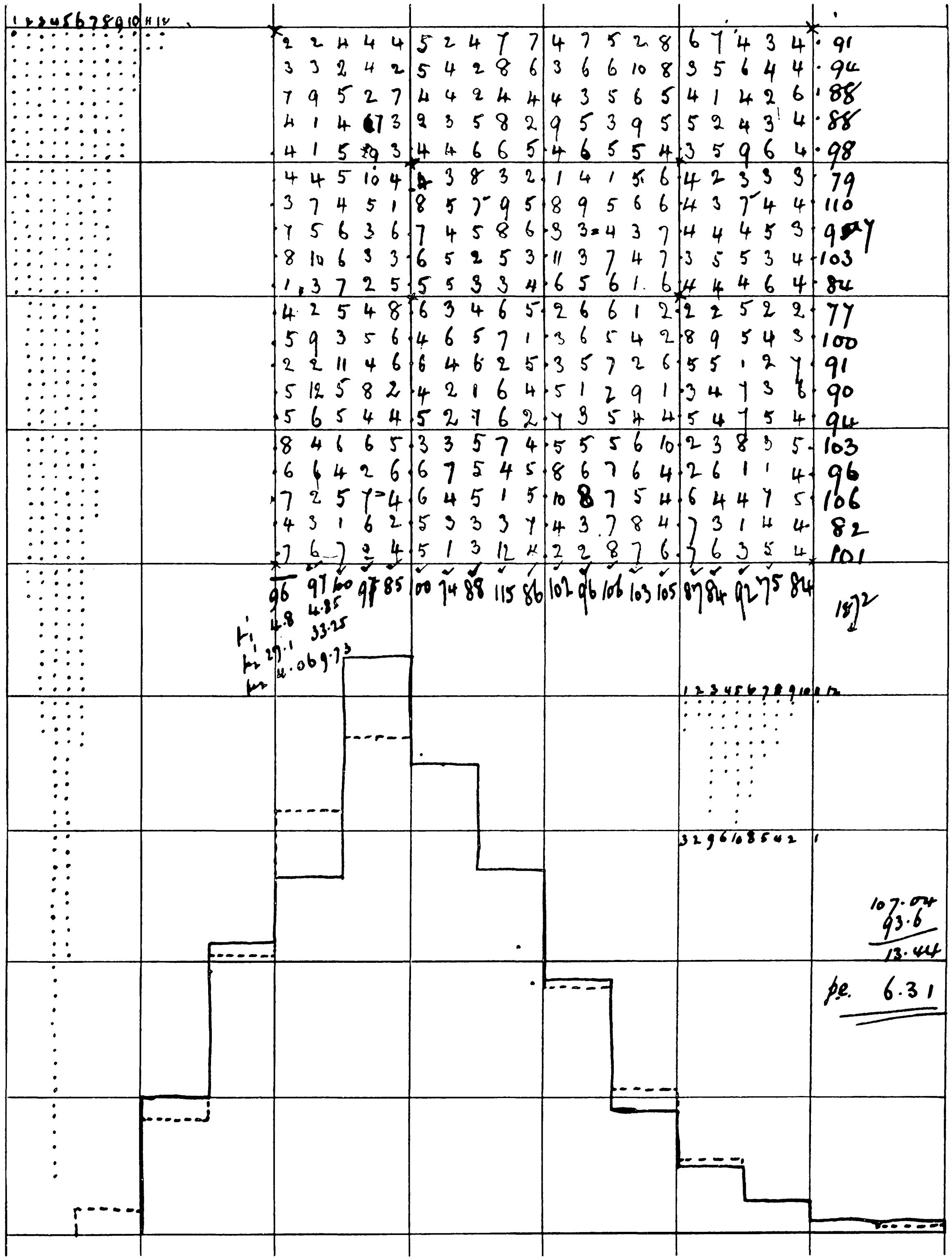
(4.68) $\quad 1+nq$

$400\left(\frac{399}{400}+\frac{1}{400}\right)^{1872}$ $\quad \left(\frac{1872}{400}\right)\times 400$ $\quad 1+\frac{1872}{400}$

$1+4.68$

$Me^{-m}\left\{1+m+\frac{m^2}{12}\right.$

Page from Gosset's notebook containing the analysis of haemacytometer counts. (Right-hand page.)



Page from Gosset's notebook containing the analysis of haemacytometer counts. (Left-hand page.)

What I have written on the back is true for large samples, and approximately so for small, and is the accepted theory. My work on small numbers may or may not modify it. We shall know later....

I go up to K.P.'s lectures from here [The Ousels, Tunbridge Wells] and on other days work at small numbers: a greater toil than I had expected, but I think absolutely necessary if the Brewery is to get all the possible benefit from statistical processes.

There could be no better illustration than these last sentences of the way in which Gosset's best work was called forth in the service of his firm.

The contents of the paper on the probable error of the mean are too well known to require more than a brief summary. Starting with a sample of n observations, $x_1, x_2, \ldots, x_n$, from a normal population with standard deviation σ and mean at the origin for x, Gosset obtained the sampling moments of $s^2 = \Sigma(x-\bar{x})^2/n$, where $\bar{x}$ is the sample mean. He showed that these moments corresponded exactly with those of a Pearson type III curve and hence inferred that the curve representing the sampling distribution of s^2 must almost certainly be

$$y = \text{constant} \times \sigma^{-n+1}(s^2)^{\frac{1}{2}(n-3)}\, e^{-ns^2/2\sigma^2}. \qquad \ldots\ldots(8)^*$$

He then showed that the correlation coefficient between $\bar{x}^2$ and s^2 was zero and, making the assumption (which does not necessarily follow though in fact it is true in this case) that this meant that $\bar{x}$ and s were absolutely independent, he deduced the probability distribution of $z = \bar{x}/s$ as

$$p(z) = \text{constant} \times (1+z^2)^{-\frac{1}{2}n}. \qquad \ldots\ldots(9)$$

He considered the properties of this curve,† gave a table of its probability integral for $n=4$ to 10 and examined its approach to a normal curve with standard deviation $1/\sqrt{(n-3)}$. He next compared the distributions (8) and (9) with the results of a sampling experiment for the case $n=4$ and finally illustrated the use of his results on four examples.

When two years ago the question of the photographic reissue of the paper had been suggested to meet a continued demand for offprints, Gosset wrote to me describing it as now "rather a museum piece". That is true, though perhaps in a different sense than he meant. It is a paper to which I think all research students in statistics might well be directed, particularly before they attempt to put together their own first paper. The actual derivation of the distributions of s^2 and z, or of $t = \sqrt{(n-1)}\, z$ in to-day's terminology, has long since been made simpler and more precise; this analytical treatment need not be examined carefully, but there is something in the arrangement and execution of the paper which will always repay study.

In the first place, in the Introduction and Conclusions we find an excellent illustration of Gosset's wise advice given to a beginner in the art of composition: "First say what you are going to say, then say it and finally end by saying that

* That this result had previously been derived by Helmert (1876), English-speaking statisticians were quite unaware till many years later.

† There are some minor errors in §§ IV and V of the paper.

you have said it."* The main part of the paper, the "saying it", is divided clearly into headed sections. The adequacy of the assumptions on which the mathematical theory rests is tested by a piece of experimental sampling; this test being satisfactorily passed, computed tables required for application are given and finally a number of well-chosen examples illustrate the purpose of the inquiry.

Before considering some other notable features of the paper and attempting to assess its influence on later work, it is important to see just what was the main purpose of the inquiry that its author had in mind. As usual with him, this was simple and practical. Having n observations, he wished to know within what limits the mean of the sampled population—the "true result" of the 1904 *Report*—probably lay. His solution involved a tacit introduction of the method of inverse probability, but I do not think he ever tried to put this into precise terms.† Thus the last sentence on the first page of the paper runs as follows:

> 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.

The results of the present investigation meant to Gosset that he could now assume in small samples a z-distribution for the population mean about the sample mean, the scale now being the sample standard deviation, s. In his examples he uses the z tables, not to test the hypothesis that the population mean is zero or has some other specified value, but to find the odds that this mean lies within specified limits, e.g. between 0 and ∞, that is to say is positive. Take for instance his *Illustration* I (pp. 20–1); the average number of hours of sleep gained by ten patients treated with *D. hyoscyamine hydrobromide* is $\bar{x}=0{\cdot}75$ while the standard deviation is $s=1{\cdot}70$. If we regard the population mean, say ξ, to be distributed about the sample mean 0·75 in the z-form, with a standard deviation of s, it follows that the chance that $\xi>0$ is the proportionate area under the z-curve between the ordinate at

$$z=\frac{0-0{\cdot}75}{1{\cdot}70}=-0{\cdot}44$$

and ∞. This is the same as the chance that $z<+0{\cdot}44$, which interpolation in his tables in the column $n=10$ shows to be 0·887. He therefore argued that the odds are 0·887 to 0·113 that the population mean ξ is positive, i.e. that the soporific will

* The advice was not originally Gosset's. Writing in 1934 he says: "This is a rule which we owe to A. J. (I think at second hand)." He then quotes the rule and adds, "It does make things so much easier for everybody concerned, besides which 'what I tell you three times is true'"; the last words are those of the Bellman in *The Hunting of the Snark*.

† In his paper on the correlation coefficient written in the same year (**3**, p. 302) Gosset states definitely that a knowledge of the *a priori* probability distribution of the population correlation coefficient, R, is needed in order to determine "the probability that R...shall lie between any given limits".

on the average give an increase of sleep. While a somewhat loosely defined conception of inverse probability seems to underlie the argument, it will be seen that as far as the practical consequences go, Gosset had reached a result which we can hardly improve on 30 years later. It is true that, using the idea of the fiducial or confidence interval, some of us would word our statement of limits and probabilities a little differently so as to avoid any appeal to inverse probability, but as practical statisticians we must, I think, admit that our conclusions would be identical.

