Chapter 2
Fisher's Testing Methodology

2.1 The Small-Sample and χ^2 Papers

We saw in Sect. 1.3 that Student in 1908a brought a new point of view to statistical inference by determining the small-sample (exact) distribution of what he called z, now called t, under the assumption of normality. Student found the correct form of this distribution but was not able to prove it.

In 1912 Ronald Fisher, then still a Cambridge undergraduate, sent Gosset a proof using n-dimensional geometry. Gosset, stating that he did not "feel at home in more than three dimensions," sent it on to Karl Pearson, with the suggestion to publish it as a note in *Biometrika*.

However, Pearson replied that, "I do not follow Mr. Fisher's proof and it is not the kind of proof which appeals to me. … Of course, if Mr. Fisher will write a proof in which each line flows from the preceding one and define his terms, I will gladly consider its publication. Of the present proof I can make no sense."

Despite this negative reaction, three years later Pearson did accept a paper in which Fisher used his geometric method to derive the small-sample distribution of the correlation coefficient from a bivariate normal distribution (and in passing also that of Student's t) (1915) [4]. For the case of zero correlation, this distribution had been correctly conjectured (but again not proved) by Student in a second *Biometrika* paper, also in 1908 (Student, 1908b).

After 1915, for a number of years Fisher did no further work on such distributional problems, but he was pulled back to them when investigating the difference between the inter- and intraclass correlation coefficients. The distribution of the latter was still missing and Fisher derived it by the same geometric method he had used previously (Fisher, 1921) [14].

An idea of Fisher's thinking at the time can be gathered from his 1922 [18] paper discussed in Sect. 1.5, where he mentions the "absence of investigation of other important statistics, such as regression coefficients, multiple correlations and the correlation ratio." He refers to these problems in the summary of the paper as "still affording a field for valuable enquiry by highly trained mathematicians."

E.L. Lehmann, *Fisher, Neyman, and the Creation of Classical Statistics*,
DOI 10.1007/978-1-4419-9500-1_2,

This passage suggests that Fisher thought these problems to be difficult, and that he had no plans to work on them himself. However, in April 1922 [20] he received two letters from Gosset that apparently changed his mind. We are lucky that these and subsequent letters from Gosset to Fisher (but unfortunately few of Fisher's replies) were preserved and in 1970 were privately published by Gosset's employer, the firm Arthur Guinness Son and Co. (Gosset, 1970).

In the first of the letters mentioned, Gosset pleaded: "But seriously, I want to know what is the frequency distribution of $r\,\sigma_x/\sigma_y$ for small samples, in my work I want that more than the r distribution now happily solved."

Fisher's solution to this problem (together with that of the two-sample problem) appeared in the *Journal of the Royal Statistical Society* in 1922 [20]. 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. Fisher acknowledges in this section that:

> 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.

The solution to Gosset's problem turned out to be surprisingly easy. It consisted essentially in showing that the regression coefficients had the same structure as Student's t, and hence were distributed according to Student's distribution, however with the degrees of freedom in the denominator reduced by one in the case of simple linear regression, and by more if more than one variable is involved. The argument was so simple that Fisher was able to send it to Gosset by return mail.

Gosset's second letter followed within a few days. It was a short note, saying:

> I forgot to put another problem to you in my last letter, that of the probable error of the partial $\left\{\frac{\text{Correlation}}{\text{Regression}}\right\}$ coefficients in small samples.

Using once more the geometric method, Fisher was able to reduce the partial correlation to the total correlation based on a smaller sample. He sent this solution to Gosset without much delay and published it in a short note in 1924 [35].

In the same year, Fisher gave a lecture[1] at the International Congress of Mathematics in Toronto, "On a distribution yielding the error functions of several well-known statistics" [36]. The distribution in question, which Fisher called the general z-distribution, is equivalent to that of the ratio of two independent χ^2-based estimates of a common variance, required for the analysis of variance. This talk and a 1925 paper on "Applications of 'Student's' distribution" outlined the new testing methodology Fisher had developed during the preceding decade and of which he wrote an account in a book that will be discussed in the next sections.

The small-sample papers reviewed in the present section constitute the second series mentioned in Sect.1.6. It should be pointed out that between 1922 and 1924,

[1] The Proceedings were not published until 1928.

Fisher also published a third series consisting of papers 19, 31, and 34 on χ^2-tests of goodness of fit. In the first of these he corrected Pearson’s error regarding degrees of freedom[2] and noted that his proof applies to χ^2-tests in “all cases in which the frequencies observed are connected with those expected by a number of relations, beyond their restriction to the same total frequency.”

