'Student' and Small-Sample Theory

E. L. Lehmann*

Abstract

The paper discusses the contributions Student (W. S. Gosset) made to the three stages in which small-sample methodology was established in the period 1908–1033: (i) the distributions of the test-statistics under the assumption of normality; (ii) the robustness of these distributions against non-normality; (iii) the optimal choice of test statistics. The conclusions are based on a careful reading of the correspondence of Gosset with Fisher and E. S. Pearson.

Keywords and phrases: History of statistics, "exact" distribution theory, assumption of normality, robustness, hypothesis testing, Neyman–Pearson theory.

*E. L. Lehmann is Professor Emeritus, Department of Statistics, University of California, Berkeley, California 94720.

1 Introduction

In an interview in Statistical Science (Vol. 4, 1989), F. N. David talks about statistics in the 1920's and 30's as developed by Gosset (Student), Fisher, Egon Pearson,[1] and Neyman. She describes herself as a contemporary observer who saw "all the protagonists from a worm's eye point of view".[2] Her surprising assessment:

(1.1) "I think he [Gosset] was really the big influence in statistics... he asked the questions and Pearson and Fisher put them into statistical language, and then Neyman came to work with the mathematics. But I think most of it came from Gosset."

Here she is of course not talking about all of statistics in this period, but of the development of the new small-sample approach. Nevertheless, her claim is surprising since Gosset is mainly known for only one, although a pathbreaking contribution: Student's t-test. The aim of this paper is to consider to what extent David's conclusion is justified.

The basis for the new methodology was established in three stages. Stage 1 (Student–Fisher) determined the distributions of the statistics used to test means, variances, correlation and regression coefficients under the assumption of normality. At the second stage (Pearson), the robustness of these distributions under non-normality was investigated. Finally at the last stage Neyman and Pearson laid the foundation for a rational choice of test statistics not only in the normal case but quite generally. In the following sections we shall

[1] In the remainder of this paper we shall usually refer to E(gon) S. Pearson simply as Pearson and to his father as Karl Pearson or occasionally as K.P.

[2] For anyone like myself who knew the feisty David later in her life it is hard to imagine her ever playing the role of a "worm".

consider the contributions Gosset made to each of these stages.

An author writing about this period is fortunate to have available a wealth of material that frequently makes it possible to trace mutual influences and the development of ideas in considerable detail. The principal sources I have used are acknowledged at the end of the paper.

2 The New Methodology

2.1 Gosset's 1908 Papers

The event that with little fanfare and no particular enthusiasm on the part of the Editor ushered in a new era in statistics was the Biometrika publication in 1908 of a paper "The probable error of the mean" by 'Student' the pseudonym of William Sealy Gosset. The reason for the pseudonym was a policy by Gosset's employer, the brewery Arthur Guinness Sons and Co., against work done for the firm being made public. Allowing Gosset to publish under a pseudonym was a concession which resulted in the birth of the statistician 'Student', the creator of Student's t-test.

Today the pathbreaking nature of this paper is generally recognized and has been widely commented upon, among others by Pearson (1939), Fisher (1939), Welch (1958), Mosteller and Tukey (1977), Section (B), Box (1981), Tankard (1984), Pearson (1990), Lehmann (1992), and Hald (1998, Section 27.5).

The core of the paper consists of the derivation of the distribution of

$$z = (\bar{X} - \mu)/S \tag{2.1}$$

where $X_1, \ldots, X_n$ are i.i.d. with normal distribution $N(\mu, \sigma^2)$ and where $S^2 = \Sigma(X_i - \bar{X})^2/n$.[3] This derivation was a difficult task for Gosset who was a chemist, not a mathematician. And although he obtained the correct answer, he was not able to give a rigorous proof.

He began with calculating the first four moments of S^2, from them conjectured that the distribution of S^2 might be a Pearson type III distribution, and concluded that

(2.2) "it is probable that the curve found represents the theoretical distribution of S^2, so that although we have no actual proof we shall assume it to do so in what follows."

To obtain the distribution of z, he next calculated the correlation of $\bar{X}$ and S and, finding them to be uncorrelated, concluded that they were independent. (He should not be blamed for this confusion since even Karl Pearson did not appreciate the distinction between independence and lack of correlation; see Reid (1982, p. 57).) From there the derivation of the distribution of z was a fairly easy calculus exercise.

However, this result—although it undoubtedly was what caused Student the most effort—was not the reason for the enormous influence of the paper. The principal contribution was that it brought a new point of view. It stated the need for methods dealing with small samples, for which the normal approximations of the theory of errors were not adequate. And it brought the crucial insight that exact results can be obtained when the form of the distribution of the observations is known.

