165

"STUDENT"

BY R. A. FISHER

THE untimely death of W. S. Gosset, at the age of 61, in October 1937, has taken one of the most original minds in contemporary science. Without being a professed mathematician, he first published, in 1908, a fundamentally new approach to the classical problem of the theory of errors, the consequences of which are still only gradually coming to be appreciated in the many fields of work to which it is applicable. The story of this advance is as instructive as it is interesting.

It was a primary achievement of the theory of errors, due to Gauss, to show that when the mean of a population is estimated from the mean of a number, n, of independent observations, the precision of this estimate might be inferred from the discrepancies among the several observations. These discrepancies naturally throw no light on the question whether the estimate is too great or too small; for this, further samples from the same population would be needed. Without further samples we may, however, form an estimate of the extent to which the means of such samples differ among themselves, for it is easily shown that the sampling variance of the means of samples of n is always equal to $1/n$ of the variance of individual observations. Moreover, an estimate of the variance of individual observations can be formed from the observed discrepancies of these from their observed mean. Thus, if $\bar{x}$ is the mean of n observations, x, the sum of the squares of the deviations

$$S(x-\bar{x})^2$$

when divided by $(n-1)$ yields an unbiased estimate of the sampling variance of the observations, that is, of the population sampled, and tends to be equal to that value as the size of the sample is increased.

Taking, therefore, the estimated variance of the population to be

$$s^2 = \frac{1}{n-1} S(x-\bar{x})^2,$$

the estimated sampling variance of the mean of n such observations is

$$s^2/n.$$

Such, in modern notation, was the theoretical basis available before "Student's" work for specifying the precision of means with a view to testing their significance. Writers on the method of least squares, following Gauss, had elaborated the problem so as to apply it to testing the significance also of differences between means, and of regression coefficients, always, however, on the basis of the adoption as true of the best available estimate of the residual variance.

"Student's" work has shown that a better course is open to us than that of adopting even the best available estimate of the variance of the population; that, by finding the exact sampling distribution of such an estimate, we may make allowance for its sampling errors, so as to obtain a test of significance which, in spite of these errors, is exact and rigorous. He thus rendered obsolete the restriction that the sample must be "sufficiently large", a restriction vaguely adumbrated by some previous writers, and ignored by others. The claim that "Student's" result is rigorous and exact has, of course, been resented by advocates of "large sample" methods. After the repeated publication of proofs of his result, this criticism has of late reduced itself to the trivial objection that it is exact only for normal distributions; that is, only for the case which he, like previous writers on the theory of errors, had thought it necessary to discuss.

If the problem had been presented to "Student" in the form in which we have stated it above, as left by the best of his predecessors, his task might seem, after the event, a straightforward, if not an easy one. For "Student" in 1907 the task was made unnecessarily complicated. By a foolish convention, biometrical work in this country had abandoned the use of the degrees of freedom as the divisor, when the variance of the population was estimated from the sum of squares. Thus they would write

$$\sigma^2 = \frac{1}{n} S(x-\bar{x})^2,$$

using the actual number of observations in the sample as the divisor. When it came to the estimated variance of the mean of n such observations the bias of the previous estimate would sometimes be repaired by taking the estimated sampling variance of the mean to be

$$\sigma^2/(n-1);$$

as though the variance of the mean of n independent observations could be other than one nth of the mean variance of each of them singly.

Sampling problems were, moreover, constantly and seriously obscured by the practice of using the same symbol for the quantity estimated and for the estimate of it derived from the observations. For example, we should now use σ^2 consistently only for the unknown variance of the population sampled, and should use some other symbol, such as s^2 for any of the estimates of it, which can be derived from the data. It is a tribute to "Student's" penetration that, without this aid in most of the literature he must have studied, he should yet have arrived at a clear insight into the nature of sampling problems.

If, then, σ^2 stands for the true variance of the population, and s^2 for the estimate,

$$\frac{1}{n-1} S(x-\bar{x})^2,$$

it has now been demonstrated, by a variety of methods of no great complexity, that the frequency distribution of s in different random samples, of the same size n, is given by the frequency element

$$df = \frac{(n-1)^{\frac{1}{2}(n-1)}}{2^{\frac{1}{2}(n-3)}(\frac{1}{2}n-\frac{3}{2})!} \frac{s^{n-2}}{\sigma^{n-1}} e^{-\frac{(n-1)s^2}{2\sigma^2}} ds.$$

This is equivalent to the statement that

$$(n-1)\, s^2/\sigma^2$$

is distributed as is χ^2 for $n-1$ degrees of freedom.