The next paper is concerned with the special case of a 2×2 table where Fisher shows that when the margins are fixed so that the whole table is determined by the entry in any one of the four cells, the χ^2-statistic for testing independence has a limiting χ^2-distribution with one degree of freedom, rather than three as Pearson had claimed.

Finally, paper 34 of 1924 deals with goodness of fit tests in which some parameters have to be estimated. Here, Fisher finds that the χ^2-distribution with the appropriate number of degrees of freedom is valid provided the parameters are estimated efficiently, and hence in particular when the estimators used are either maximum likelihood or minimum chi-squared.

2.2 “Statistical Methods for Research Workers” I

The impulse to write a book on the statistical methodology he had developed came not from Fisher himself but from D. Ward Cutler, one of the two editors of a series of “Biological Monographs and Manuals” being published by Oliver and Boyd. We owe this information to a letter Fisher wrote in 1950 to Robert Grant, then a director of Oliver and Boyd (Bennett, 1990, pp. 317–318.) In this letter, he also points out that writing the book did not require new research but, “I only had to select and work out in expository detail the examples of the different methods proposed. It was often quite a job to find a good example.”

The book came out in 1925 with the title, “Statistical Methods for Research Workers” (SMRW), and it revolutionized the practice of statistics. The opening introductory chapter lays out Fisher’s vision of statistics. He states that

> statistics may be regarded as (i) the study of populations, (ii) as the study of variation, (iii) as the study of methods of the reduction of data.

In the latter category, he again mentions, as he had done in his 1922 [18] foundational paper, that the problems arising in this reduction can be divided into three types: problems of specification, estimation, and distribution.

He goes on to discuss the role of probability in this work, and to attack the method of “inverse probability” (what today we would call the Bayesian approach), stating his “personal conviction…that the theory of inverse probability is founded upon an error, and must be wholly rejected.” As an alternative, he suggests that “the

[2] A careful analysis of Pearson’s evolving progress on this issue is given in Chap. 19, “Karl Pearson and degrees of freedom,” of Stigler (1999).

mathematical quantity which appears to be appropriate for measuring our preference among different possible populations" is that of likelihood.

Next, he defines a number of concepts of the theory of estimation, such as consistency, efficiency, sufficiency, and maximum likelihood.

It comes as somewhat of a surprise when the next section of the introduction entitled "the scope of this book" states that

> the prime object of this book is to put into the hands of research workers…the means of applying statistical tests accurately to numerical data accumulated in their own laboratories…

and later refers to

> the exact distributions with the use of which this book is chiefly concerned…

Thus, the book does not primarily deal with estimation but with significance testing. In fact, estimation is never again mentioned.

Under these circumstances, one wonders why the introduction mentions distributions as the second principal concern of statistics rather than testing. As was already indicated in Sect. 1.6, this is actually quite reasonable. Unlike estimation, where the choice of an appropriate estimate was the central problem, the corresponding issue for testing did not arise for Fisher. Assuming normality, the choice of test statistic was intuitively obvious to him, and once its distribution had been obtained, the test consisted simply of declaring significance for observations in the appropriate tail of the distribution. Thus, the only serious problem was that of distribution, i.e., of determining the distribution of the test statistic.

The introduction concludes with a brief discussion of the tables "which form a part essential to the use of the book."

The introductory chapter is followed by two preparatory chapters on diagrams and distributions respectively. The latter introduces the reader to various aspects of three basic distributions: the normal, Poisson, and binomial.

Chapter IV deals with χ^2-tests of goodness of fit, independence, and homogeneity. After reminding the reader that such tests were already used in the preceding chapter to test the goodness of fit of Poisson and binomial distributions, Fisher states that the general class of problems for which it is appropriate is "the comparison of the number actually observed to fall into any number of classes with the numbers which upon some hypothesis are expected." The proposed test statistic, denoted by χ^2, is the sum over all the classes of x^2/m where "m is the number expected and $m + x$ the number observed in any class."

Fisher points out that "the more closely the observed numbers agree with those expected the smaller χ^2 will be; in order to utilize the table it is necessary to know the value of n with which the table is to be entered. The rule for finding n is that n is equal to the number of degrees of freedom in which the observed series may differ from the hypothetical; in other words, it is equal to the number of classes the frequencies in which may be filled up arbitrarily." This definition, which an uninitiated reader may have found somewhat cryptic, is made clearer through a number of examples.