There are some other features of the paper which are interesting historically. Gosset remarks on p. 13 that before he succeeded in solving the problem analytically, he had endeavoured to do so empirically. The sampling experiment which he carried out for this purpose involved the drawing of 750 samples of 4 by means of shuffled slips of cardboard, from W. R. Macdonell's (1901) correlation table containing the distribution of height and middle-finger length of 3000 criminals. As far as I know this was the first instance in statistical research of the random sampling experiment which since has become a common and useful feature in a large number of investigations where precise analysis has failed. The results of this same experiment were used by Gosset in a number of later papers. On p. 16 he draws attention to a difficulty in the application of Pearson's χ^2-test of goodness of fit which was later to lead to R. A. Fisher's modification in terms of degrees of freedom. On p. 19 he gives reasons for believing that even when the population sampled is not normal the sampling distribution of z will be very little modified; this was a prediction which experimental and theoretical investigations carried out in recent years have confirmed.

Finally we may note the introduction of a difference in notation to distinguish between sample and population characters, viz. s for the sample and σ for the population standard deviation. The need for this distinction seems obvious to us to-day, but it is interesting to notice that it was only when attention was directed to the problem of small samples that statisticians grasped the clarification resulting from this innovation.

As the theory of mathematical statistics has developed, the significance of "Student's" test has been elaborated from many angles and deeper meanings associated with it than its author had ever dreamed of. This is a common feature of scientific progress, but as Neyman very appropriately remarked on a recent occasion (1937, p. 142): "The role of a rigorous scientific theory is frequently very modest and is reduced to explaining to the practical man—and this sometimes with a certain difficulty—how good is what he himself knew to be good long ago." To understand the reason for the historical importance that has rightly been associated with this paper, it is not however necessary to discuss the abstract conceptions of the mathematical statistician and their relation to forms of critical regions in hyperspace; it can be explained much more simply

15-2

than that. As Gosset wrote on the second page of the paper, referring to the inadequacy for certain purposes of the statistical technique available in 1908:

> There are other experiments, however, which cannot easily be repeated very often; in such cases it is sometimes necessary to judge of the certainty of the results from a very small sample, which itself affords the only indication of the variability. Some chemical, many biological, and most agricultural and large scale experiments belong to this class, which has hitherto been almost outside the range of statistical inquiry.

It is probably true to say that this investigation published in 1908 has done more than any other single paper to bring these subjects within the range of statistical inquiry; as it stands it has provided an essential tool for the practical worker, while on the theoretical side it has proved to contain the seed of new ideas which have since grown and multiplied an hundredfold.

The sampling experiment used to test the accuracy of the theoretical distributions of s^2 and z was also planned to throw light on the distribution of the correlation coefficient r, in very small samples. In this second problem (3) Gosset was forced to rely much more on his empirical approach than before, since the mathematical solution lay beyond his powers. In suggesting the probable form of the distribution of r when sampling from a population in which the two variables were uncorrelated (i.e. $R = 0$)* he could get no clue from known values of moments as in the case of s^2. He started from the following basis: (*a*) the distributions must be symmetrical about $r = 0$ and be limited within the range -1 to $+1$; (*b*) he had available the distributions of r found from his experiment for 745 samples of 4 and 750 samples of 8; (*c*) of these, he noticed that the former was approximately rectangular.

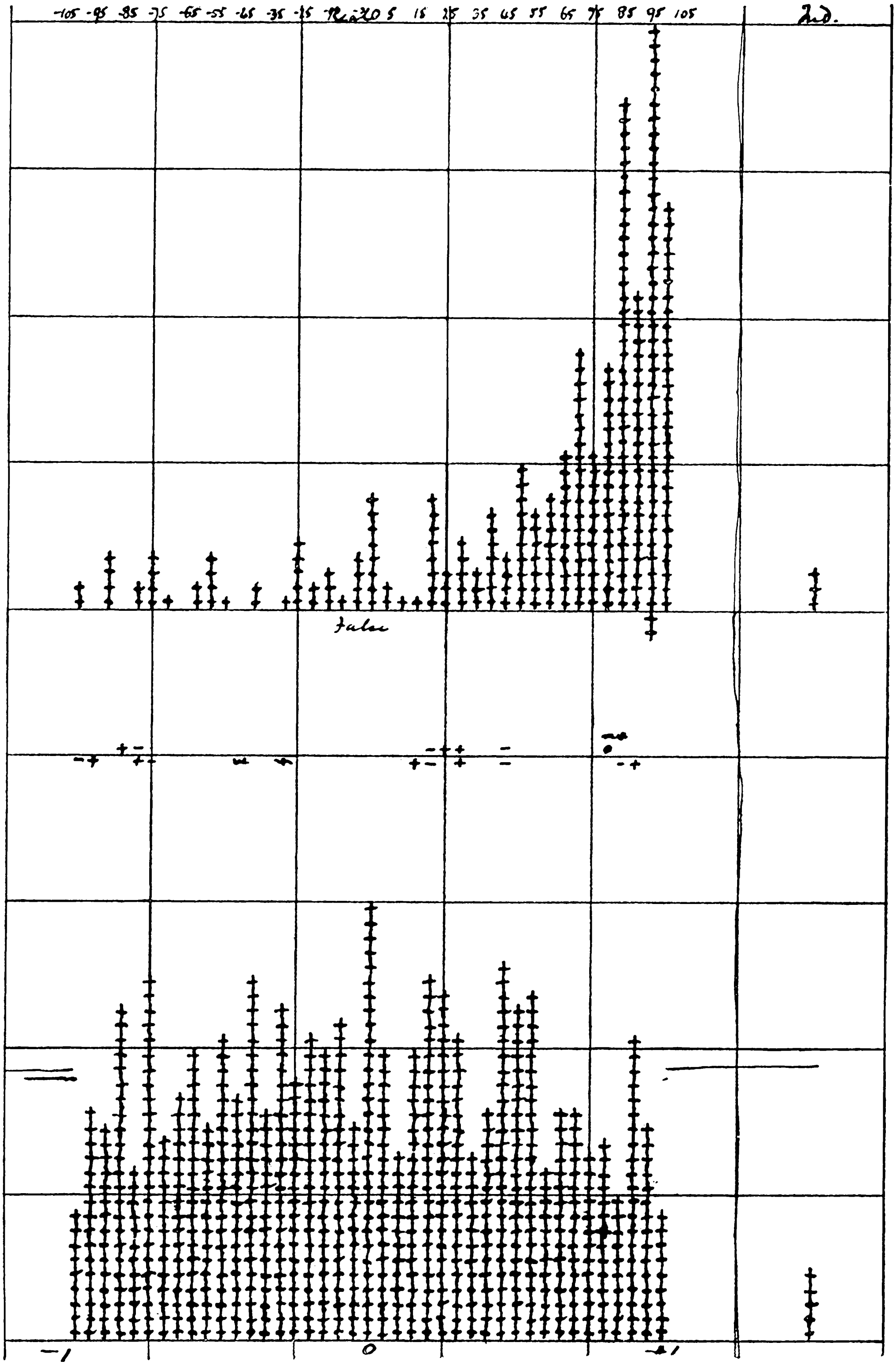
As in the case of s^2, his training at the Biometric Laboratory naturally suggested that he should try to use a Pearson curve for the unknown distribution; a type II curve was the only one suitable, and therefore in his own simply expressed phrase, "working from $y = y_0(1-x^2)^0$ for samples of 4 I guessed the formula

$$y = y_0(1-x^2)^{\frac{1}{2}(n-4)}\text{''}. \qquad \ldots\ldots(10)$$

He then showed that for $n = 8$ this formula represented his empirical sampling distribution very well, and pointed out that the result agreed with large sample theory, since the standard deviation $\sigma_r = 1/\sqrt{(n-1)}$ would equal Pearson and Filon's value of $(1-R^2)/\sqrt{n}$ when $R = 0$ and $n \to \infty$. He also gave the correct limiting result, which he had been able to establish for any R, when $n = 2$, suggesting that this might furnish a clue for the distribution when $n > 2$. It was a brilliant piece of guessing and all the more striking because of the forceful way in which the supporting evidence was marshalled.

In the case where the population correlation, R, was not zero Gosset provided three empirical sampling distributions for the cases $R = 0{\cdot}66$ and $n = 4$, 8 and 30

* He used R for the population correlation; the notation, ρ, seems to have been first used by H. E. Soper (1913).



Distribution of the correlation coefficient in samples of 4, tabled in Gosset's notebook.
Above, $R = 0{\cdot}66$; *below*, $R = 0$.

He also set out very clearly the conditions which his work showed must be satisfied by the true distribution. "I hope", he concluded, "they may serve as illustrations for the successful solver of the problem". Six years later R. A. Fisher was able to demonstrate the substantial accuracy of all Gosset's predictions both in the r and the z paper.

In the notebook containing the original samples of 4 from Macdonell's correlation distribution, there are given what I think must be the original distributions built up by Gosset as he tabled his calculated values of r. Two of these are shown in facsimile in Plate III ($n=4$, $R=0{\cdot}66$ and $n=4$, $R=0$). It is hard to believe that Gosset did not experience a very pleasurable excitement as these distributions gradually took shape on the paper, for he was exploring a region entirely unmapped and the discovery of the rectangular distribution in the case when $R=0$ must have been a complete surprise.*

One of the curious things that must strike us now about these two papers of Gosset's (2, 3) is the small influence that their publication had for a number of years on current statistical literature and practice. The z-test was used in the brewery at once, but I think very little elsewhere for probably a dozen years. Perhaps because he realized that it showed how little reliability could be placed on a correlation coefficient based on small numbers, Gosset does not seem to have recommended the use of the r-test even to his colleagues and he made no tables of the probability integral for the distribution (10). I have come across, however, one reference to the work in a letter of 3 April 1912 to E. S. Beaven, in which the following remarks occur:

> By the way, don't be *too* cock-a-hoop about your 0·95 correlation with 7 cases. Such a thing might occur more than once in a hundred trials of 7 cases, even if there were no correlation. (I haven't got tables to evaluate
>
> $$\int_{\sin^{-1}0{\cdot}95}^{\frac{1}{2}\pi}\cos^4\theta\,d\theta\Big/\int_0^{\frac{1}{2}\pi}\cos^4\theta\,d\theta,$$
>
> but you get that fraction of N at each end 0·95 or over in N trials); and I guess its about 2 % at each end.† All the same it seems very reasonable to suppose that it is right.