This important result was first found by Helmert about 1875, and it is remarkable that its discovery at that date led to no material advance in statistical technique. This must be ascribed, at least in part, to the peculiar circumstance that the earlier formula for estimating the variance, due to Bessel, the superiority of which had been demonstrated by Gauss, had been overshadowed by a method of estimation known as Peter's, using the equation,

$$s' = \sqrt{(\tfrac{1}{2}\pi)}\, \frac{S\,|x-\bar{x}|}{n-1},$$

of which the only recommendation seems to be that for some types of work it is more expeditious than the use of squares. The formula was, at the time, much used in astronomy, and, in the difficult problem of investigating the sampling properties of this mode of estimation, Helmert seems to have lost sight of the value of his discovery respecting the mean square.

It has sometimes been implied that Helmert had anticipated "Student's" work. It is true that in 1907 Helmert's discovery was unknown to English biometricians, and that "Student" had to solve his problem without the material aid that it would have afforded him. It was, however, the χ^2 distribution, named by Pearson, about 1900, and tabulated by P. Elderton, and not "Student's" t distribution, that Helmert really anticipated; what is remarkable is that, eight years after the χ^2 distribution had come into use as a test of goodness of fit, its simple connexion with the sampling distribution of the estimated variance should have been still unknown.

"Student" evidently attacked this preliminary obstacle in a practical spirit. He calculated the first four moments about zero of the sampling distribution of s^2, using an algebraic approach of great simplicity and generality; he recognized that these moments were those of the distribution given above; and adopted this distribution, perhaps with full confidence that it was exact, perhaps feeling sure only that it was a good approximation for the further use to which he wished to put it.

A second theoretical point was treated even more brusquely. The sampling distribution of s is in reality completely independent of that of $\bar{x}$. In order to derive the distribution of the significance test

$$t = \frac{\bar{x}-\mu}{s\sqrt{n}},$$

where μ is the true mean of the population sampled, it was necessary that this independence should be established. "Student" perceived this necessity, and any capable analyst could have shown him the demonstration he needed. As it was, he satisfied himself with showing somewhat laboriously that the distribution of s^2 was uncorrelated both with $\bar{x}$ and with $\bar{x}^2$. This was the most striking gap in his argument, for, in truth it was not merely the distribution of s found by Helmert, but the exact simultaneous distribution of s and $\bar{x}$ that "Student" needed to develop his test.

From this simultaneous distribution

$$df = \frac{\sqrt{n}}{\sigma\sqrt{(2\pi)}} e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2} d\bar{x} . \frac{(n-1)^{\frac{1}{2}(n-1)}}{2^{\frac{1}{2}(n-3)}(\frac{1}{2}n-\frac{3}{2})!} \frac{s^{n-2}}{\sigma^{n-1}} e^{-\frac{(n-1)s^2}{2\sigma^2}} ds$$

it follows, merely by substituting ts for $(\bar{x}-\mu)\sqrt{n}$, and integrating with respect to s from 0 to ∞, that the sampling distribution of t is

$$\frac{(\frac{1}{2}n-1)!}{(\frac{1}{2}n-\frac{3}{2})!\sqrt{\{(n-1)\pi\}}} \cdot \frac{dt}{\left(1+\frac{t^2}{n-1}\right)^{\frac{1}{2}n}}.$$

The frequency with which t falls within any assigned limits is thereby known, since the distribution involves only the known quantity n. Since, moreover, the only unknown quantity involved in the expression for t is the parameter μ, the mean of the population sampled, direct frequency statements respecting the value of μ become immediately possible. In fact, the fiducial distribution of μ deducible from the data only has been determined.

The introduction and use of fiducial probability was, however, a later development. "Student" was immediately concerned only with developing an exact test of significance. That he was, in the 1908 paper, thinking of probability statements respecting μ cannot, however, be doubted. In the case $n=2$, for example, he points out that the quartiles of the distribution of μ lie at the two observed points. This conclusion is notable as illustrating a type of inference entirely independent of the form of the distribution sampled; for if μ stands for the median of such a distribution, it follows, since the three probabilities that two independent observations (*a*) should both exceed the median, (*b*) should lie on each side of it, and (*c*) should both fall short of it, must occur in the frequency ratio 1 : 2 : 1, that we may infer that the two observations are the quartiles of the fiducial distribution of the median.