There still remains an important piece of business (relating to the table) to deal with before the tests are illustrated by means of seven examples covering twelve pages. A table of the χ^2-distribution had been published in 1902 in *Biometrika*, but it was protected by copyright and Fisher could not get permission to reprint it. So instead of giving a table of the probability[3]

$$P = P(\chi^2 \geq x),$$

he tabled x as a function of P for selected values of P between 0.99 and 0.01. He comments that:

> In preparing this table we have borne in mind that in practice we do not want to know the exact value of P for any observed χ^2, but, in the first place, whether or not the observed value is open to suspicion. If P is between .1 and .9 there is certainly no reason to suspect the hypothesis tested. If it is below .02 it is strongly indicated that the hypothesis fails to account for the whole of the facts. We shall not often be astray if we draw a conventional line at .05 and consider that higher values of χ^2 indicate a real discrepancy.

This statement has been quoted in full because of its great influence. Fisher's recommendation of 5% as a fixed standard took hold and, for good or ill, has permeated statistical practice. Its advantages and disadvantages have been much discussed and will be compared with a more flexible alternative approach in Chap. 3.

Fisher gives further support to the proposal in the examples which illustrate the χ^2-test in a number of representative situations. In the first of these, in particular, he finds a p-value between 0.01 and 0.02 and concludes: "If we take $P=0.05$ as the limit of significant deviation, we shall say that in this case the deviations from expectation are significant."

The χ^2-chapter concludes with a section on the partitioning of χ^2 into components "to test the separate components of a discrepancy." This foreshadows the analysis of variance treated in the last chapter of the book.

2.3 "Statistical Methods for Research Workers" II

The four chapters described so far make up nearly half of SMRW. The second half is concerned with the "exact" small-sample tests which are the avowed purpose of the book. Chapter V in particular deals with the problems for which the t-distribution provides the solution: tests of significance of means, differences of means, and regression coefficients.

Fisher, however, starts the chapter with a brief discussion of the large-sample tests for a mean or a difference of two means, pointing out that these procedures are what today we call distribution-free, i.e., valid for all distributions with finite variance.

Only then does he consider the small-sample t-test, stating that

> The distribution of t for random samples of a normal population about zero as mean, is given in the table of t.

[3] Today such p-values are usually denoted by a lower-case rather than capital p.

After illustrating the procedure on a sample of ten patients that had already been used by Student in 1908, he next discusses the "significance of difference of means of small samples." This case is of particular interest since it is one of the few in which today's approach differs from that presented by Fisher.

The heading of Fisher's treatment of the two-sample problem reads:

(*) To test whether two samples belong to the same population, or differ significantly in their means.

The test he proposes for this problem is the two-sample t-test.

However, the dichotomy (*) is clearly incomplete. Realizing this, Fisher in the second (1928) edition of the book adds the following paragraph:

> It will be noted…that a difference in variance between the populations from which the samples are drawn will tend somewhat to enhance the value of t obtained.[4] The test, therefore, is decisive, if the value of t is significant, in showing that the samples could not have been drawn from the same population; but it might conceivably be claimed that the difference indicated lay in the variances and not in the means. The theoretical possibility, that a significant value of t should be produced by a difference between the variances only, seems to be unimportant…; as a supplementary test, however, the significance of the difference between the variances may be tested directly by the method of § 41.

In still later editions, Fisher adds yet another paragraph:

> It has been repeatedly stated, perhaps through a misreading of the last paragraph, that our method involves the "assumption" that the two variances are equal. This is an incorrect form of statement; the equality of variances is a necessary part of the hypothesis to be tested, namely that the two samples are drawn from the same normal distribution. The validity of the t-test is therefore absolute, and requires no assumption whatever.

Fisher concludes this later discussion by pointing out that one could of course ask the question: "Might these samples have been drawn from different normal populations having the same mean?" He states that this problem has been solved (it is what is today known as the Behrens-Fisher problem), but that "the question seems somewhat academic."

Despite Fisher's protest, the modern view considers the testing of the difference of two means in two versions:

(i) Assuming nothing about the two variances (the Behrens-Fisher problem) or
(ii) Assuming the two variances to be equal, in which case the t-test is the appropriate (and by a number of criteria the best possible) procedure.

This disagreement serves to emphasize the fact that for most of the procedures set out in SMRW, Fisher's way of seeing and doing things still holds sway today. It also illustrates that in an argument Fisher rarely gave an inch. Those holding views different from his own had "misread" him and their statements were "incorrect." We shall see this attitude repeated in later controversies.

The remainder of Chapter V is concerned with regression. It is shown how the t-test also applies to the testing of regression coefficients in linear regression. Fisher then

---

[4] This statement is correct for some balanced designs but is not correct in general; see, for example, Scheffé (1959, p. 353).

turns to the fitting of "curved regression lines" for the case "when the variability of the independent variate is the same for all values of the dependent variate, and is normal for each such value." The chapter concludes with the testing of partial regression coefficients, where the distribution of the test statistic is again found to be Student's t.