Student summarized this new approach in the Introduction of his paper as

[3] The definition of z and S^2 do not agree with current usage but are those used by Student.

follows:

> "Although it is well known that the method of using the normal curve is only trustworthy when the sample is 'large', no one has yet told us very clearly where the limit between 'large' and 'small' samples is to be drawn.
>
> (2.3) The aim of the present paper is to determine the point at which we may use the tables of the probability integral in judging of the significance of the mean of a series of experiments, and to furnish alternative tables for use when the number of experiments is too few."

The calculation of these tables requires knowing the shape of the underlying distribution of the observations. However, as Student points out, for small samples

> (2.4) "the sample is not sufficiently large to determine what is the law of distribution of individuals. It is usual, however, to assume a normal distribution...: since some law of distribution must be assumed it is better to work with a curve whose area and coordinates are tabled, and whose properties are well known. This assumption is accordingly made in the present paper, so that its conclusions are not strictly applicable to populations known not to be normally distributed; yet it appears probable that the deviation from normality must be very extreme to lead to serious error."

What does Student mean when he writes that "it is usual to assume a normal distribution"? He learned statistics by reading two books: Airy's "Theory of Errors of Observations" (1879) and Merriman's "Method of Least Squares" (1884). Both emphasize that errors are typically sums of a large number of independent small components and hence are approximately normally distributed. In fact, the normal distribution is called the Law of Probability of Errors. Merriman says about it (p. 33): "Whatever may be thought of the theoretical deductions of the law of probability of error, there can be no doubt that its practical demonstration by experience is entirely satisfactory." Airy (p. 24) is slightly more cautious since he warns: "It must always be borne in mind that the law of frequency of errors does not exactly hold except the number of errors [i.e. components of error] is infinitely great. With a limited number of errors the law will be imperfectly followed; and the deductions, made on the supposition that the law is strictly followed, will be or may be inaccurate or inconsistent."

Thus, Student's assumption of normality is grounded in a well-established tradition.

Student's work leading up the 1908 paper was done largely while he was on a study leave in Karl Pearson's Department at University College, where among other new ideas, he learned about the (Karl) Pearson system of curves. Commenting on Student's derivation of his distribution Egon Pearson (1939) says that "it is doubtful… if he could have reached the distribution of S^2 (and hence of z) if he had not had available for use Pearson's type III curve". In the same vein, Fred Mosteller once commented to me that Student might not have discovered his distribution had he been born in Scandinavia instead of England, since then he would have tried to fit a Gram–Charlier distribution rather than a curve from the Pearson system.

Student illustrated the use of his distribution by three examples, including one of a paired comparison experiment which he reduces to the one-sample situation by taking differences. Finally there is also a table of the z-distribution for sample sizes 4 to 10. He later extended it to sample sizes 2 to 30 (Student, 1917). The change from z to $t = z\sqrt{n-1}$ which is now called Student's t, is due to Fisher (1925a and b) and is discussed in Eisenhart (1979). Student provided tables for t in 1925.

The paper on z was followed by another paper (1908b) in which Student tackled the small-sample distribution of the sample correlation coefficient in the normal case when the population correlation coefficient is 0. Since a mathematical derivation was beyond his powers, he decided to "fit a Pearson curve" and, using some elementary properties of correlation coefficients, he "guessed" (his own word) the correct form.

For the case that the population correlation coefficient is different from 0, he came to the conclusion that it "probably cannot be represented by any of Professor Pearson's types of frequency curves" and admits that he "cannot suggest an equation which will accord with the facts".

Gosset wrote no further papers on small-sample distributions (except for providing tables of z and t). An obvious explanation is that he had a full-time job as brewer. However, he himself denies that this was the reason, explaining to Fisher (August 14, 1924): "By the way it is not time but ability which has prevented me following up my work by more on your line."

Fisher (1939) suggests still another reason:

"Probably he felt that, had the problem really been so important as it had once seemed, the leading authorities in English statistics [i.e. Karl Pearson] would at least have given the encouragement of recommending the use of his method; and better still, would have sought again similar advantages in more complex problems. Five years, however, passed without the writers in

Biometrika, the journal in which he had published, showing any significance of his work. This weighty apathy must greatly have chilled his enthusiasm."

E. S. Pearson (1990, p. 73) explains this lack of interest by the fact that K.P.'s laboratories were not carrying out small-scale experiments.... In K.P.'s fields of interest, sound conclusions appeared to require the analysis of large-scale data. Thus, he [K.P.] jokingly remarked: 'Only naughty brewers deal in small samples!' (E. S. Pearson discusses this issue further in Pearson (1967)).