A similar inference may be drawn with respect to a sample of any number, n, of observations. The probability that the median shall exceed r of these and be less than $n-r$, is easily seen to be

$$\frac{n!}{r!(n-r)!} 2^{-n}.$$

Thus, for a sample of 6 the chance that it lies outside the observed range is only 0·03125; while for a sample of 9 the fiducial probability of it lying outside the penultimate pair is only 0·03906. Similarly, for a sample of 12, the chance is only 0·0386 of it lying outside the range of the central 8 observations.

More generally, we may draw similar inferences with respect to any percentile point of a continuous distribution, without reference to its theoretical form. For, if μ_p stand for the hypothetical value which exceeds the fraction p of the population sampled, the terms of the expansion of

$$(q+p)^n, \quad p+q=1,$$

will give in succession the probabilities that μ_p is less than all the observed values, exceeds one of them, exceeds two, or finally, exceeds them all.

The same method further yields the simultaneous fiducial distribution of two or more such hypothetical percentile values. For, if these divide the hypothetical population in the ratio $p:q:r:s:\ldots$ the terms of the multinomial expansion

$$(p+q+r+s+\ldots)^n$$

will give the probabilities of any simultaneous distribution of the percentile points among the $n+1$ spaces within and beyond the range of the observations.

More precisely, if $p_1, \ldots, p_n$ are the fractions of the population exceeded by the n values observed in order of increasing magnitude, the simultaneous distribution of these n unknowns is simply

$$n!\, dp_1, \ldots, dp_n, \quad 0 \leqslant p_1 \leqslant p_2 \ldots \leqslant p_n \leqslant 1.$$

The hint implicit in "Student's" remark was not, however, taken up by other writers, and the results of the general method outlined above seem to be almost unknown.

It is doubtful if "Student" ever realized the full importance of his contribution to the Theory of Errors. From correspondence with him before the War, during which I sent him the proofs of his method, derivable from the consideration of the sample as a point in Euclidean hyperspace, I should form a confident judgement that at that time certainly he did not see how big a thing he had done. That the successful solution of the problem that had engaged his interest gave him satisfaction, there can be no doubt. Nor can it be doubted that he appreciated the convenience, in his own experimental studies, of having an exact test of significance to rely on. Probably he felt that, had the problem really been so important as it had once seemed, the leading authorities in English statistics would at least have given him the encouragement of recommending the use of his method; and, better still, would have sought to gain similar advantages in more complex problems. Five years, however, passed, without the writers in *Biometrika*, the journal in which he had published, showing any sign of appreciating the significance of his work. This weighty apathy must greatly have chilled his enthusiasm. In his correspondence with me he did not even suggest that the completed proofs should be published. In consequence, they finally saw the light, in 1915, only incidentally, as a preliminary step in the determination of the exact sampling distribution of the correlation coefficient.

The fruition of his work was, therefore, greatly postponed by the lack of appreciation of others. It would not be too much to ascribe this to the increasing dissociation of theoretical statistics from the practical problems of scientific research. For "Student" had only opened the door with the result he had published. The significance of the mean of a unique sample could now be tested exactly. But the same could not be said, even of the significance of the difference between two experimental samples grown for comparison; or, of the significance of a regression coefficient, which may be treated as a weighted mean; though the exact solution of both of these problems was later shown to be reducible to a t test. Further, although requiring more extensive tables, a host of problems involving variances, now generally handled by the technique known as the Analysis of Variance, really involved no more than applications to problems of a different kind of the principle of making exact allowance for the sampling errors, which "Student" had introduced for his problem of the mean of a unique sample. It was sixteen years before, in 1924, the system of tests of which "Student's" was the prototype was logically complete. Only during the thirteen years which have since passed has "Student's" work found its proper place as an experimental resource.