Chapter VI has a somewhat narrower focus: correlation coefficients. It defines the correlation coefficient ρ in a bivariate normal distribution and discusses its estimation, and from there progresses to partial correlation. It next takes up the significance of an observed correlation coefficient r and states that when $\rho=0$, then

$$t = r\sqrt{n-2}\Big/\sqrt{1-r^2},$$

has a t-distribution with $n-2$ degrees of freedom, and extends this result to partial correlation coefficients.

Having thus provided the means for testing that these coefficients are zero, Fisher turns to the problem of testing $\rho=\rho_0$ for $\rho_0\neq 0$. For this purpose, he proposes the transformation

$$z = \tfrac{1}{2}\log\,[(1+r)/(1-r)]$$

and states that z is approximately normally distributed with mean ρ and standard error $1/\sqrt{n-3}$. He cautions that

> The distribution of z is not strictly normal, but it tends to normality rapidly as the sample is increased, whatever may be the value of the correlation.

After considerable further discussion, Fisher shows in two examples that the approximation is reasonably accurate. As an additional application, he points out that this transformation also makes it possible to test the difference between two observed correlations.

Chapter VII extends these considerations to the case of intraclass correlations, but it does so apologetically. It starts out with the admission that the data to be analyzed in this chapter by means of correlations "may be more usefully and accurately treated by the analysis of variance, that is by the separation of the variance ascribable to one group of causes, from the variance ascribed to other groups."

And again, after discussing and illustrating a generalized z-transformation of the intraclass correlation coefficient, Fisher states that "a very great simplification is introduced into questions involving intraclass correlation when we recognize that in such cases the correlation merely measures the relative importance of two groups of factors causing variation."

Still later, he points out that "the data provide us with independent estimates of two variances; if these variances are equal the correlation is zero;... . If, however, they are significantly different, we may if we choose express the fact in terms of a correlation."

The final sections of the chapter then at last deal directly with tests which arise in the analysis of variance. "The test of significance of intraclass correlations is thus simply," Fisher writes, "an example of the much wider class of tests of significance which arise in the analysis of variance. These tests are all reducible to the single problem of testing whether one estimate of variance derived from n_1 degrees of freedom is significantly greater than a second such estimate derived from n_2 degrees of freedom. This problem is reduced to its simplest form by calculating z equal to half the difference of the logarithms of the estimates of the variance." Fisher provides a table for $P = 0.05$ and selected values of n_1 and n_2.

The remainder of the chapter illustrates the use of the table and contains a brief discussion of "analysis of variance into more than two portions."

The last chapter (Chapter VIII) consists of applications of the analysis of variance to a few important situations. The first half of the chapter is concerned with testing the structure of regression lines; this includes a discussion of the multiple correlation coefficient, and is an extension of earlier material.

The second half strikes out in a new direction, resulting from Fisher's work at Rothamsted. As he writes, "The statistical procedure of the analysis of variance is essential to an understanding of the principles underlying modern methods of arranging field experiments. The first requirement which governs all well-planned experiments is that the experiment should yield not only a comparison of different manures, treatments, varieties, etc., but also a means of testing the significance of such differences as are observed."

In the last pages of the book, he illustrates such analyses first on the case that the agricultural plots are assigned to the different treatments (with replication) completely at random, then into randomized blocks, and finally into Latin squares.

The expansion of this program was to be one of Fisher's most important tasks during the next decade.

2.4 The Reception of SMRW

Fisher was greatly disappointed in the reviews of his book. The most important British journal of statistics publishing book reviews was the *Journal of the Royal Statistical Society (JRSS)*, and its January issue of 1926 carried a review of Fisher's book signed L. I. The reviewer was Leon Isserlis, a statistician at the Chamber of Shipping, with a strong Cambridge mathematics background and graduate studies under Karl Pearson.

Isserlis' review was not very enthusiastic: "We have presented [here]," he wrote, "a very full account of the statistical methods favored by the author and of the conclusions he has reached on topics, some of which are still in the controversial stage. Much is lacking if the book is to be regarded as an authoritative record of achievement in statistical method apart from Mr. Fisher's own contributions."

After briefly describing the content of the book, the review reaches the conclusion that

> The book will undoubtedly prove of great value to research workers whose statistical series necessarily consist of small samples, but will prove a hard nut to crack for biologists who attempt to use it as a first introduction to statistical method."