In this atmosphere Gosset's ideas might have continued to go unnoticed had they not acquired a new champion of exceptional brilliance and enormous energy.

2.2 Fisher's Proof

In 1912 R. A. Fisher, then 22 years old and a Cambridge undergraduate, was put into contact with Gosset through Fisher's Tutor, the astronomer Frederick Stratton. As a result, Gosset received from Fisher a proof of the z-distribution and asked Karl Pearson to look at it, admitting that he could not follow the argument (which was based on n-dimensional geometry) and suggesting: "It seemed to me that if it's alright perhaps you might like to put the proof in a note [in Biometrika of which K.P. was the Editor]. It's so nice and mathematical that it might appeal to some people. In any case I should be glad of your opinion of it..."

Pearson was not impressed. "I do not follow Mr. Fisher's proof and it is not the kind of proof which appeals to me", he replied. As a result, the proof was only published in 1915 together with the corresponding proof for the distribution of the correlation coefficient which Student had conjectured in his second 1908 paper. In the correlation case, the n pairs of observations are considered as the coordinates of a point in $2n$-dimensional space, in which the two sample means, two sample variances, and the sample covariance have, as Fisher writes, "a beautiful interpretation", from which the desired density can be obtained.

During the next few years, Fisher did no further work on such distributional problems, but he was pulled back to them when undertaking an investigation of the difference between the intro- and intra-class correlation coefficients. The distribution of the latter was still missing, and Fisher derived it by the same geometrical method he had used previously (Fisher, 1921).

A clue to Fisher's thinking about such problems at the time can be gleaned from his fundamental paper, "On the mathematical foundations of theoretical statistics" (1922a), which was submitted in June 1921 and read in November of that year. As the principal problems in statistics he mentions the problems of specification, estimation, and distributions. He lists the work on χ^2 by Karl Pearson and himself, the papers by Student and his own papers of 1915 and 1921 as "solving the problem of distribution" for the cases which they cover.

He continues:

(2.5) "The brevity of this list is emphasized by the absence of investigations of other important statistics, such as the regression coefficients, multiple correlations, and the correlation ratio",

and he takes up this theme again in the Summary of the paper where he states

(2.6) "In problems of Distribution relatively little progress has hitherto been made, these problems still affording a field for valuable inquiry for highly trained mathematicians."

2.3 Extensions

These two passages suggest that Fisher thought the outstanding distributional problems were difficult, and also that he had no plans to work on them himself. However, in April 1922 he received two letters from Gosset which apparently changed his mind. In the first letter Gosset urged:

(2.7) "But seriously I want to know what is the frequency distribution of $r\sigma_X/\sigma_Y$ [the regression coefficient] for small samples, in my work I want that more than the r distribution [the correlation coefficient] now happily solved."

In his later summaries of the correspondence, Fisher comments on this letter: "inquiry about the distribution of an estimated regression coefficient (a problem to which he [Gosset] presumably received the solution by return)".

This solution (together with that of the two-sample problem) appeared in JRSS (1922). The paper is primarily concerned with a different problem, that of testing the goodness of fit of regression lines. At the end Fisher appends a section which in view of the dates must have been added at the last moment and of which he later states in his Author's Note: "Section 6 takes up a second topic, connected with the first only by arising also in regression data". He introduces this second topic by explaining that

(2.8) "an exact solution of the distribution of regression coefficients... has been outstanding for many years, but the need for its solution was recently brought home to the writer by correspondence with 'Student' whose brilliant researches in 1908 form the basis of the exact solution."

A comparison of (2.8) with the statement (2.6) of a year earlier indicates the change of mind brought about by Gosset's letter. The earlier statement suggests that Fisher thought the problem was hard and that he had no intention of working on it himself. After reading Gosset's letter, he must have looked at the problem again and realized that it easily yielded to the geometric method he had used earlier; in fact so easily that he was able to send Gosset the solution "by return [mail]".

This seems to be the point at which Fisher realized the full power of his method, and the opportunity to apply this new-found confidence arose immediately. For within days there followed another request from Gosset (April 12):

> "I forgot to put another problem to you in my last letter, that of the probable error of partial $\left\{\begin{array}{c}\text{correlation}\\ \text{regression}\end{array}\right\}$ coefficients for small samples."

And Fisher in his later summary comments: "this was probably quickly answered for on May 5 he [Gosset] refers to the solution". (The result was published in 1924.)