From Gosset's point of view, he had developed the tools which he needed for practical application in Dublin and he was not primarily interested in their wider use. If Pearson failed to realize the importance of the work and did not assimilate the results into current practice and teaching, it was because he too was mainly interested in what appeared to be of value in the research investigations of his laboratories. To him all small sample work was dangerous and should be avoided. But it would be wrong to suppose that there was a lack of sympathy between the two; except at a far later stage when opposite views over z found their way into print, Pearson's attitude towards Gosset's small sample

* Yet Mrs Gosset, who was helping him at the time, writes: "Whatever thrill he may have got out of that experiment he showed nothing whatever of it, and his amanuensis never realized that there was anything original about it!"

† Gosset was wrong here. The fraction is actually 0·001.

work was one of humorous protest, well conveyed in the quotation I have given about "naughty brewers" who take n too small (p. 218 above). The readiness with which he would talk to Gosset over his problems and at times refer to him on matters of difficulty shows how highly he rated his ability and insight Although Gosset launched off along independent lines of investigation directly he had mastered the elements of statistical theory, it is clear that he owed a great deal to the early guidance that he received in London. In the first place he had that very great advantage of being freed for a year from his official duties and of spending that time in close contact with persons who were enthusiasts in the study of statistics. Although, as he wrote at a later date, "I am bound to say that I did not learn very much from his [K. P.'s] lectures; I never did from anyone's and my mathematics were inadequate for the task", he obtained from the Biometric Laboratory a number of things which were not to be found in Airy or Merriman: the theory of correlation, the χ^2-test, and above all Pearson's system of frequency curves. It is doubtful for instance if he could have reached the distribution of s^2, and hence that of z, if he had not had available for use Pearson's type III curve.

After his year in London was over Gosset kept in close touch with Pearson for 29 years, and to his intimate friends would speak with admiration of his teacher. Some sentences which he spoke at the opening meeting of the Industrial and Agricultural Research Section of the Royal Statistical Society in November 1933 were composed, I know, with this aspect of the relationship between professor and student in mind:

> Another point arises from the peculiar nature of statistics. It is impossible to apply statistical methods to industry or anything else unless one has a certain amount of intelligent experience as a background. That works both ways. The practical man has to go and talk to his Professor partly in order that the Professor himself should share his experience. ...The whole art of statistical inference lies in the reconciliation of random mathematics with biassed samples. Every new problem has some fresh kind of bias and might contain some new pitfall. The only way not to fall into these pitfalls is to talk over the problem with some intelligent critic; and so the practical man, if he is not entirely foolish, talks over his problems with the Professor, and the Professor does not consider himself to be a competent critic unless he has had some experience of applying the statistics to industry and has learned the difficulties of that application.

Miscellaneous papers, 1909–21

Before considering the very important part that Gosset played in the development of agricultural experimentation, it is desirable to give a brief account of six papers on a variety of subjects which were published in *Biometrika* between 1909 and 1921.

(i) The first of these papers on "The distribution of the means of samples which are not drawn at random" (**4**, 1909) dealt with one aspect of that theme which, as I have already mentioned, runs through so much of his work. He had realized at an early date how frequently there existed a correlation between successive observations either in time or space. Thus if x and y are two contiguous observations it would follow that

$$\sigma^2_{x+y} > \sigma^2_x + \sigma^2_y > \sigma^2_{x-y}.$$

Hence if x and y were successive duplicate chemical analyses of the same quantity their mean would be less reliable than we should expect on the usual theory of random sampling. On the other hand were x and y the yields from plots of two different cereals which were to be compared, by placing the plots side by side in space, the difference $x-y$ would be more reliable than on the classical error theory. In this paper he considers the distribution of the mean not of two but of n observations, so selected that they are correlated, i.e. more like one another than individuals randomly selected from the population. It is the problem of fraternities which Pearson had termed homotyposis in his biometric work. Gosset gave the second, third and fourth moments of the sample mean, the second having the value

$$M_2 = \frac{\sigma^2}{n}\{1+(n-1)\rho\}, \qquad \ldots\ldots(11)$$

where σ is the population standard deviation of x and ρ the correlation between the x's in a sample, which Fisher has termed the intraclass correlation. From the values of the third and fourth moments he deduced that in general it was likely that the distribution of the mean would tend to normality less rapidly than when $\rho = 0$.

From the practical point of view he was concerned to warn the chemist that "repetition of analyses in a technical laboratory should never follow one another, but an interval of at least a day should occur between them. Otherwise a spurious accuracy will be obtained which greatly reduces the value of the analyses".

(ii) The next paper (**6**) published in 1913 dealt with "The correction to be made to the correlation ratio for grouping", an investigation no doubt connected with Pearson's work (1913) on the same subject published in the same number of *Biometrika*.

(iii) Volume x of *Biometrika* (1914) contains a short note on "The elimination of spurious correlation due to position in time or space" (**7**). In this, Gosset showed that the difference correlation method used by F. E. Cave (1904) and R. H. Hooker (1905) could be extended to differences of higher order than the first. This paper was the basis of later investigations on the variate difference correlation method.

(iv) In 1917 (**8**) Gosset published an extension of his tables of the probability integral of z; the range covered now ran from $n=2$ to $n=30$. In the introductory remarks he again gave advice "as to the best way of judging the accuracy of physical or chemical determinations". He wrote:

After considerable experience, I have not encountered any determination which is not influenced by the date on which it is made; from this it follows that a number of determinations of the same thing made on the same day are likely to lie more closely together than if the repetitions had been made on different days. It also follows that if the probable error is calculated from a number of observations made close together in point of time much of the secular error will be left out and for general use the probable error will be too small. Where then the materials are sufficiently stable, it is well to run a number of determinations on the same material through any series of routine determinations which have to be made, spreading them over the whole period.

(v) Gosset's paper of 1919 (**9**) on "An explanation of deviations from Poisson's law in practice" answered some questions regarding the relation of this series to the positive and the negative binomial raised by Lucy Whittaker (1914) in a paper published five years earlier from the Biometric Laboratory. Since the rather severe criticisms of the latter paper directed against the applications of the Poisson law made by Bortkiewicz and Mortara might have discouraged its use in other directions, Gosset pointed out that the object of his own earlier paper (**1**) was to give the user of the haemacytometer a guide to the error of his count. From this first practical point of view it made little difference whether, theoretically, the better fitting distribution was a positive or negative binomial, although as a further point it was of interest to consider what such departures implied if the data were sufficient to establish them.

(vi) The final paper (**10**) of this group on "An experimental determination of the probable error of Dr Spearman's correlation coefficients", was written in the first instance for reading at one of the early meetings (13 December 1920) of the newly formed Society of Biometricians and Mathematical Statisticians. Gosset had many years before realized the value of the method of rank correlation in assessing quickly the order of relationship between two short series of numbers. Probably while working at the Biometric Laboratory he had developed the proof quoted by Pearson (1907, p. 13), that the standard error of the coefficient

$$\rho = 1 - \frac{6\Sigma D^2}{n(n^2-1)} \qquad \ldots\ldots(12)$$

is $1/\sqrt{(n-1)}$, in the case of independence in the population. In a Report written in 1911 for his colleagues in the brewery he illustrated the use of the method and gives what is substantially the correction for "ties" described in the present paper of 1921. Apart from the publication of this correction, the paper is of interest because Gosset again made use of his sampling experiment of 1907. For the 375 samples of 8 from a population having correlation 0·66 he calculated both of Spearman's rank correlation coefficients, in their raw and corrected form and, in the case of his 100 samples of 30 added Sheppard's estimate of correlation obtained from a median fourfold division. He uses these results to make a number of comparisons between the methods, in particular paying regard to the amount of additional sampling needed if one of these more rapid methods of "assay" is to give as reliable an estimate of the population correlation coefficient as that obtained from the usual product-moment formula. He concludes by suggesting to mathematicians a problem which has still remained unsolved, that of determining the sampling distribution of the rank coefficient of equation (12) above, in random samples from a bivariate normal population, in which the correlation is not zero.

THE APPLICATION OF STATISTICAL METHOD TO AGRICULTURAL PLOT EXPERIMENTS

It is a feature commonly noticeable in the advance along any new line of scientific inquiry that the first steps in that progress are made hesitatingly and with difficulty, accompanied by much trial and error; and then after many years of what seems, looking back, to have been a painfully slow advance to an obvious goal, a stage is reached where the way forward has been almost cleared so that the introduction, perhaps, of some new tool or some fresh personality leads to a rapid advance into fresh country. In later years the casual student may well attribute the beginning of an epoch to that moment of rapid advance, partly because few records of the earlier struggle have found their way into print and partly because the later workers themselves have hardly realized the amount of thought that has gone into the creation of ideas which have formed the groundwork of their own further progress.

The history of the introduction of statistical methods in the planning and interpretation of agricultural experiments provides an illustration of these points. The large extension of technique with the accompanying stimulus to scientific planning which followed R. A. Fisher's introduction of the methods of analysis of variance in the years following 1923, may have caused the present-day statistician to overlook the essential pioneer work of the preceding years, without which it is certain that the later advance would have been impossible.* It therefore seems appropriate to take this opportunity of giving rather special attention to this aspect of Gosset's contribution to statistics and to do so by following out the gradual stages by which he advanced from simple beginnings to the analysis of a balanced block experiment.

A number of persons contributed to this early work and, as is often the case when methods of attack are in an imperfect or trial stage, ideas were worked out in correspondence or by word of mouth rather than in print. The brewery, as a very large consumer of barley, was naturally interested in agricultural problems and in particular in certain large-scale experiments undertaken in Ireland under the supervision of the Irish Department of Agriculture. Gosset was not, however, concerned with giving advice in these experiments till a

* Fisher himself has on many occasions paid a warm tribute to the help he received both from "Student's" published work and from correspondence and discussion.

number of years after he had specialized in statistics, and I think his first real interest in agricultural work arose from his contact with E. S. Beaven, who as a maltster was from time to time in Dublin on official business. Beaven had started experimental work in the nineties and about 1905 approached Gosset for an interpretation of apparently anomalous results, afterwards seen to be due to interference, that he found in comparing the yields of two varieties of barley in his "cage" at Warminster. From that date until Gosset's death there was a continuous flow of correspondence between them in which ideas were exchanged and thrashed out, and the more mathematical approach of the younger man was influenced by the practical experience of his older friend.