How did it come about that a man of "Student's" interests and training should have made an advance of fundamental mathematical importance, the possibility of which had been overlooked by the very brilliant mathematicians who have studied the Theory of Errors? Of his personal qualifications the most obvious are a clear head, and the practice of forming independent judgements. Without these it is certain that he would not have perceived the problem he was to solve. But, in varying degrees, these are qualities possessed by many mathematicians better able than "Student" to overcome the analytic difficulties of this task. One immense advantage which "Student" possessed was his concern with, and responsibility for, the practical interpretation of experimental data. If more mathematicians shared this advantage there can be no doubt that mathematical research would be more fruitfully directed than it often is. Many mathematicians must possess the penetration necessary to perceive, when confronted with concrete experimental results, that it must be possible to use them, by rigorously objective calculations, to throw light on the plausibility or otherwise of the interpretations that suggest themselves. A few must also possess the pertinacity needed to convert this intuition into such a completed procedure as we know as a test of significance. It is, I believe, nothing but an illusion to think that this process can ever be reduced to a self-contained mathematical theory of tests of significance. Constructive imagination, together with much knowledge based on experience of data of the same kind, must be exercised before deciding on what hypotheses are worth testing, and in what respects. Only when this fundamental thinking has been accomplished can the problem be given a mathematical form. "Student" possessed the constructive imagination and the pertinacity, and in these respects he was fitted both to perceive and to solve his problem.

The same qualities appear in some of his other scientific papers. Without knowledge of previous work he recognized the importance for experimentalists of the Poisson Series, which had been ignored in biometry, though some writers on probability had given it much attention. The influence of the statistical tests in remodelling the design on which experiments are planned came suddenly, and took him, perhaps, somewhat by surprise. Probably he could never easily accommodate himself to the view that statistical conclusions must often be vague and unsatisfactory, unless the experimenter deliberately plans his methods so as to make the estimate of error a valid one for his purpose; that by inattention to this requirement the experimenter might falsify just those elements of his data on which the statistician needed to rely. Certainly, though he practised it, he did not consistently appreciate the necessity of randomization, or the theoretical impossibility of obtaining systematic designs of which both the real and the estimated error shall be less than those given by the same plots when randomized; this special failure was perhaps only a sign of his loyalty to colleagues whose work was in this respect open to criticism.

His interest in the theory of evolution is less well known than is his work on mathematical statistics, although it was, in fact, a subject which constantly fascinated him.

The discovery of Mendel's work at the close of the century, when Gosset was 24, produced for more than a decade a rather wide loss of confidence in the Theory of Natural Selection. The retrogression of opinion is not perhaps in the long run to be regretted, since Darwin's work had never been comprehended by more than a minority, and the teaching of the biological theory had been encumbered by much unintelligent repetition of half-understood phrases. The forensic attack which developed might reasonably be expected to lead to greater clarity of meaning. The Mendelian discovery, since it embodied some facts unknown to Darwin, was eagerly seized on, and an antagonism between these facts and Darwin's theory was assumed and asserted, though never conscientiously examined. The early advocates of Mendelism, such as Bateson, had already before its discovery embroiled themselves in anti-Darwinian controversy.

The early geneticist to whom the elucidation of a situation involving only two factors, such as rose and pea in the combs of poultry, seemed a major achievement, found it easy to assume that natural populations contained heredity differences for only very few factors. Since their breeding technique was inadequate to handle factors having small effects, they often ignored the probability that on any given character most of the factors make only small differences. This early bias seems to be responsible for the belief, widely held among geneticists not so many years ago, that the selection of small differences, in this connexion commonly called "fluctuations", can only lead to unimportant evolutionary effects. It underlies the use of the oft-repeated statement that selection can do no more than select the best of the existing variety of genotypes; a statement obviously true, though irrelevant in its application, since it needs the support of the belief, also widely taught, that the diversity available for selection is easily exhausted. It was, of course, inferred that evolutionary progress must wait upon the occurrence of new mutations; and in the so-called "Mutation Theory", which still has its adherents, the argument is so telescoped that evolution is represented as consisting in such mutations.

Mendel himself, it should be explained, anticipated that even closely related species must differ in a large number of factors, and clearly set forth the immense number of genotypes which variation in comparatively few factors made possible. Moreover, experimental breeding continued on an increasing scale, and, whenever it was devoted to practical objects, the importance of fluctuating inheritance consistently showed itself. Many critical minds must have felt as Gosset came to feel—I quote from a letter of January 1933:

For at least twenty-five years I've been reading that the continued accumulation of infinitesimal variations can do nothing, and all the time I've felt in my bones that Darwin was right.