This series of papers by Fisher on "exact" small-sample distributions culminates in two summary papers in 1924 and 1925. In the first of these he introduces the distribution of the F-statistic for testing the equality of two normal variances and points and how χ^2 and t are special cases of F. He also shows how this distribution can be used for an "analysis of variance". In the second paper he surveys the many uses to which the t-distribution can be put, and points out that this distribution applies whenever one is dealing with the ratio of two independent variables of which the numerator is normally distributed (with mean 0) and the denominator as $\sqrt{\chi^2}$. He also gives for the first time an algebraic proof of the t-distribution.

However, this is still not quite the end, for in 1928 Fisher obtained the distribution of the multiple correlation coefficient, the last "of the problems of the exact distribution of statistics in common use to have resisted solution", as he writes in the opening sentence. The paper is remarkable in that it obtains the distribution not only under the hypothesis but also in the non-central case (a term Fisher introduced here) and as a result also the non-null distribution for the analysis of variance and all the other statistics treated earlier.

The solutions of the indicated distributional problems led Fisher to some further developments, in particular the analysis of variance and the design of experiments (for details see Pearson (1939)). In these, Gosset participated but no longer in the earlier role of catalyst and we shall therefore not discuss them here.

3 Robustness

3.1 Student's Questions

The small-sample "exact" methodology discussed in the preceding section is based on the assumption of normally distributed observations. This was emphasized by Student in his first 1908 paper where he stated that

(3.1) "the conclusions are not strictly applicable to populations known not to be normally distributed; yet it appears probable that the deviation from normality must be very extreme to lead to serious error."

He was naturally curious about the effect of non-normality and in 1923 (July 3) wrote to Fisher:

> "What I should like you to do is to find a solution for some other population than a normal one. It seems to me you might assume some sort of an equation for the frequency distribution of x which would lend itself to treatment besides the Gaussian. I tried $y = a$ [i.e. the uniform distribution] once but soon got tied up.
>
> I had hoped to go on to a (right-) triangular distribution but having been defeated by I hadn't the heart to try."

Later that year (October 1), he acknowledges a reply by Fisher, unfortunately lost: "I like the result for z in the case of that horrible curve you are so fond of.[4] I take it that in skew curves the distribution of z is skew in the opposite direction."

At that time no further discussion between Gosset and Fisher on robustness is recorded. However, the problem continues to concern Gosset, so he raises it again, this time with E. S. Pearson. In response to a letter in which Pearson inquires about a different matter (which will be taken up in the next section). Gosset writes (May 11, 1926):

(3.2)

> "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 correlations between s and m [the denominator and numerator of t] with any other population distribution.
>
> Still I wish you'd tell me what happens with the even chance population [rectangular] or such as Δ [symmetrical triangular]: it's beyond my analysis."

[4] I have made various inquiries about what this distribution might have been, but without any definite answers. George Barnard has suggested the log normal as a possibility.

Pearson describes his reaction to this appeal in [Pearson (1990), p. 90]:

(3.3) "The existence of these random numbers [i.e. Tippet's table of random numbers (1927)] opened out the possibility scarcely dreamed of before, of carrying out a great variety of experimental programmes particularly of answering in considerable depth and breadth the kind of questions about the robustness of the 'normal theory' tests based on z (or t), s^2, r and χ^2 *raised by Gosset* (my italics) in his letter to me of 11 May 1926. This programme I started on in 1927 and the results began to appear, as they became available, in Biometrika papers published between 1928 and 1931."

3.2 A Fisher–Gosset Debate

As this robustness work of Pearson and his coworkers was progressing during 1926–1931, a heated argument broke out between Pearson and Fisher. Trying to mediate, Gosset was drawn into a lengthy debate with Fisher, which is of interest here because it produced several statements that clarify the views of the three participants on some aspects of the robustness question. The dispute was sparked by a critical remark that Pearson made in his 1929 review in Nature of the second edition (1928) of Fisher's book "Statistical Methods for Research Workers" and involved two issues:

(i) Whether Fisher's writing had been misleading;

(ii) How robust the normal theory tests actually are under non-normality.

Fisher, in a letter to Gosset of June 20, 1924, made it clear that he was concerned only with the first of these, but Gosset was less interested in bruised egos than in the validity of the new methods, and replied:

> (3.4) "The really important point is not your misunderstanding of Pearson, or, if there was one, his of you, but the crying practical problems of *How much does it matter*? And in fact *that* is *your* business: none of the rest of us have the slightest chance of *solving* the problem: we can play about with samples [i.e. perform simulation studies], I am not belittling E.S.P.'s work, but it is up to you to get us a proper solution."

This is a remarkable statement, particularly from Gosset whose statistical concerns are practical and who himself was a pioneer in the use of simulation. Here he makes clear the inadequacy of simulation alone and the need to supplement it by theory.