It will be noticed that three out of Gosset's four illustrations in the paper on the probable error of the mean (2) deal with agricultural topics; the data were taken from published accounts of Woburn farming experiments and Gosset shows how, by taking appropriate differences and using his z-test, a more precise interpretation of such results could be obtained than had hitherto seemed possible. Beaven was in touch with the agricultural work both at Rothamsted and Cambridge and it was no doubt owing to his report of Gosset's keen interest in these problems that both of those classical papers by Wood & Stratton (1910) and by Mercer & Hall (1911), dealing with the analysis of what we now term uniformity trial data, passed through Gosset's hands before publication. The first was only "an affair of a day or two's glancing at" after which he "made one or two suggestions, most of which were quite rightly turned down as being too refined for the purpose".* But in the second case he "brooded over the paper for months", and made suggestions which were incorporated, as well as adding an Appendix (5). If we compare the two statistical contributions, that of Stratton to the first paper and that of Gosset to the second, it is possible, I think, to see without difficulty the latter's special contribution to the subject. Stratton is following the approach of the classical theory of errors, which he had learnt and applied as an astronomer; he shows that variation in plot yields can be represented by the error curve and hence that the results of that theory regarding the probable error of a mean are applicable. These results are used to show the relation of size and number of plots (or animals) to the reliability of the results. No reference is made to "Student's" paper of 1908.

Gosset, writing his Appendix a year later, brings to the problem the added insight that he has gained from an understanding of correlation theory and from much discussion of the Warminster results with Beaven. He shows how it is possible to bring the changing fertility level or "patchiness" of the experimental field into service (*a*) by scattering the varieties to be compared in small plots over the field, and (*b*) then taking as the statistical variable for analysis the difference between the characteristics of two varieties on neighbouring plots. Thus the standard error, by way of formula (2), p. 212 above, can be very much

* These quotations come from a letter of 4 June 1922 from Gosset to Beaven.

reduced. The illustration which he gives deals only with the case of two varieties A and B, and at this date he had probably not thought out a technique for dealing with more comparisons.

There is another point of difference that may perhaps be noted; Wood and Stratton by raising the question, "What is the probable error of a single field experiment?" seemed to suggest that it might be possible to determine a single value, σ, which it would be appropriate to apply to future experiments of a given type. Gosset however emphasized a rather different idea. He writes (**5**, p. 130):

> But, it will be asked, why take all this trouble? The error of comparing plots of any given size has been found by the authors of the paper, and all that has to be done is to apply this knowledge to the particular set of experiments.
>
> The answer to this is that there is no such thing as the absolute error of a given size of plot. We may find out the order of it, be sure perhaps that it is not likely to be less than (say) 5 per cent. nor more than 15 per cent.... but the error of a given size of plot must vary with all the external conditions as well as with the particular crops upon which the experiment is being conducted, *and it is far better to determine the error from the figures of the experiment itself; only so can proper confidence be placed in the result of the experiment.**

His own z-distribution was available, if the number of observations was scanty.

If the field were divided into m pairs of plots and x_i and y_i were the yield, say, of varieties A and B on contiguous ith plots, then Gosset's test for a difference in yield may be summarized as follows:

Write $d_i = x_i - y_i$ and $\bar{d} = \sum_i d_i/m$.

Calculate the ratio
$$z = \frac{\bar{d}}{\sqrt{\left\{\sum_i (d_i - \bar{d})^2/m\right\}}}, \qquad \text{......(13)}$$

and if $m \leqslant 10$ refer this to the z-tables (**2**, p. 19). Otherwise, if $m > 10$, since z has a standard deviation of $1/\sqrt{(m-3)}$, refer $z\sqrt{(m-3)}$ to Sheppard's tables of the normal probability integral.

In the years 1912 and 1913 at Beaven's suggestion plot experiments of similar design, each comparing eight varieties of barley, were carried out at three centres, viz. Warminster, Cambridge and Ballinacurra in Co. Cork. The experiments were carried out in cages, and there were twenty replications of each of the eight varieties in square-yard plots. The arrangement of the varieties in a "chess-board" pattern was effectively what we should now term balanced; a plan of one of the schemes has been shown in Gosset's paper of 1923 "On testing varieties of cereals" (**11**, p. 277) and I have reproduced a portion of this below, only adding some thicker rules to separate the different sets of eight plots.

Beaven suggested that the results might be analysed by using as a statistical

* The italics are mine.

variable the difference between (1) the yield on a plot of A, say, and (2) the mean yield for the eight varieties (including A) on the 9-plot area in which this A-plot lay at the centre.* This was a rough and ready procedure but, as Gosset pointed out, owing to correlation there would be difficulty in the statistical interpretation. The method which he preferred was a very natural extension of his difference method advocated in the case where there were only two varieties. He could still clearly use that method to compare any two of the eight varieties,

E 230·1	*B* 249·3	*G* 312·2	*D*	*A*	*F*	
D 255·9	*A* 222·6	*F* 218·7	*C*	*H*	*E*	
C 265·6	*H* 205·0	*E* 246·7	*B*	*G*	*D*	
B 265·9	*G* 236·7	*D* 295·8	*A*	*F*	*C*	*H*
A 236·5	*F* 210·4	*C* 291·1	*H* 223·9	*E*	*B*	*G*

Fig. 1.

say A and D, taking the corresponding pair of plots from each set of eight, and differencing the character measured, although the plots would not now be generally contiguous. This would mean that changes in soil fertility, etc. would make the comparison less accurate than before,† but that could not be helped if eight varieties were to be compared in a single experiment in place of two. He saw, however, that it was possible to compensate to some extent in another direction for this loss in accuracy, by getting a single combined estimate of error from all the $\frac{1}{2}n(n-1) = 28$ possible sets of differences between $n = 8$ varieties, a method which he described as "hotchpotching" the comparisons. The reasoning which he used in reaching his result may be set out as follows:

Let there be n varieties each repeated m times and denote by $d_{uv\cdot i}$ the difference obtained from the ith comparison of the uth and vth varieties ($i = 1, 2, \ldots, m$) and by $\bar{d}_{uv}$ the mean of these m differences. Thus in Fig. 1, if u and v stand for varieties A and D, respectively, then

$$d_{uv\cdot 1} = 236{\cdot}5 - 255{\cdot}9 = -19{\cdot}4, \quad d_{uv\cdot 2} = 222{\cdot}6 - 295{\cdot}8 = -73{\cdot}2, \text{ etc.}$$

To obtain a common estimate of the standard deviation of differences, say σ, proceed now, he argued, as follows: (1) calculate the $\frac{1}{2}n(n-1)$ possible values of $s_d^2 = \sum_i (d_{uv\cdot i} - \bar{d}_{uv})^2/m$; (2) multiply each by a factor $m/(m-1)$ so that its

* One variety would appear twice in this mean and its yield must be suitably weighted.

† Gosset at a later date made comments on this point and on the assumption involved in getting a pooled estimate of standard errors that might differ; see (**11**, pp. 285 and 282).

expectation becomes σ^2; (3) sum these quantities and divide by their number. Thus the final estimate of σ^2 becomes

$$s^2 = \frac{2\sum_{u,v}\sum_{i}(d_{uv\cdot i}-\bar{d}_{uv})^2}{n(n-1)(m-1)}. \qquad \text{......(14)}$$

As I shall explain later, this is exactly the estimate which would now be used, only it would be calculated in a more direct manner. The division of Beaven's plots into sets of eight which I have shown in Fig. 1, would to-day be termed a division into blocks (though the blocks are not similar in shape), and the arrangement of the different varieties within a block would be called balanced rather than random. Thus already in 1912 Beaven and Gosset together had gone a long way towards reaching one form of the present-day experimental technique.

Having obtained the estimate s^2 of (14), Gosset was then able to consider the significance of the difference between any pair of varieties by calculating the ratio

$$x = \frac{d_{uv}\sqrt{m}}{s}, \qquad \text{......(15)}$$

and referring to Sheppard's tables.* His method was to place the eight varieties in order of magnitude of the character under consideration and, by applying the test as a foot-rule to selected differences, draw reasoned conclusions as to the existence or absence of real variety differences. A test (R. A. Fisher's z-test) which would determine whether as a whole the eight variety means differed significantly would clearly have been useful, but sound common sense could make the difference test yield reliable results.

This method was applied to the English and Irish chess-board results; the computation was lengthy and many pages of a large notebook of Gosset's are filled with the calculations. G. U. Yule carried out the Cambridge computations in consultation with Gosset. But, however laborious the work, the conclusions obtained from the analysis combined with results of large scale tests played an important part in securing the steady improvement that was being effected in the quality of Irish grown barley.

It is perhaps of historical interest to note a more general formula that Gosset was using at this time to obtain a common estimate of standard deviation from data classified into a number of groups with possibly different means.† The formula would not now be regarded as satisfactory, but it illustrates well the slow progress of the human mind to its final goal.

Suppose that N observations of a variable x are divided into n groups of unequal size, that x_{ti} is the ith observation in the tth group; further that m_t is

* The common estimate, s^2, of (14) is based on so many observations that Gosset probably had not considered whether $\bar{d}_{uv}/s$ could be referred to the z-distribution.