If Isserlis was not overly enthusiastic, Fisher himself must bear part of the blame. In the author's preface, he asserted:

> Little experience is sufficient to show that the traditional machinery of statistical processes is wholly unsuited to the needs of practical research. Not only does it take a cannon to shoot a sparrow, but it misses the sparrow.

Such wholesale and insulting dismissal of traditional procedures can hardly have endeared him to the British establishment.

However, even reviewers who were not offended by Fisher's attack on traditional methods found much to criticize. In particular, they complained about Fisher's dogmatism, the lack of proofs, the emphasis on small samples, and the difficulty of the book.

The only review that did justice to Fisher's achievement came not from Great Britain but from America. The reviewer was Harold Hotelling (1927), who had obtained his Ph.D. in 1924 from Princeton in topology, and who then took up a position as research associate at the Food Research Institute at Stanford University. He was so impressed with the book that he submitted a review to the *Journal of the American Statistical Association* without having been asked to do so.

He opens the review by stating that

> Most books on statistics consist of pedagogic rehashes of identical material. This comfortably orthodox subject matter is absent from the volume under review, which summarizes for the mathematical reader the author's independent codification of statistical theory and some of his brilliant contributions to the subject, not all of which have previously been published.

After some discussion of the content of the book, the review concludes with:

> The author's work is of revolutionary importance and should be far better known in this country.

The first edition was soon sold out, and in 1928 Fisher brought out a second edition. He was pleased with this success and considered it a vindication. He gave expression to his feelings in the preface to the new edition, when he wrote:

> The early demand for a new edition has more than justified the author's hope that use could be made of a book which, without entering into the mathematical theory of statistical methods, should embody the latest results of that theory, presenting them in the form of practical procedures appropriate to those types of data with which research workers are actually concerned.

He goes on to defend his much-criticized decision not to include proofs by pointing out that, "The practical application of general theorems is a different art from their establishment by mathematical proof, and one useful to many to whom the other is unnecessary."

The remainder of the preface outlines the changes from the first edition. The most important of these is the addition of a new chapter on "The Principles of Statistical Estimation." The presentation is very unusual. Fisher illustrates the concepts of consistency, efficiency, relative efficiency, and so on by means of a single example: a contingency table arising in genetics, in which all cell probabilities are expressible in terms of a single parameter θ.

Before discussing the estimation of θ, Fisher notes that, "It is a useful preliminary before making a statistical estimate…to test if there is anything to justify estimation at all." He therefore tests for independence (which would specify the value of θ) and rejects the hypothesis. He then points out that, "Nothing is easier than to invent methods of estimation," and proposes to consider four estimates of θ (T_1 to T_4), the fourth being the maximum likelihood estimate.

The first two are linear functions of the cell frequencies, and Fisher shows how to calculate their variances. The other two are more complicated, so he provides large-sample approximations to their variances. These asymptotic variances for T_3 and T_4 turn out to be equal, and of this common value he writes that "it is of great importance for our problem, for it has been proved that no statistic can have a smaller variance, in the theory of large samples, than has the solution of the equation of maximum likelihood. This group of statistics…are therefore of particular value, and are designated *efficient* statistics."

He goes on to consider the reciprocal of this minimum asymptotic variance "as a numerical measure of the total amount of information, relevant to the value of θ, which the sample contains." He further states that "the actual fraction utilized by inefficient statistics in large samples is obtained by expressing the random sampling variance of efficient statistics as a fraction of that of the statistic in question," and exemplifies the calculation on the estimates of T_1 and T_2.

The review of the second edition of SMRW in the *Journal of the Royal Statistical Society* was considerably more favorable than that of the first, and refers to Fisher's contributions as having "already had, and are still more likely to have in the future, a far-reaching influence on the subject." The review must, however, be read with some caution for possible bias, since the reviewer, J.O. Irwin (1929), although a student of Karl Pearson, was at the time of the review a member of Fisher's department at Rothamsted.

As he had done for the first edition, Hotelling also volunteered a review of the second edition (in fact for each of the first seven editions). By that time he was an associate professor in the Stanford Mathematics Department.

He refers to the book as "this unique work" and concludes with the advice that

> An American using statistics will do well to work through the book, translate it into his own tongue, and change his habits accordingly.

A third review by Pearson (1929) led to a controversy which will be considered in the next section.

2.5 The Assumption of Normality

Egon Pearson reviewed the second edition of Fisher's book in the 8 June 1929 issue of *Nature*. The review was more positive than an earlier review he had written of the first edition, but it also contained the following critical paragraph:

> There is one criticism, however, which must be made from the statistical point of view. A large number of the tests developed are based…on the assumption that the

> population sampled is of the "normal" form. That this is the case may be gathered from a careful reading of the text, but the point is not sufficiently emphasized. It does not appear reasonable to lay stress on the "exactness" of the tests when no means whatever are given of appreciating how rapidly they become inexact as the population sampled diverges from normality. That the tests, for example, connected with the analysis of variance are far more dependent on normality than those involving Student's z (or t) distribution is almost certain, but no clear indication of the need for caution in their application is given.