But Fisher will have none of it. In a long reply on June 27 he brushes off Gosset's suggestion

> "I do not think what you are doing with non-normal distributions is at all my business, and I doubt if it is the right approach. What I think is my business is the detailed examination of the data, and of the methods of collection, to determine what information they are fit to give, and how they should be improved to give more or other information. In this job it has never been my experience to want to make the variation more normal; I believe even in extreme cases a change of variate [i.e. a transformation] will do all that is wanted, but that of course depends on the limitation of my own experience. I have fairly often applied a z-test to crude values, and to log values, even when the translation is a severe strain, ..., but have never found it to make an important difference.
>
> Where I differ from you, I suppose, is in regarding normality as only a part of the difficulty of getting good data; viewed in this collection of difficulties I think you will see that it is one of the least important.

> You want to regard it as a part of the mathematical problem, which I do not, because a mathematical problem must start with precise data, and data other than normally [sic] are either not precise or very uninteresting."

To bring greater clarity to the issue, Gosset in his next letter makes use of a distinction that Fisher (in a letter of June 27) had introduced with respect to an Assistant of Gosset's, and which Gosset now turns around and applies to Fisher himself (June 28):

> "I think you must for the moment consent to be analysed into α-Fisher the eminent mathematician and β-Fisher the humble applier of this formulae.
>
> Now it's α-Fisher's business, or I think it is, to supply the world's needs in statistical formulae: true α-Fisher doesn't think the particular ones that I want are necessary but between ourselves that's just his cussedness. In any case I quite agree that what *we* are doing with non-normal distributions is no business of either of them; it is merely empirical whereas α-Fisher is interested in the theoretical side and β-Fisher in whatever seems good to him. But when β-Fisher says that the detailed examination of the data is his business and proceds to examine them by means of tables which are strictly true only for normally distributed variables I think I'm entitled to ask him what difference it makes if in fact the samples are not taken from this class of variables."

As a result of Gosset's intervention, Fisher did not publish his planned (apparently rather intemperate) rebuttal (which has not been preserved) to Pearson's review. Instead, at Fisher's suggestion, Gosset submitted a diplomatic response which was published in Nature (July 20). Much of it was concerned with Pearson's comment that Fisher's text could be misleading but he also addressed the substantive issue of the robustness of the t-test:

> "The question of the applcability of normal theory to non-normal material is, however, of considerable importance and merits attention both from the mathematician and from those of us whose province it is to apply the results of his labours to practical work. Personally, I have always believed, without perhaps any very definite grounds for this belief, that in point of fact, 'Student's' distribution will be found to be very little affected by the sort of small departures from normality which obtain in most biological and experimental work, and recent work on small samples confirms me in this belief. We should all of us, however, be grateful to Dr. Fisher if he would show us elsewhere[5]on theoretical grounds what *sort* of modification of his tables we require to make when the samples with which we are working are drawn from populations which are neither symmetrical nor mesokurtic [i.e. whose coefficient of kurtosis is not zero]."

Fisher was not willing to leave Gosset's challenge unanswered and published a reply in Nature (August 17) which, as he mentions in a letter to Gosset "seems free from Billingsgate [i.e. abusive language] and may even help members and others to understand better where we stand". In this reply Fisher does not address the question of robustness of the tests but in response to the last sentence of Gosset's letter considers alternatives which would avoid Pearson's criticism. Two comments are of particular interest.

In the first of these he considers what would happen if it were possible to generalize the normal-theory distributions and discusses the criticisms to which such an extension would be exposed. The most interesting of these is:

[5]The word "elsewhere" had been inserted by the Editor and greatly annoyed Fisher.

> "that the particular statistics, means and mean squares entering into these tests are only efficient for normal distributions and that for Pearson curves quite other statisics are required, and not only revised distributions of the familiar statistics appropriate to normal material."

This statement is of course correct and interesting in light of the later Neyman–Pearson theory.

Fisher makes another interesting suggestion in the next paragraph of his letter:

> "Beyond all questions of metrical variates there are, largely undeveloped, a system of tests which depend only on the frequency and an order of magnitude. Examples exist in 'Student's' writing and in my own. They are free from all taints of normality, but are too insensitive to be very useful; still, their development would be of more interest than the programme of research first considered."

These two comments show the enormous breadth of Fisher's vision. They foreshadow two of the most significant later developments of the small-sample approach: the Neyman–Pearson theory and the nonparametric methodology of rank tests.