† I have taken the expression from a letter of 1912 to Beaven.

the number and $\bar{x}_t$ the mean in that group. Then Gosset took as an estimate of a supposed common within-group variance, σ^2, the expression

$$s^2 = \frac{1}{N}\sum_{t=1}^{n}\frac{m_t}{m_t-1}\sum_{i=1}^{m_t}(x_{ti}-\bar{x}_t)^2. \qquad \text{......(16)}$$

Since the expectation of $\sum_i (x_{ti}-\bar{x}_t)^2$ is $(m_t-1)\,\sigma^2$ and $N=\sum_t m_t$ it will be seen that the expectation of s^2 is σ^2. Except in the case where m_t is the same for every group, which was the case he was concerned with in the chess-board analysis, the factors weighting the sums of squares are not, however, those which we now know give an estimate of σ^2 having minimum sampling error. When however $m_t = m$ his estimate assumed the correct form

$$s^2 = \frac{1}{n(m-1)}\sum_t\sum_i(x_{ti}-\bar{x}_t)^2. \qquad \text{......(17)}$$

Had he applied formula (16) to the chess-board problem in a case where the number of plots was not the same for all varieties, his final estimate would have been less satisfactory.

During the war period of 1914–19 the analysis of the chess-board results was discontinued. In 1920 Gosset took over responsibility for the statistical aspects of the barley experiments conducted at a number of centres by the Irish Department of Agriculture, and this made him particularly interested in the possibilities of Beaven's new half drill strip method of arrangement. Correspondence with Beaven is full of discussion of the possibilities of this method and of the best way of analysing the results. At the same time he was in touch with R. A. Fisher who was beginning to turn his great mathematical powers to similar problems at Rothamsted.

The next reference I can find to the chess-board analysis is early in 1923, when Beaven had asked Gosset to explain again the procedure he had used ten years before. The final lap of the long passage to an "analysis of variance" is of sufficient historical and personal interest to place on record. On 29 March 1923 Gosset writes:

> I enclose a note on the chess-board error. I was using the formula before the war and see no reason to repent of it. I am writing Fisher asking him to look it over and if necessary criticize.

The method given is that which I have described above, involving the calculation of the $\frac{1}{2}n(n-1)$ squares of differences. It was naturally a lengthy procedure, and I find a brief note of Beaven on the papers, after working through an example: "Conclusion (if any possible) from above is that P.E. with chess-boards might be guessed at almost as well as calculated." It needed a "Student" with his facility for doing calculations in spare moments on the back of an envelope to cope with such computations. But the author of the method himself was not content and on 9 April in the second half of a letter started on the 6th, he writes again to Beaven:

Since writing the above I have had a vision on the subject of chess-board error and enclose a rough proof of my new method. I have written to Yule asking him whether he is in fact working at chess-board error and enclosing a similar proof. If he is *not* I shall be inclined to write it up and shall ask your leave to use the No. 1 chess-board of 1913 as an illustration. If he *is*, he has doubtless got something as good or better, and he can put mine in the W.P.B.

To use my new method with 15 plots, each of 8 varieties (1) find the square of the S.D. of the whole 120 plots, Σ^2; (2) after calculating the averages of the eight varieties, find the square of the S.D. of these eight figures, σ_8^2; (3) after calculating the averages of the fifteen groups of eight, find the square of the S.D. of these fifteen figures, σ_{15}^2. Then the P.E. of the error of a comparison should be

$$0{\cdot}6745\sqrt{\frac{2\times 8(\Sigma^2-\sigma_8^2-\sigma_{15}^2)}{120-8-15}}.*$$

In calculating the S.D.'s do not use the $(n-1)$ divisor.

The "rough proof" of the method which he enclosed was as follows: it will be seen to be on similar lines to that given in the paper "On testing varieties of cereals" (**11**, pp. 282–3) except for the omission of the term $-\sigma_e^2/mn$ referred to in the published paper, which resulted in a divisor of $mn-m-n$ instead of $mn-m-n+1$.

Memorandum

Let m plots of each of n varieties be chessboarded. There will be m groups each containing one of each of the n varieties. If Σ^2 be the variance of the nm plots, it may be considered to be composed of three parts which as a first approximation may be taken as uncorrelated:

(1) The real differences between the varieties, σ_v^2,
(2) The errors common to each group of n, σ_c^2,
(3) The remaining casual errors, σ_e^2.

Of these the last is the only part that affects the comparison of varieties since the differences which we intend to measure compose (1), and (2) is eliminated by the process of chess-boarding.

It remains to find the best estimate of (1), (2) and (3) given Σ^2, the averages of the n varieties, and those of the m groups.

Now if σ_n be the S.D. of the averages of the n varieties

$$\sigma_n^2=\sigma_v^2+\sigma_e^2/m,\dagger$$

and if σ_m be the S.D. of the averages of the m groups

$$\sigma_m^2=\sigma_c^2+\sigma_e^2/n.$$

Also
$$\Sigma^2=\sigma_v^2+\sigma_c^2+\sigma_e^2.$$

* This is the P.E. of the difference between two means of fifteen plots. It must be squared and multiplied by $m=15$ to get into the form of (18) below. [E. S. P.]

† The expression on the right-hand side should have been $\sigma_v^2+\sigma_e^2(n-1)/mn$; this is equal to the expectation of σ_n^2. Similar corrections to the σ_e^2 term are required in the next two equations. [E. S. P.]

Hence $$\Sigma^2 - \sigma_n^2 - \sigma_m^2 = \sigma_c^2\left(1 - \frac{1}{n} - \frac{1}{m}\right),$$

therefore $$\sigma_e^2 = \frac{mn(\Sigma^2 - \sigma_n^2 - \sigma_m^2)}{mn - m - n}.$$

Whence the others follow, and the error of a comparison between a pair of varieties is

$$\sqrt{\frac{2}{m}}\,\sigma_e = \sqrt{\frac{2n(\Sigma^2 - \sigma_n^2 - \sigma_m^2)}{mn - m - n}}.$$

In the next letter to Beaven of 20 April Gosset writes:

> Now as to chess-board error. About a week after I sent the proposed simplified method to you and Yule, I got a note from Fisher via Somerfield giving the same method in rather more technical language. Next I got a reply from Yule saying that the method was new and giving it his blessing more or less, and finally I got a p.c. from Fisher this morning saying that the divisor should be $mn - m - n + 1$ not $mn - m - n$. Anyhow the thing seems to have some weight behind it now.
>
> It should give the same result as my original method....

That the agreement between the two results depends on the identity*

$$\frac{2\sum\limits_{u,v}\sum\limits_{i}(d_{uv\cdot i} - \bar{d}_{uv})^2}{n(n-1)\,(m-1)} = 2 \times \frac{\sum\limits_{u}\sum\limits_{i}(x_{ui} - \bar{x})^2 - m\sum\limits_{u}(\bar{x}_u - \bar{x})^2 - n\sum\limits_{i}(\bar{x}_i - \bar{x})^2}{(n-1)\,(m-1)} \quad \text{......(18)}$$

was shown by Fisher in the letter Gosset quotes in the footnote to p. 283 of his paper (11). The expression on the left-hand side is taken from formula (14) above, while that on the right represents the estimate of the sampling variance of the difference between two single plot yields obtained by the usual analysis of variance method.

Fisher's application of the method was given in a joint paper with W. A. Mackenzie on "Studies in crop variation", received by the *Journal of Agricultural Science* on 20 March 1923 and published in July. The theory was illustrated on an experiment with potatoes "planted in triplicate on the 'chess-board' system"; the arrangement of the plots was not so well balanced as in Beaven's chess-board and as yet no question of randomization was considered. The paper contained what was I think the first published arrangement of numerical data in an analysis of variance table (then described as analysis of variation), and a method was given of testing for the significance of the treatment (or variety) sum of squares, taken as a whole.

"Student's" paper (11) was read before the Society of Biometricians and Mathematical Statisticians on 28 May 1923 and published in *Biometrika* in the following December. In obtaining the formula of the memorandum even with the slip which no doubt he would later have found out himself, and in the description of the method of procedure given to Beaven, he had so evidently after long searching reached the essential conception of breaking up a total sum

* In this notation $d_{uv\cdot i} = x_{ui} - x_{vi}$ or is the difference between uth and vth varieties in the ith block. $\bar{x}_u$, $\bar{x}_i$ and $\bar{x}$ are the variety, the block and the grand mean respectively. There are n varieties and m blocks.

of squares into parts* that I feel his achievement should be put on record. As we have seen, in his modest way he was ready to have his results thrown into the waste paper basket, if another statistician could improve on his work! Whether his mathematics could ever have shown unaided that if no variety differences existed:

(1) the expressions $\Sigma^2 - \sigma_n^2 - \sigma_m^2$ and σ_n^2 of his memorandum were independent,

(2) were each distributed in a modified form of the distribution he had discovered in 1908,

(3) gave a ratio whose distribution law was a Pearson type VI curve;

all this is doubtful. But, as he would have said himself, why speculate, these further results were derived by Fisher; the problem was therefore solved and a new chapter opened.

The 1923 paper (**11**) contains much else of interest besides this handling of the chess-board type of experiment. It starts with an historical survey of the development of experiments aiming at the comparison of cereals and concludes with a critical discussion of the half drill strip method. The simple theme which I have referred to on many occasions runs through the whole and takes form in a final concluding sentence:

It is shown that methods (2) [chess-board] and (3) [half drill strip] depend for their accuracy on the fact that the nearer two plots of ground are situated, the more highly are the yields correlated, so that we are able to increase the effect of the last term of the equation

$$\sigma_{A-B}^2 = \sigma_A^2 + \sigma_B^2 - 2r_{AB}\sigma_A\sigma_B$$

(where A and B are the varieties to be compared) by placing the plots to be compared with one another as near together as possible.

LATER PAPERS

In his later papers Gosset tended to avoid, as far as possible, the introduction of mathematics and he would ask his friends to regard him as a non-mathematician. Thus he forwarded his paper on the Lanarkshire milk experiment (**17**) to Karl Pearson with the words:

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

Or again, writing to me in 1926 regarding the original χ^2 paper (Karl Pearson, 1900) he remarked:

I have now read the χ^2 paper in *Phil. Mag.* **50**. 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 penknife might appreciate a model yacht cut and rigged to scale, the second I can

* His original approach to statistics through Airy's book made this a natural way of regarding things; see the formula (5) I have quoted above. There are points in Gosset's proof in (**11**, p. 282) also reminiscent of Airy, *Theory of Errors of Observations* (1875, p. 46).

only compare to a conjuring trick of which I haven't got the key (such for mexaple as the transformation to polar co-ordinates on p. 158) and lastly quite a small part which I think I can understand.