The trend of his thoughts, on this as on many other subjects, followed much the same lines as my own, and it will illustrate both the charm of his manner, and the real generosity of his attitude, if I quote from another letter of the following month:

It is narrated in the Classics that Kai Lung, having expended his last tael in publishing the outcome of his many versatile and highly ornamental ideas, found to his stupefaction that his mind had been cast so nearly in the mould of the great Lo Kuan Chan that there was nothing original in his own publication. The analogy is admittedly inexact, for Kai Lung's trouble arose from the fact that he had been so engrossed in a literary career that he had had no time to read the great poet's work, while mine is due to my exceedingly bad habit of reading without paying the least attention to what I read. Moreover my unconscious doubtless takes in more than I realize so that in this way the many ingenious and "original" suggestions in Student's last note in the *Eugenics Review* can really be traced to their origin in p. 96 of the *Genetical Theory of Natural Selection*. Seriously, it was not until your reticence impressed it on me, that I realized that of course you must have already said much if not all of what I was trying to say, and I had just found the reference when your last note came....

And so I hope I may be forgiven: the plagiarism was quite unconscious, even the last two paragraphs of which I was (and am!) inordinately proud are mere corollaries of p. 96, which is as compressed as a Chinese poem....

What, with his absurd generosity, Gosset chooses to speak of as a plagiarism had not indeed struck me in this light. It seemed to me the natural result of applying statistical reasoning to experimental data on the Mendelian theory. I hoped and thought that his ideas on the subject were aided by the general arguments developed in my book, but was none the less honoured that his conclusions agreed so closely with my own in rejecting the so-called "Mutation Theory" of evolution, traditionally supported by geneticists in England and America since the beginnings of the century, and in emphasizing the unsuspected efficacy of Darwinian selection among the Mendelian genotypes. Gosset followed up this work in the following year with "A calculation of the minimum number of genes in Winter's selection experiment", the first example, I believe, of a paper in which the difficulties of estimating such a minimum number have been fairly faced and overcome.

In writing this brief memoir of a loyal and generous friend I have purposely laid emphasis on one scientific achievement with which his name will always be associated. His life was one full of fruitful scientific ideas, and his versatility extended beyond his interests in research. In spite of his many activities it is the "Student" of "*Student's*" *test of significance* who has won, and deserved to win, a unique place in the history of scientific method.

BIBLIOGRAPHY

(1) "On the error of counting with a haemacytometer" (1907). *Biometrika*, 5, 351.
(2) "The probable error of a mean" (1908). *Biometrika*, **6**, 1.
(3) "Probable error of a correlation coefficient" (1908). *Biometrika*, 6, 302.
(4) "The distribution of the means of samples which are not drawn at random" (1909). *Biometrika*, 7, 210.
(5) "Appendix to Mercer and Hall's paper on 'The experimental error of field trials'" (1911). *J. Agric. Soc.* **4**, 128.
(6) "The correction to be made to the correlation ratio for grouping" (1913). *Biometrika*, **9**, 316.
(7) "The elimination of spurious correlation due to position in time and space" (1914). *Biometrika*, **10**, 179.
(8) "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" (1917). *Biometrika*, **11**, 179.
(9) "An explanation of the deviations from Poisson's law in practice" (1919). *Biometrika*, **12**, 211.
(10) "An experimental determination of the probable error of Dr Spearman's correlation coefficients" (1921). *Biometrika*, **13**, 263.
(11) "On testing varieties of cereals" (1923). *Biometrika*, **15**, 271.
(12) "Note by 'Student' with regard to his paper 'On testing varieties of cereals'" (1924). *Biometrika*, **16**, 411.
(13) "New tables for testing the significance of observations" (1925). *Metron*, **5**, 105.
(14) "Mathematics and agronomy" (1926). *J. Amer. Soc. Agron.* **18**, 703.
(15) "Errors of routine analysis" (1927). *Biometrika*, **19**, 151.
(16) "The Lanarkshire milk experiment" (1931). *Biometrika*, **23**, 398.
(17) Article on "Yield Trials" in Baillière's *Encyclopaedia of Scientific Agriculture* (1931).
(18) "On the 'z' test" (1931). *Biometrika*, **23**, 407.
(19) "Evolution by selection. The implications of Winter's selection experiment" (1933). *Eugen. Rev.* **24**, 293.
(20) "A calculation of the minimum number of genes in Winter's selection experiment" (1934). *Ann. Eugen., Lond.*, **6**, 77.
(21) "Co-operation in large-scale experiments. A discussion opened by W. S. Gosset" (1936). Supplement to *J. Roy. Statist. Soc.* **3**, 115.
(22) "Random and balanced arrangements" (1937). *Biometrika*, **29**, 363.