While the Fisher–Gosset debate concerning the robustness (against non-normality) of the t-test and the tests Fisher had developed in its wake brought no meeting of minds, some clarification was achieved by Pearson's simulation studies. They indicated that the t-test and those of the model I analysis of variance are fairly insensitive under non-normality, but that this is not true for the F-test for variances or some of the tests for variance components. These suggestions were later confirmed by theoretical results of George Box and others as well as by additional simulation work. Since the vulnerable F-tests were included in "Statistical Methods" without any warning about their unreliability, Gosset's insistence on verification seems justified.

4 Choice of Test

On the robustness question Gosset clearly was the driving force. It was his suggestion that led to Pearson's empirical investigations, and he tried repeatedly, though unsuccessfully, to get Fisher to study the issue theoretically.

At the next stage which led to the Neyman–Pearson theory, Gosset's role was quite different. As Pearson recalled the origin of this development in his obituary of Gosset [ESP (1929)]:

> "I had been trying to discover some principle beyond that of practical expediency which would justify the use of Student's z."

He addressed some of his questions in a letter to Gosset and

> (4.1) "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 explains the observed events with a greater degree of probability."

Pearson goes on to quote the relevant paragraph of Gosset's letter of May 11, 1926. The crucial point which was to have such far reaching consequences is contained in a single sentence. Speaking about the observation of a very unlikely event which has probability, say, .0001, Gosset writes:

(4.2) "that doesn't in itself necessarily prove that the sample is not drawn randomly from the population [specified by the hypothesis]: 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."

Pearson passed the suggestion on to Neyman who was spending the year in Paris and who acknowledged it in a letter of December 6, 1926, agreeing that "to have the possibility of testing, it is necessary to adopt such a principle as Student's" [quoted in Reid, p. 70].

Gosset's suggestion led Pearson to the idea of likelihood ratio tests as a reasonable method of test construction, and the result was a pair of joint papers by Neyman and Pearson in the 1928 volume of Biometrika: "On the use and interpretation of certain test criteria for purposes of statistical inference" which together took up 98 pages. In it the authors introduced not only the likelihood ratio principle, but also the concept of 1$^{\text{st}}$ and 2$^{\text{nd}}$ kinds of error. The formulation of both ideas required not only the hypothesis H but also a class of alternatives to H.

Neyman and Pearson followed the likelihood ratio paper with an attack from first principles on how to choose a test not only in the normal case but quite generally. This work resulted in their 1933 paper "On the problem of the most efficient tests of statistical hypothesis", which formed the basis of the theory of hypothesis testing as we now know it. The approach made use not only of the class of alternatives suggested by Student, but also of an innovation introduced by Fisher in his Statistical Methods book, namely to define significance in terms of a preassigned level instead of reporting p-values. This proposal (which later was overused and as a result attracted strong opposition) has a curious connection with Gosset.

A few years after starting to work for Guinness and learning the statistics he needed from Airy and Merriman, Gosset in 1905 (before he became "Student") made an appointment to consult Karl Pearson about some questions he was unable to resolve. In a report to the brewery (see Pearson (1939)) he states as one of these questions that

> "none of our books mentions the odds, which are conveniently accepted as being sufficient to establish any conclusion."

Had Gosset addressed this question twenty years later to Fisher, we might credit (or blame) him for having suggested the idea of fixed levels, a concept which constituted a crucial element of the Neyman–Pearson formulation.

5 Conclusion

Let us now return to the question posed at the beginning of this paper regarding the influence of Gosset on Fisher, Pearson and Neyman.

5.1 Gosset and Pearson

Pearson's contributions to small sample theory are twofold. They consist on the one hand of his simulation studies of robustness culminating in his 1931 paper "The analysis of variance in cases of non-normal variation". The other strand is his joint work with Neyman in which they developed what is now called the Neyman–Pearson theory. For both aspects, the crucial ideas came from Student. As Pearson himself acknowledges (Pearson, 1990, p. 82) in comment on Student's letter from 1926

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

Thus, with respect to Pearson, F. N. David's assessment seems essentially correct: The main ideas leading to Pearson's research were indeed provided by Student.

5.2 Gosset and Fisher

A corresponding conclusion does not apply to Gosset's influence on Fisher. It is of course true that the central idea of Gosset's 1908 papers—the possibility of determining the exact distribution of various statistics by assuming a known (for example normal) underlying distribution—provided the inspiration, and their mathematical incompleteness the opportunity, for Fisher's basic 1915 paper. And in addition Gosset provided the impetus for Fisher's later distributional work by urging him to determine the distribution of regression and correlation coefficients.

However, Fisher also made many highly original and influential contributions to small-sample theory which owed nothing to Student, such as variancestabilizing and normalizing transformation, permutation tests, the design of experiments (including randomization), the concepts of sufficiency and of likelihood.