When at last, after the war, an increasing number of men trained as mathematicians began to turn their attention to statistics, it was not perhaps surprising that one whose mathematical training had ceased with Oxford Mods. in the nineties should refuse to regard himself as a mathematician. Besides, the increasing responsibilities of his work as a brewer left him little time or inclination to follow out in detail the continuous elaboration of the theory of mathematical statistics. As a result, in his relatively rare publications he tended to concentrate on simple exposition of the function of statistical method. The best examples of such work are:

(1) The paper on "Errors of routine analysis" of 1927 (**15**) which develops more fully a theme he had touched on before (**4** and **8**), and shows how some recent theoretical work on the distribution of "range" in small samples might be made to give a useful working tool for the analyst.

(2) Two admirable papers on the use of statistical methods in agriculture, both unfortunately rather inaccessible to the ordinary student: "Mathematics and Agronomy", 1926 (**14**), and the article on "Yield Trials" in *Baillière's Encyclopedia of Scientific Agriculture* 1931 (**16**).

This recession from the mathematical approach of his earlier papers had other consequences. In the first place it meant that during a period of rapid advance in statistical technique there was available, for almost anyone in need of advice, a statistician of great practical experience and unusual insight, whom the inquirer could be sure would not be carried away by the fascination of any mathematical model into allowing abstract theory to step beyond its proper sphere. On the other hand there were certain disadvantages; Gosset's avoidance of a mathematical statement of his case sometimes, as in his last two papers (**21**), (**22**), made it difficult for others to grasp an idea or method which probably was clear enough in his own mind. The theory of probability is based on mathematics, and beyond a certain point there are dangers in introducing it into practice without a precise mathematical statement of the assumptions underlying the method of procedure.

If we return to 1923, it is clear that Gosset welcomed with enthusiasm the new methods that R. A. Fisher was developing. The neatness of the arrangement of calculations in an analysis of variance table for example, appealed to him. It brought to the rather laborious calculation methods of his own a simplification whose value he was quick to realize. The introduction of t as the ratio of a deviation to an estimate of its standard error, in place of his own criterion z, and the use of degrees of freedom, appealed to him at once because of the greater generality; as a result he calculated extended values of the probability integral of t to replace his old z tables and published these in 1925 (**13**) in conjunction with a theoretical contribution of Fisher's. In print and in correspondence he emphasized the importance of randomness. "The experiments", he wrote in 1926 (**14**, p. 711), "must be capable of being considered to be a *random* sample of the population to which the conclusions are to be applied. Neglect of this rule has led to the estimate of the value of statistics which is expressed in the crescendo 'lies, damned lies, statistics'."

This paper of 1926 contains perhaps the extreme limit to which he ventured in allowing the toss of a coin or a die to decide the arrangement of plots in an agricultural experiment. On the last page (p. 719) he suggests the arrangement of four varieties in an 8×8 square, in which two plots of each variety are to fall in each row and each column. Subject to this restriction the arrangement was to be obtained in a random manner.

He must soon, however, have realized the disadvantages of such a procedure. If A, B, C and D represent the varieties, a possible if unlikely run of luck might lead to the following pattern of plots in one corner of the square:

A	*A*	*C*	*D*	*C*
A	*A*	*C*	*D*	*B*
B	*C*	*A*	*A*	*D*
C	*C*	*C*	*B*	*A*

Fig. 2.

Should this chance juxtaposition of many A-plots happen to coincide with a "fertility summit" or "depression" in the field, the resulting statistical analysis of plot yields might easily attribute a characteristic to the variety A which it did not possess. His practical mind could not accept such a state of affairs. To know in advance that if an experiment was carried out with a particular pattern of plots there was quite a chance that it would be misleading, and to continue with this pattern—this was a course he was not prepared to follow. It was no compensation to be told that in the long run, if the verdict of the random toss was accepted and the 5% significance level of mathematical tables used in the statistical analysis, then misleading results would be obtained only 5 times in 100. In his own words (**22**, p. 366):

> It is of course perfectly true that *in the long run*, taking all possible arrangements, exactly as many misleading conclusions will be drawn as are allowed for in the tables, and anyone prepared to spend a blameless life in repeating an experiment would doubtless confirm this; nevertheless it would be pedantic to continue with an arrangement of plots known beforehand to be likely to lead to a misleading conclusion.

His withdrawal from the out and out randomization position is illustrated in his article of 1931 on "Yield Trials" (16). Here he speaks of the Latin square arrangement as ideal in the types of experiment for which it is suited, because it combines the elements of balance and randomness, but he is critical of the randomized block arrangement because of the risk involved of getting misleading results. He gives the following illustration of a balanced or equalized block design which he had recommended to a horticultural correspondent, comparing ten treatments with five replications:

G	*H*	*E*	*C*	*A*	Block I
F	*D*	*J*	*B*	*I*	
H	*J*	*D*	*F*	*E*	Block II
B	*G*	*I*	*A*	*C*	
E	*I*	*A*	*G*	*D*	Block III
J	*B*	*C*	*H*	*F*	
C	*F*	*B*	*I*	*J*	Block IV
A	*E*	*G*	*D*	*H*	
D	*A*	*F*	*J*	*B*	Block V
I	*C*	*H*	*E*	*G*	

Fig. 3.

In this example the assignment of treatments to plots in Block I is random, but each successive block has its arrangement more and more controlled, so that (i) each of the five columns contains one plot only of the ten varieties, (ii) *A*, *D*, *E*, *F* and *J* occur in the top row of their block three times and in the lower row twice, while for *B*, *C*, *G*, *H* and *I* the position is reversed, an arrangement as nearly balanced as possible for an odd number of blocks.

In advocating the introduction of this element of balance, he did not consider that the random element could be dispensed with; but he believed that if a regular pattern was used to equalize the more probable variations in fertility there were still sufficient complications to leave the residual variations random enough to justify from the practical point of view the application of probability theory. It was here that he disagreed and was eventually forced into open controversy with R. A. Fisher and the Rothamsted school.

This is not the place to enter into detail regarding the nature of this controversy, which resulted in Gosset's last paper published a few months after his death (22). It is however well to emphasize that his attitude was closely related to the type of agricultural problem with which he had had most experience, the development of improved strains of barley. In such a case as this he saw that success was only likely to result from a comparison of two or more strains in a number of years and in a number of different localities. Small scale investigations must be followed by others in which the technique conformed as far as possible to ordinary agricultural practice. In each case some experimental plan was needed which would give the yields, let us say, of variety A and variety B on the experimental area with as little error as possible, that is to say freed from bias such as might be introduced by changes in fertility, patches of weed, etc. Provided that the error of the difference (yield of A − yield of B) could be kept low, he was satisfied with a knowledge of its probable upper limit and did not mind if he was told that the ratio of this difference to the estimate of its standard error in a particular experiment could not be referred with mathematical precision to a table of probabilities. He was interested primarily in the behaviour of the difference from farm to farm and year to year, and experience had shown him, beyond any possibility of doubt, that small scale balanced plot experiments followed by larger scale tests with the half drill strip method of Beaven's, the purpose of which any intelligent farmer could understand, had achieved remarkable success in the improvement of barley. If it were argued that fully randomized experimental designs would have achieved the same or better results he would not have denied this dogmatically, but he felt doubtful on the point because his perusal of reports on such experiments showed to his mind an unduly high proportion of inconclusive results. He would also have added that with the staff, the ground and other facilities available in the investigations for whose planning he was responsible, fully randomized designs could not have been carried out. This was his attitude in writing the Statistical Society paper of 1936 (21).

In his final paper (22) he attacked his critics on their own ground by pointing out that in the experiment at a single station a balanced arrangement of plots in blocks was on the whole more likely to detect variety differences than a random arrangement when those differences were really large and therefore important, although for small differences the reverse would be true.

The ultimate decision on these points can hardly be expected as yet; it will come in time, perhaps after 10 or after 20 years, when there has been ample opportunity for the practical experimenter, freed from the weight of authority, from fear of mathematics on the one side and from the fascination of a new technique on the other, to judge from accumulated experience what methods have been most worth while having regard to the results they have achieved.

In addition to these papers on agricultural subjects a brief reference may be given to some other published work of the last few years:

(1) A paper on "The Lanarkshire milk experiment", 1931 (17); his suggestion

that the experiment should be repeated on a more precise but far less expensive scale by using pairs of twins involves a characteristic introduction of his paired difference plan.

(2) Two papers on certain implications of F. L. Winter's selection experiments with maize 1933 (19) and 1934 (20). The plant breeder's problem of improving varieties of cereal by continued selection had long been of interest to him in connexion with barley and in these papers he discusses the bearing of these experimental results upon evolutionary theory.

(3) A number of short but suggestive contributions to the discussion of papers read before the Industrial and Agricultural Research Section of the Royal Statistical Society (see references on p. 249 below).

Extracts from letters

I have spoken more than once of Gosset's correspondence; the professional statistician, whether he be attached to a university or research station, receives and expects to receive appeals for advice which will continue to increase through life as his circle of contacts grows. But with Gosset the position was somewhat different; to provide advice to correspondents all over the world was in no way part of his job. Yet he gave that help unstintingly and unless it could be described as brewery business, he gave it out of his own time. Advice as to how to plan a particular experiment, or explanations of misunderstood points in statistical theory, while of extreme value at the time to the individual who receives them are rarely of interest to the general reader. Nevertheless, I believe that a few quotations from letters will add to the record of Gosset's personality by showing something of his patience, his practical mind, his suggestiveness and his characteristic freedom of expression.