Fisher was deeply offended; he felt that his honesty had been impugned, and he wrote a blistering reply to *Nature* which has not been preserved. The editor sent it on to Pearson for comment, and Pearson drafted a reply which he showed to Gosset. In defense of the criticism in his review, Pearson quoted from a 1929 paper by the American economist Howard R. Tolley which showed that at least one reader had been misled by Fisher's book. In this paper, Tolley claimed that

> Recently the English school of statisticians has developed formulas and probability tables to accompany them which, they state, are applicable regardless of the form of the frequency distribution. These formulas are given, most of them without proof, in Fisher's book (1925). … If we accept the statements of those who have developed those newer formulas [i.e., Student and Fisher], skew frequency functions and small samples need cause us no further difficulty as far as measurement of error is concerned.…

Gosset offered to write to Fisher in an effort to mediate the dispute and in his letter referred Fisher to Tolley's paper. Fisher, somewhat shaken by Tolley's complete misunderstanding, gave a very conciliatory reply. After some more correspondence back and forth, Fisher, in a letter of 27 June 1929 [80], suggested that Gosset should write to *Nature* in his stead and that he (Fisher) would withdraw his letter.

Fisher's letter to Gosset is unusually revealing and the following paragraphs shed light on Fisher's attitude toward his book:

> In rereading the review, you must have noticed that there was a criticism which must be made on statistical grounds. I take this (and so do you) to mean that there was something wrong with it as statistics; what was wrong? The claim of exactness for the solutions and tests given was wrong, although a careful reader would find that I had kept within the letter of the law by hidden allusions to normality.
>
> …
>
> However important the question of normality may be, it is certainly irrelevant to the book he [E. S. Pearson] was reviewing. The reviewer takes up the position that he could have admitted this if I had mentioned normality more often or in larger type, or if I had made the meaningless claim that the solutions are approximate to those of certain undefined problems.
>
> About my examples I think you are right, they are nearly all very imperfect from the point of view of a purely theoretical treatise. In a practical treatise, I submit that the only question a critic has a right to raise is: – Are the right conclusions drawn, and by the right methods?

Gosset did write, as Fisher had suggested, and his letter was published in the 20 July (1929) issue of *Nature*. After quoting Tolley and absolving Fisher from subscribing to Tolley's views, he continues:

> The question of the applicability of normal theory to non-normal material is, however, of considerable importance. … I have always believed…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. … We should, however, be grateful

> to Dr. Fisher if he would show us elsewhere on theoretical grounds what sort of modification we require to make when the samples…are drawn from populations which are neither symmetrical nor mesokurtic.

In his reply of August 17, Fisher rejects Gosset's suggestion that he should give some guidance on how to modify the *t*-test for data from non-normal populations for a variety of reasons, but he hints at the existence of distribution-free tests and at taking higher moments into account. He concludes his letter by reiterating that

> On the practical side there is little enough room for anxiety, especially among biologists. … I have never known difficulty to arise in biological work from imperfect normality.

Thus, in his letter of 27 June 1929 [80], Fisher admits that his claim of exactness of various tests was wrong, but he says that he "had kept within the letter of the law by hidden allusions to normality." This suggests a conflicted attitude towards the presentation of this material. On the one hand, he wanted to teach his intended readers, mostly biologists, who would know little mathematics, how to use his methods. For this purpose, the presentation should be as simple and straightforward as possible, avoiding complications that would cause confusion and raise doubts.

On the other hand, he was a scientist and it bothered him to put down statements that were false, even if they held "for all practical purposes." So, to salve his conscience, he would sometimes add a "hidden allusion" that would prove his honesty without disturbing his readers.

A striking example of this strategy can be found in Chapter VI (Section 35). After introducing the transform z of the correlation coefficient, Fisher states:

> The standard error of z is simpler in form
>
> $$\sigma_z = 1/\sqrt{n-3} \quad (n \text{ is the sample size}),$$
>
> and is practically independent of the value of the correlation in the population from which the sample is drawn.

The verb "is" before σ_z seems to claim unambiguously that the formula for the standard error is exact. But if so, σ_z is completely independent of the population correlation, not only "practically independent." The added phrase is a hidden admission of the fact that the formula is only an approximation. (It is noteworthy that in later editions Fisher added the word "approximately" before the formula for σ_z).