5.3 Gosset and Neyman

Although no conversations or correspondence between Neyman and Student are reported, Student of course affected Neyman indirectly through his influence on Pearson. However, Neyman's and Pearson's recollections of the origin of the Neyman–Pearson theory are at variance. While Pearson attributed the basic idea leading to their work to Student's letter of 1926, Neyman cites

> "remarks of Borel that served as an inspiration to Egon S. Pearson and myself in our effort to build a frequentist theory of testing hypothesis." (Neyman, 1977)

Neyman's first reference to Borel, who incidentally does not mention alternatives explicitly, occurs in Neyman (1929). The discrepancy between these two different views has been discussed in some detail in Lehmann (1993). Pearson's recollections are quite specific and since he was clearly the leader in the work up to 1928 (this was to change later), it seems fair to consider Student's contribution as the decisive one.

5.4 Concluding Remark

This paper has a limited aim: To assess the contributions made by Student to the three stages of small-sample theory listed at the end of Section 1. Thus, in particular, it is not an account of Student's work as a whole and does not cover his remaining papers, the wealth of ideas and suggestions contained in his correspondence, and his work for Guinness both as a statistician and a brewer.

Such a more comprehensive account would, I believe, support the statement made by Fisher (1939) in his obituary of Student in which he describes Gosset as "one of the most original minds in contemporary science".

Acknowledgement of Sources

The primary sources for Gosset are not only his papers [Student's collected papers (1942), E. S. Pearson and J. Wishart (Eds.)] but also his correspondence of which Pearson (1939) says: "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". Gosset's biographer McMullen (1939) notes: "'Student' had many correspondents, mostly agricultural and other experimenters, in different parts of the world. He took immense pains with these... [and they] contain some of his clearest writing...".

A discussion of the letters exchanged between Gosset and Pearson (including some important extracts) can be found in Pearson (1939) and (1990).

Some early correspondence between Gosset and Fisher is discussed in Pearson (1968), and the bulk of the surviving letters from 1915 (when Fisher was 25) to shortly before Gosset's death in 1937 were privately published by Gosset's lifelong employer, the firm of Arthur Guinness Sons and Co. Unfortunately most of Fisher's letters are lost although there are some important exceptions. On the other hand, the volume includes valuable summaries of the letters (sometimes with added comments) which Fisher made at a later date. Some of the most important passages from the letters are reprinted, with comments, in Pearson (1990).

In addition to the papers and letters, there exists a wealth of secondary literature. A long obituary “W. S. Gosset, 1876–1937”, was published in the 1939 volume of Biometrika in two parts: “‘Student’ as a man” by Gosset’s brewery colleague L. McMullen, and “‘Student’ as a statistician” by E. S. Pearson.

There are other obituaries, book chapters, and so on, but we mention here only the most important of these: the book “‘Student’ – statistical biography of William Sealy Gosset” (1990) which was “edited and augmented” from an incomplete manuscript of E. S. Pearson’s, by R. L. Plackett with the assistance of G. A. Barnard. This book contains many additional references.

For Fisher, Pearson and Neyman we again have available convenient collections of their papers: The five volumes of Fisher’s papers, many with helpful later “Author’s Notes”, edited by Bennett and published by the University of Adelaide, and three volumes, one each of papers by Pearson, Neyman, and Neyman–Pearson, published by the University of California Press. Of the secondary literature we mention only the biographies of Fisher by his daughter Joan Fisher Box and of Neyman by Constance Reid.

References

Airy, G. B. (1861, 1875, 1879). On the Algebraic and Numerical Theory of Errors of Observations and the Combination of Observations. MacMillan. London.

Box, J. F. (1978). R. A. Fisher: The Life of a Scientist. Wiley. New York.

Box, J. F. (1981). Gosset, Fisher and the t-distribution. American Statistician **35**, 61–66.

Box, J. F. (1987). Guinness, Gosset, Fisher, and small samples. Statist. Science **2**, 45–52.

Eisenhart, C. (1979). On the transition from "Student's" z to "Student's" t. American Statistician **33**, 6–10.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika **10**, 507–521.

Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. Metron **1**, 3–32.

Fisher, R. A. (1922a). On the mathematical foundations of theoretical statistics. Phil. Trans. Royal Soc. London A **222**, 309–368.

Fisher, R. A. (1922b). The goodness of fit of regression formulae, and the distribution of regression coefficients. J. Roy. Statist. Soc. **85**, 597–612.

Fisher, R. A. (1924a). The distribution of the partial correlation coefficient. Metron **3**, 329–332.