The first quotations are taken from a long letter written to me in 1926. At that time I had been trying to discover some principle beyond that of practical expediency which would justify the use of "Student's" ratio $z=(\bar{x}-m)/s$ in testing the hypothesis that the mean of the sampled population was at m. Gosset's reply had a tremendous influence on the direction of my subsequent work, for the first paragraph contains the germ of that idea which has formed the basis of all the later joint researches of Neyman and myself. It is the simple suggestion that the only valid reason for rejecting a statistical hypothesis is that some alternative hypothesis explains the observed events with a greater degree of probability. The second part of the letter probably put into my mind the very extensive plan of sampling from non-normal populations which we carried out in the Department of Statistics at University College during the next few years.

Letter I

From a letter of W. S. G. to E. S. P., dated 11 May 1926.

In your large samples with a known normal distribution you are able to find the chance that the mean of a random sample will lie at any given distance from the mean of the population. (Personally I am inclined to think your cases are best considered as mine taken to the limit n large.) That doesn't in itself necessarily prove that the sample is not drawn randomly from the population even if the chance is very small, say ·00001: what it does is to show that if there is any alternative hypothesis which will explain the occurrence of the sample with a more reasonable probability, say ·05 (such as that it belongs to a different population or that the sample wasn't random or whatever will do the trick) you will be very much more inclined to consider that the original hypothesis is not true.

I can conceive of circumstances, such for example as dealing a hand of 13 trumps after careful shuffling *by myself*, in which almost any degree of improbability would fail to shake my belief in the hypothesis that the sample was in fact a reasonably random one from a given population.

* * * * * * *

I'm more troubled really by the assumption of normality and have tried from time to time to see what happens with other population distributions, but I understand that you get correlation between s and m with *any* other population distribution.

Still I wish you'd tell me what happens with the even chance population ▭ or such a one as △: it's beyond my analysis.

* * * * * * *

If Student is wrong it is up to you to give us something better. You see one must experiment and frequently it is quite out of the question, from considerations of cost or of impossibility of duplicating conditions in the time scale, to do enough repetitions to define one's variability as accurately as one could wish. It's no good saying "Oh these small samples can't prove anything". Demonstrably small samples *have* proved all sorts of things and it is really a question of defining the amount of dependence that can be placed on their results as accurately as we can. Obviously we lose by having a poor definition of the variability but *how much do we lose*?

Letter II, with its enclosure which, for reasons I have forgotten, was never published, was written shortly after K. P. had made an editorial comment on "Sophister's"* (1928) interpretation of the distribution of "Student's" ratio in samples from a non-normal population. It had been found that in such cases the distribution of t was asymmetrical, but that the distribution of $|t|$ (or of t^2) followed very closely the standard normal-theory form, i.e. if the distribution of t was curtailed on one side of the origin this was balanced by a corresponding extension on the other side. The letter also refers to a suggestion of bringing up at a meeting of the International Statistical Institute the question of differentiating between the symbols used for probable error and standard error.

* "Sophister" like "Mathetes" was the *nom de plume* of a disciple of "Student". The particular sampling investigation in question had been sketched out by Gosset and myself before "Sophister" came to spend a year in the Biometric Laboratory.

Letter II

Holly House,
Blackrock,
Co. Dublin.
May 18th, '29.

Dear Pearson,

I was rather amused to see your letter open with an apology for delay in writing as I have for some time been acutely conscious that I have been in arrears. However, last things first.

(i) I agree that Z's second suggestion though sound is not workable. Your idea of raising the question at Warsaw seems to me to suggest the right way of getting to work. I think they should raise the question on the grounds (*a*) that $\pm$ is being used in two senses, (*b*) that the prob. error is no longer the slightest use to anyone and (*c*) that as the tables are in terms of the S.D. a simple notation such as : or ; or anything of the sort is required.

(ii) I fancy you give me credit for being a more systematic sort of cove than I really am in the matter of limits of significance. What would actually happen would be that I should make out P_t (normal) and say to myself "that would be about 50 : 1; pretty good but as it may not be normal we'd best not be too certain", or "100 : 1; even allowing that it may not be normal it seems good enough" and whether one would be content with that or would require further work would depend on the importance of the conclusion and the difficulty of obtaining suitable experience.

One so often finds that the importance (and even occasionally the direction of the result) of varying one factor, change from experiment (or experience) to experiment according to the accompanying variations in other factors, that it often doesn't pay to make too certain of any one result.

E.g. You may have two varieties of barley one of which will give the best yield in one season or place while the other will win in another season or place; hence we have to sample places and seasons widely rather than aim at being meticulously accurate at all places sampled: there must be economy of effort.

* * * * * * *

Lastly I am enclosing a short note in reply to the Editorial footnote. Probably you are going to say all that is at all useful in it in your next paper, and in any case I haven't the least intention of indulging in a controversy, so suppress it unless you think it will clear up our position. All the same I think it is a pity to let the thing go by default without any comment.

Yours v. sincerely,
W. S. Gosset.

Suggested Note for "Biometrika"

17th May, 1929.

In his footnote on page 422 of Sophister's paper the Editor asks, "Supposing 50 per cent 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 judgment?" This, if I may say so, is hardly what Sophister proposes to do. If I may deal first with the Editorial analogy the position is rather, "The evidence before the court is such that the chances are even that the prisoner committed the murder". Doubtless if more evidence were forthcoming we should know more about it; as it is, an English Court will acquit, though the inexorable Justice of Shan Tien would condemn the prisoner to piecemeal slicing, unless of course sufficiently weighty evidence for the defence could be imperceptibly introduced within the Mandarin's sleeve. But, seriously, a better illustration can be drawn from the practice of

Insurance where in the first place the premium is calculated on the Healthy Male table and, I suppose, originally this was the only basis after a medical examination. But the material which supplied the experience for the H.M. table can be subdivided into various classes, by professions and occupations, by stature or eye colour, total abstainers or moderate drinkers and so forth, which further investigation may find to have expectations of life which do not accord with the table. The life expectation of some of these classes is probably taken into consideration by the Companies—I doubt whether a Lion Tamer, however healthy, could insure at the ordinary rates—but no company, as it well might, charges a lower rate of premium for the descendants of centenarians or a higher for orphans; they are most unfairly lumped together just as Sophister proposes to do with his samples from unknown populations. In effect he says, "This small sample is from an unknown population, which *may* be normal; it probably is not far from normal; if it *is* normal we use the table justly, if it is anormal but symmetrical we can still use the table with sufficient accuracy; even if it is skew, about which we cannot be sure—much less about the direction of the skewness—we shall in the long run draw much the same proportion of correct inferences as if it were normal." Admittedly our ignorance of the nature of the population introduces an element of uncertainty which no sensible person will ignore when using the tables, but recent work, and not least Sophister's, shows that this uncertainty, while not altogether negligible, is much less than we had any right to expect.

Student.

The suggestion in Letter III of 1932 ultimately led to the production of tables of percentage limits of the ratio of (*a*) range in a sample of *n* observations to (*b*) an independent estimate of standard deviation, which are to be published shortly in *Biometrika*. From the beginning of his analysis of the results of the chess-board experiments, Gosset had wondered how best to judge what differences among variety means were significant. While the ratio of (*a*) the difference between any two means selected at random to (*b*) the estimate of standard error could be referred to "Student's" distribution or, if desired, the significance of the set as a whole could be judged by Fisher's z-test, it was not possible to treat selected differences in either of these ways. In the article in *Baillière's Encyclopedia* (**16**, p. 1358) he refers to a method suggested by Fisher of taking the differences between individual variety yields and the mean yield. He felt however that a knowledge of the probability levels of "studentized" range would in addition be very useful; on this could be based a rough test of the kind he had suggested in his paper on "Errors of routine analysis" (**15**, p. 161).

Letter III

St. James's Gate,
Dublin.
Jan. 29th, '32.

Dear Pearson,

Many thanks for your letter and enclosure: as I am at the moment

"The Cook and the Captain bold
And the mate of the Nancy brig",

I have handed all the lot to Mathetes till such time as I can get a chance of dealing with it which should be sometime next week.

I have been meaning to write to you for some time re the proposals for the use of range and sub-range which I made in my last letter to you. Of course there is a serious crab which

I had at one time recognised and then forgotten in that the thing would have to be "Studentised": the only measure of the S.D. is provided by a limited number of degrees of freedom. Whether one could get an approximate correction for this with moderately small numbers by reducing still further the degrees of freedom or whether it would be necessary as Fisher suggested when I mentioned the matter to him (he was here lecturing) to dive into the depths of hyperspace to produce the jewel I am not clear, but obviously something would have to be done about it.

* * * * * * *

Yrs. v. sincerely
W. S. Gosset.

Letters IV and V of 1936, which Dr Beaven has kindly allowed me to reproduce, deal with the interpretation of the results of half drill strip barley experiments carried out at six stations in England; the two varieties compared were Plumage Archer and Beaven's 35/7. The second letter followed a reply from Beaven discussing the position in terms of betting on two horses, whose form varies on different courses. The argument illustrates Gosset's outlook on the function of large scale experiments to which I have already referred.

Letter IV

From a letter of W. S. G. to E. S. B., dated 8 January 1936.

If you derive the S.E. from a set of 10 strips at one station, you are sampling "comparisons between plots grown at a certain station in the weather of 1935" and can draw the appropriate conclusion, e.g. that at Sprouston it is quite certain that Beaven's 35/7 would have beaten Plumage Archer in any sound arrangement of plots in 1935.

When however you regard the six stations as a small sample of the barley land of England you can very nearly draw the conclusion that Beaven's 35/7 would on the average have beaten Plumage Archer if compared all over the barley land of England in 1935.

The chance that so favourable a result would have happened if there were really no difference between them is only 1/38, i.e. the odds are 37 to 1 against it's happening. This is very nearly significant but as you know, what odds are to be considered significant is a matter of convention—or taste.

Naturally, in calculating the S.E. (not really an *error* at all) of the second conception where the variation from Station to Station depends as much, (or much more. . . than), on the differential response to weather and soil as on the soil errors taken account of in each station, one takes no particular account of the S.E.'s at the individual stations: one merely rejoices because the Half Drill Strip method has largely eliminated the errors due to soil position and left us mainly the differential response aforesaid, which would have affected the result to a greater or less extent in every field of barley-growing England and which we have assumed that we have sampled by the six results which we have examined.