As another example of the same strategy, consider Fisher's presentation of the distribution of Pearson's χ^2-statistic for goodness of fit in Chap. 4 (Sect. 4.1). After defining the statistic, Fisher states

> The form of the distribution of χ^2 was established by Pearson in 1900; it is therefore possible to calculate…

This statement gives the impression that χ^2 is one of the exact tests with which the book is chiefly concerned. Fisher does not mention that it is in fact a large-sample test based on the normal approximation to the multinomial distribution.

However, in Example 9 in this section, when testing the goodness of fit of a Poisson distribution, he casually mentions that:

> In applying the χ^2-test to such a series, it is desirable that the number expected should in no group be less than 5, since the calculated distribution of χ^2 is not very closely realized for very small classes.

What is the reader to make of the vague statement, "not very closely realized," when earlier it had been stated unequivocally that the distribution in question was a χ^2–distribution? It is again one of those hidden corrections, this time tucked away in an example separated by several pages from the original statement.

Unlike the later correction regarding the formula for σ_z, Fisher kept the presentation of the χ^2–distribution unchanged throughout all the editions.

2.6 The Impact of Fisher's SMRW

Fisher's SMRW brought together the two testing strands whose starting points were discussed in Chap. 1. Both Pearson's χ^2 and Student's small-sample tests had been extensively developed by Fisher in the intervening years, and the book thus made a wealth of new methods available to laboratory workers. The result was a complete change in statistical methodology.

Although reviewers were slow in recognizing this revolutionary development, the book made its way. The first edition of 1,050 copies was sold out after three years, and the second edition of 1,250 copies in another two. Every two to three years necessitated a new edition, which usually contained some improvements and often additions. The size of the editions steadily increased and the eleventh edition of 1950 ran to 7,500 copies. The last edition, the fourteenth, was published posthumously in 1970 from notes Fisher had prepared before his death in 1962.

In 1972, to commemorate Fisher's death ten years earlier, Oxford University Press brought out a one-volume edition of Fisher's three statistical books, SMRW, "The Design of Experiments," and "Statistical Methods and Scientific Inference," with an introduction by Fisher's coworker and later successor at Rothamsted, Frank Yates. Thus, SMRW is still in print today, 83 years later, a testament to its enduring value.

The enormous impact this book had on the field of statistics is indicated by a number of later developments, of which I shall mention two.

On 10 May 1950, Fisher received a letter from W. Allen Wallis of the University of Chicago[5]:

> This year marks the twenty-fifth anniversary of the publication of *Statistical Methods for Research Workers*. The *Journal of the American Statistical Association*, of which I have

[5] Published in Bennett (1990).

> recently become editor, hopes to mark this event by one or two articles on the character and consequences of that volume. In that connection, I would appreciate your help on two points.
>
> First, can you suggest two or three persons whom I might invite to prepare articles? … The two names that come first to my mind are Hotelling and Cochran.[6]…
> Second, would it be possible for you to prepare a paper for us on the history of SMRW? …

Fisher replied that

> I am sure that the choice of Hotelling and Cochran is an excellent one … . They are, however, both professional mathematicians, and the book was, of course, written not for mathematicians but for practitioners, who, insofar as they understand their fields of application, are good judges of the kind of statistics which aids them in their work. They constitute, I believe, the real and ultimate judges of such a book as mine, though, of course, the majority of them are rather inarticulate in mathematical circles.
>
> I should suggest, therefore…that others from one or more fields of application should be invited to show what the book or the ideas which it was intended to express have done for their own subjects.

In the event, the March 1951 issue of the Journal carried four articles on SMRW. Their authors were Frank Yates from Rothamsted; the mathematician, statistician, and economist Harold Hotelling; the chemist Jack Youden; and the biologist Kenneth Mather. Together, they provided the comprehensive view that Fisher had hoped for. Wallis's second suggestion of a paper by Fisher himself unfortunately was not realized.

A second example of the stature the book had attained can be found in a book published in 2005, "Landmark Writings in Western Mathematics, 1640–1940," edited by Grattan-Guinness. Each of its 77 chapters is devoted to a pathbreaking book or paper in pure or applied mathematics. The subject of Chapter 67 by the Cambridge Statistician A. W. F. Edwards is Fisher's SMRW.

The chapter begins with some details of the publication history of the book, including a list of its translations into French, German, Italian, Spanish, Japanese, and Russian. It then provides sections on "The Author" and the "Writing of Statistical Methods." The third section summarizes the contents of the first edition, and this is followed with a fairly detailed account of the later editions. The final section (except for a brief epilog) deals with the impact of the book. It begins by stating

That Fisher is the father of modern statistics, no one will dispute…

and then imagines how statistics would have developed if Fisher had not written this book.