Fisher, R. A. (1924b). On a distribution yielding the error functions of several well known statistics. Proc. Intern. Congress of Mathematics, Toronto, **2**, 805–813.

Fisher, R. A. (1925a). Applications of 'Student's' distribution. Metron **5**, 90–104.

Fisher, R. A. (1925b). Statistical Methods for Research Workers. Oliver and Boyd. Edinburgh (2nd Ed. 1928).

Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. Proc. Roy. Soc. London, A, **121**, 654–673.

Fisher, R. A. (1929). Statistics and biological research. Nature **124**, 266–267.

Fisher, R. A. (1939). "Student". Ann. Engenics **9**, 1–9.

Fisher, R. A. (1971–1974). Collected Papers of R. A. Fisher (J. H. Bennett, Ed.). University of Adelaide.

Gosset, W. S. (1970). Letters from W. S. Gosset to R. A. Fisher, 1915–1936 With Summaries by R. A. Fisher and a Foreword by L. McMullen. Printed for private circulation.

Hald, A. (1998). A History of Mathematical Statistics (from 1750–1930). Wiley. New York.

Laird, N. M. (1989). A conversation with F. N. David. Statist. Sci. **4**, 235–246.

Lehmann, E. L. (1993). The Bertrand–Borel debate and the origins of the Neyman–Pearson theory. In Statistics and Probability: A Raghu Raj Bahadur Festschrift (J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy and B. L. S. Prakasa Rao, Eds.). Wiley Eastern Limited. New Delhi.

McMullen, L. (1939). "Student" as a man. Biometrika **30**, 205–210.

Merriman, M. (1884). A Textbook on the Method of Least Squares. Wiley. New York.

Mosteller, F. and Tukey, J. W. (1977). Data Analysis and Regression. Addison–Wesley. Reading.

Neyman, J. (1929). Méthodes nouvelles de vérification des hypothèses. Comptes Rendus du I Congrès de Mathématiciens des Pays Slaves, Warsaw, 355–366.

Neyman, J. (1967). A Selection of Early Statistical Papers of J. Neyman. University of California Press. Berkeley.

Neyman, J. (1977). Frequentist probability and frequentist statistics. Synthèse **36**, 97–131.

Neyman, J. and Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. Biometrika **20A**, 175–240 and 263–294.

Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. Phil. Trans. Roy. Soc. A **231**, 289–337.

Neyman, J. and Pearson, E. S. (1967). Joint Statistical Papers of J. Neyman and E. S. Pearson. University of California Press. Berkeley.

Pearson, E. S. (1929a). Review of R. A. Fisher: Statistical Methods for Research Workers (2nd Ed). Nature.

Pearson, E. S. (1929b). Statistics and biological research. Nature **124**, p. 615.

Pearson, E. S. (1931). The analysis of variance in cases of non-normal variation. Biometrika **23**, 114–133.

Pearson, E. S. (1939). 'Student' as a statistician. Biometrika **30**, 210–250.

Pearson, E. S. (1966a). The Selected Papers of E. S. Pearson. University of California Press. Berkeley.

Pearson, E. S. (1966b). The Neyman–Pearson Story: 1926–1934. In Research Papers in Statistics. Festschrift for J. Neyman (F. N. David, Ed.), 1–23. Wiley. London.

Pearson, E. S. (1967). Some reflexions on continuity in the development of mathematical statistics, 1885–1920. Biometrika **54**, 341–355.

Pearson, E. S. (1968). Some early correspondence between W. S Gosset, R. A. Fisher and Karl Pearson with notes and comments. Biometrika **55**, 445–457.

Pearson, E. S. (1990). 'Student' – A Statistical Biography of William Sealy Gosset (R. L. Plackett and G. A. Barnard, Eds.). Clarendon Press. Oxford.

Reid, C. (1982). Neyman – From Life. Springer–Verlag. New York.

Student (1908a). The probable error of a mean. Biometrika **VI**, 1–25.

Student (1908b). Probable error of a correlation coefficient. Biometrika **VI**, 302–310.

Student (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 **XI**, 414–417.

Student (1925). New table for testing the significance of observations. Metron **5**, 105–108.

Student (1929). Statistics in biological research. Nature **124**, p. 93.

Tankard, J. W., Jr. (1984). The Statistical Pioneers. Schenkman Publishing Co. Cambridge.

Tippett, L. H. C. (1927). Random Sampling Numbers. Tracts for Computers **15**. Cambridge University Press. Cambridge.

Welch, B. L. (1958). 'Student' and small sample theory. J. Amer. Statist. Assoc. **53**, 777–788.