I hope I have made the distinction clear between the S.E. of the result at one station, which is rightly derived from the plots grown at that station but which only enables us to judge whether the result is significant for that station, and the S.E. of the whole series, derivable only from the six mean results of the six stations but which enables us to make an estimate of the result of comparing the barleys "everywhere", where "everywhere" represents the whole extent of country that may properly be considered to be sampled by the six stations.

Letter V

Davan Hollow,
Denham,
Bucks.
14. 1. 36.

Dear Beaven,

I don't think your analogy is quite exact: this is mine.

The two horses 35/7 and P.A. are known to vary somewhat from day to day and also to be very much affected by the particular course on which they are running.

They have raced ten times at Sprouston and 35/7 has won every time by amounts varying from one furlong to two furlongs. At Sprouston then you may lay longish odds on 35/7. At Cambridge they raced ten times and on the average 35/7 won by 50 yds, the amounts varying from 270 yds in favour of 35/7 to 170 in favour of P.A. You would not therefore bet very heavily on 35/7 at Cambridge. At four other places 35/7 beat P.A. on average by various amounts. What odds is to be given on another hitherto untried course?

You are surely as much influenced by the narrowness of the margin at Cambridge as by the width of it at Sprouston: the new course may resemble the one with just as much likelihood as the other and may even as far as you can see favour P.A. rather than 35/7, since your knowledge of the difference between courses rests on only six cases.

Furthermore a new method of training may reduce the variation so that the Sprouston results may lie between $1\frac{1}{4}$ and $1\frac{3}{4}$ furlongs and the Cambridge between 160 yds in favour of 35/7 and 60 yds in favour of P.A., without altering very much* the odds on a series of races on a new course, since the chief source of variation remains the reaction of the horses to the courses and not the day to day variation which alone is measured by the variation on a single course.

* * * * * * *

Yours v. sincerely
W. S. Gosset.

* But since the smaller day to day variation prevents an accidentally high or low value of mean obscuring the real value of the course there is a better chance of getting the right odds—not of getting higher odds.

Letter VI was written at the time when Gosset was putting together his last paper (**22**).

Letter VI

Dart Cottage,
Postbridge,
Devon.
19. iv. 37.

Dear Pearson,

Many thanks for yours of 10th; I feel I'm rather wasting your time but as long as you ask questions you must expect to get answers. You have given my reason for not changing the level of significance *viz.* that while balancing certainly *tends* to produce a lower real error and consequently higher calculated error one cannot say how much one has succeeded in any particular case. I therefore content myself with pointing out that the tendency is beneficial, not only are the cases missed of comparatively little value but one actually gets more conclusions of real value.

* * * * * * *

Now I was talking about Cooperative experiments and obviously the important thing in such is to have a low real error, not to have a "significant" result at a particular station. The latter seems to me to be nearly valueless in itself. Even when experiments are carried out only at a single station, if they are not mere five finger exercises, they will have to be part of a series in time so as to sample weather and the significance of a single experiment is of little value compared with the significance of the series—which depends on the real error not that calculated for each experiment.

But in fact experiments at a single station *are* almost 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 only "In the old cow field at Rabbitsbury". What you really want to find out is "In what soil and under what conditions of weather do potatoes utilise the addition of potash manures?"

To do that you must try it out at a representative sample of the farms of the country and correlate with the characters of the soil and weather. It may be that you have an easy problem, like our barleys which come out in much the same order wherever—in reason—you grow them or like Crowther's cotton which benefitted very appreciably from nitro-chalk in seven stations out of eight, but even then what you really want is a low real error. You want to be able to say not only "We have significant evidence that if farmers in general do this they will make money by it", but also "we have found it so in nineteen cases out of twenty and we are finding out why it doesn't work in the twentieth". To do that you have to be as sure as possible which *is* the 20th—your real error must be small.

* * * * * * *

Tedin:* Somerfield sent me the number and I have just had time to glance at it. T. put down three kinds of patterns of Latin Squares (5×5) on various uniformity trials. There were

Two Knight's moves:

A	*B*	*C*	*D*	*E*
D	*E*	*A*	*B*	*C*
B	*C*	*D*	*E*	*A*
E	*A*	*B*	*C*	*D*
C	*D*	*E*	*A*	*B*

Two Diagonals:

A	*B*	*C*	*D*	*E*
E	*A*	*B*	*C*	*D*
D	*E*	*A*	*B*	*C*
C	*D*	*E*	*A*	*B*
B	*C*	*D*	*E*	*A*

and a number of randoms.

Of course all Latin squares are "balanced" but one wouldn't care too much for the "Diagonal" arrangement and the Knight's move would, I think, be preferred to all others. In conformity with this Tedin found a slight tendency for the Knight's move to give a low actual and a high calculated error while the diagonal tends to give a high actual and a low calculated error. The whole thing is not worth worrying about but is interesting as an illustration of what actually happens when we depart from artificial randomisation: I would Knight's move every time!

Yours

W. S. G.

P.S. Beaven after all got some slight ailment which prevented his being in the chair for Bartlett's paper: I proposed the vote of thanks....I was heard without enthusiasm but there were no cat calls!

Such are my impressions of Gosset and of his work. Others will have different views on the relative importance of his many contributions to statistics; on his rightness or wrongness. The experimentalist will have seen him in a different light from the mathematician; his personal friends will have realized aspects of his character which his correspondents could not see. But all who have known him will agree that he possessed almost more of the characteristics of the perfect

* A reference to the paper by O. Tedin (1931).

statistician than any man of his time. They will agree, too, on the essential balance and tolerance of his outlook, and on that something which a friend of his schooldays has described as an "immovable foundation of niceness" which made him through life the same friendly dependable person, quiet and unassuming, who worked not for the making of personal reputation, but because he felt a job wanted doing and was therefore worth doing well.

BIBLIOGRAPHY OF "STUDENT'S" PAPERS

(1) 1907. "On the error of counting with a haemacytometer." *Biometrika*, **5**, 351.

(2) 1908. "The probable error of a mean." *Biometrika*, **6**, 1.

(3) 1908. "Probable error of a correlation coefficient." *Biometrika*, **6**, 302.

(4) 1909. "The distribution of the means of samples which are not drawn at random." *Biometrika*, **7**, 210.

(5) 1911. Appendix to paper by W. B. Mercer and A. D. Hall on "The experimental error of field trials." *J. Agric. Sci.* **4**, 128.

(6) 1913. "The correction to be made to the correlation ratio for grouping." *Biometrika*, **9**, 316.

(7) 1914. "The elimination of spurious correlation due to position in time or space." *Biometrika*, **10**, 179.

(8) 1917. "Tables for estimating the probability that the mean of a unique sample of observations lies between $-\infty$ and any given distance of the mean of the population from which the sample is drawn." *Biometrika*, **11**, 179.

(9) 1919. "An explanation of deviations from Poisson's law in practice." *Biometrika*, **12**, 211.

(10) 1921. "An experimental determination of the probable error of Dr Spearman's correlation coefficients." *Biometrika*, **13**, 263.

(11) 1923. "On testing varieties of cereals." *Biometrika*, **15**, 271.

(12) 1924. Note by "Student" with regard to his paper "On testing varieties of cereals." *Biometrika*, **16**, 411.

(13) 1925. "New tables for testing the significance of observations." *Metron*, **5**, 105.

(14) 1926. "Mathematics and Agronomy." *J. Amer. Soc. Agron.* **18**, 703.

(15) 1927. "Errors of routine analysis." *Biometrika*, **19**, 151.

(16) 1931. Article on "Yield Trials" in *Baillière's Encyclopedia of Scien* **2**, 1342.

(17) 1931. "The Lanarkshire milk experiment." *Biometrika*, **23**, 398.

(18) 1931. "On the 'z' test." *Biometrika*, **23**, 407.

(19) 1933. "Evolution by selection. The implications of Winter's selection experiment." *Eugen. Rev.* **24**, 293.

(20) 1934. "A calculation of the minimum number of genes in Winter's selection experiment." *Ann. Eugen., Lond.*, **6**, 77.

(21) 1936. "Co-operation in large-scale experiments." A discussion opened by W. S. Gosset. *J.R. Statist. Soc.* Suppl. **3**, 115.

(22) 1937. "Random and balanced arrangements." *Biometrika*, **29**, 363.

A FEW SHORTER CONTRIBUTIONS

(*a*) *Letters to "Nature"*.

29 November 1930, **126**, 843: "Agricultural Field Experiments".

14 March 1931, **127**, 404: "Agricultural Field Experiments".

5 December 1936, **138**, 971: "The half drill strip system of Agricultural Experiments".

(*b*) *Contributions to discussions at meetings of the Industrial and Agricultural Research Section of the Royal Statistical Society.*

J.R. Statist. Soc. Suppl. (1934), **1**, 18; (1936), **3**, 173; (1937), **4**, 89, 170.

REFERENCES TO PAPERS BY OTHER AUTHORS

Cave, F. E. (1904). *Proc. Roy. Soc.* **74**, 403.
Helmert, F. R. (1876). *Astr. Nachr.* **88**, S. 122.
Hooker, R. H. (1905). *J.R. Statist. Soc.* **68**, 696.
Macdonell, W. R. (1901). *Biometrika*, **1**, 219.
Mercer, W. B. & Hall, A. D. (1911). *J. Agric. Sci.* **4**, 107.
Neyman, J. (1937). *Lectures and conferences on mathematical statistics.* Graduate School of U.S. Dept. of Agriculture, Washington.
Pearson, K. (1900). *Phil. Mag.* **50**, 157.
Pearson, K. (1907). *Drapers' Company Research Memoirs.* Biometric Series, **4**.
Pearson, K. (1913). *Biometrika*, **9**, 116.
Soper, H. E. (1913). *Biometrika*, **9**, 91.
"Sophister" (1928). *Biometrika*, **20***a*, 389.
Tedin, O. (1931). *J. Agric. Sci.* **21**, 191.
Whittaker, L. (1914). *Biometrika*, **10**, 36.
Wood, T. B. & Stratton, F. J. M. (1910). *J. Agric. Sci.* **3**, 417.