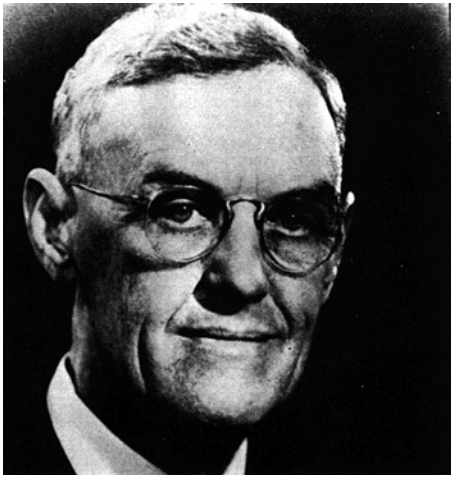
The author concludes that

> This would have made little difference to the dissemination of Fisher's more theoretical work, which would have still formed the backbone of 20th-Century mathematical statistics; but the effect of that work on applied statistical practice would have been felt more slowly, particularly in biology, including genetics, medicine and the design of agricultural experiments. If, as seems only just, one includes *The Design of Experiments* in the assessment, the impact of *Statistical Methods* throughout the biological sciences was profound and permanent, and from biology the influence spread rapidly into the social sciences.

---

[6] William Cochran.

2.7 Snedecor's "Statistical Methods": (George Snedecor)



The dissemination of Fisher's ideas set forth in his "Statistical Methods for Research Workers" was greatly furthered in an unexpected way, through a book with the shorter title, "Statistical Methods" (and the subtitle, "Applied to Experiments in Agriculture and Biology"), by a friend and admirer, George Snedecor (1881–1974).

Snedecor, who in 1933 had become the founding director of a statistical laboratory at Iowa State, in 1936 wrote to Fisher that he was working on an elementary text "designed to lead the beginner to an appreciation of your books." Snedecor's text came out in 1937 and covered roughly the same material as Fisher's SMRW: χ^2-tests, the normal one- and two-sample problem, regression, correlation, and the analysis of variance and of covariance. In addition, it included some of the ideas of Fisher's 1935 book on design of experiments, such as factorial experiments, randomized blocks, Latin squares, and confounding.

What distinguished it from Fisher's book was its style. The mode of presentation was much easier, much more user-friendly than Fisher's very terse way of writing.

Much of Snedecor's book is written as a conversation between author and reader. He constantly asks questions, points out to the reader how a new topic is related to ideas covered earlier, and illustrates each idea with numerous examples and exercises. It is also helpful that the material is divided into bite-size sections of usually a page or two.

As an example of the style, consider the beginning of Section 2.2, entitled, "The experiment described. The mean."

> Imagine a newly discovered apple, attractive in appearance, delicious in flavor, having apparently all the qualities of success. It has been christened "King." Only its yielding capacity in various localities is yet to be tested. The following procedure is decided upon. King is planted adjacent to Standard in 15 orchards scattered about the region suitable for their production. Years later, when the trees have matured, the yields are measured and recorded in Table 2.1.

This table, it is pointed out, shows that King is superior in every orchard and that the average difference over the 15 orchards is 6 bushels.

After some preliminary material that makes up Chap. 2, "An experiment designed to compare measurements of individuals," Chap. 3 discusses "Sampling from a normal distribution," and in Sect. 3.8 introduces the *t*-distribution. In this material, the apple example is occasionally used to illustrate the concepts. It resumes its central role in the next section which brings the *t*-test and states:

> At last we have reached the goal of this chapter. The aim is to test the significance of the mean difference, 6 bushels, between the yields of 15 pairs of apple trees.

The difference is found to be significant at the 1% level, and the section ends by asking:

> How great is the difference between the two population yields? We do not know. The best estimate we have is 6 bushels. In addition, we have the convincing circumstantial evidence of significance that King is the better yielder. Most people are satisfied with that information. If you are not, we shall try to give you something to think about in a section to follow (3.11).

It reads like a suspense story and keeps the reader interested.

The book was an enormous success. It has sold more than 200,000 copies in many editions, and was for some years one of the most cited books in the Science Citation Index. Fisher's own books were too difficult for mass appeal. As Fisher's daughter Joan Fisher Box wrote in the biography of her father (p. 313), "It was George W. Snedecor … who was to act as midwife in delivering the new statistics in the United States".

Fisher, Neyman, and the Creation of Classical Statistics
Lehmann, E.L.
2011, VIII, 115 p. 8 illus., Softcover
ISBN: 978-1-4419-9499-8