The Generalization of `Student's' Problem when Several Different Population Variances are Involved
Author(s): B. L. Welch
Source: *Biometrika*, Vol. 34, No. 1/2 (Jan., 1947), pp. 28-35
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/2332510
Accessed: 21-10-2019 11:38 UTC

THE GENERALIZATION OF 'STUDENT'S' PROBLEM WHEN SEVERAL DIFFERENT POPULATION VARIANCES ARE INVOLVED

By B. L. WELCH, B.A., Ph.D.

1. *Introduction and summary.* Let η be a population parameter which is estimated by an observed quantity y, normally distributed with variance σ_y^2. Let $\sigma_y^2 = \sum_{i=1}^{k} \lambda_i \sigma_i^2$, where the λ_i are known positive numbers and the σ_i^2 are unknown variances. Suppose that the observed data provide estimates s_i^2 of these variances, based on f_i degrees of freedom, respectively, so that the sampling distribution of s_i^2 is

$$p(s_i^2)\,ds_i^2 = \frac{1}{\Gamma(\frac{1}{2}f_i)}\left(\frac{f_i s_i^2}{2\sigma_i^2}\right)^{\frac{1}{2}f_i-1} \exp\left[-\frac{1}{2}\frac{f_i s_i^2}{\sigma_i^2}\right] d\left(\frac{f_i s_i^2}{2\sigma_i^2}\right), \tag{1}$$

and that these estimates are distributed independently of each other and of y.

A very simple particular case of this set-up occurs when we have samples of n_1 and n_2, respectively, from two normal populations with true means α_1 and α_2 and standard deviations σ_1 and σ_2. If η is the true difference $(\alpha_1 - \alpha_2)$ between the means, the estimated difference is $y = (\bar{x}_1 - \bar{x}_2)$. The variance of the estimate is $\sigma_y^2 = (\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2)$, where $\lambda_1 = 1/n_1$ and $\lambda_2 = 1/n_2$. The estimated values of σ_1^2 and σ_2^2 are $s_1^2 = \Sigma_1/f_1$ and $s_2^2 = \Sigma_2/f_2$, where Σ_1 and Σ_2 are the respective sums of squares of observations from the individual sample means and $f_1 = (n_1 - 1)$ and $f_2 = (n_2 - 1)$. These s^2 are distributed in the form (1) and the postulated conditions of independence hold.

Another particular case, again with $k = 2$, arises when we wish to compare two regression coefficients, fitted to independent sets of data, without making the assumption that the population residual variance about the true regression line is the same for both sets.

The present paper is written mainly with these practical applications of the case $k = 2$ in mind, but the results are expressed generally for any k, since no further analytical difficulties are involved. It will be shown how probability statements about y, considered as an estimate of η, may be made similar in character to those which W. S. Gosset derived for the mean of a single sample of n observations ('Student', 1908). We shall, in effect, seek a quantity h, calculable from the observations, with the property that the chance of the difference $(y - \eta)$ falling short of h is a given probability P. It is clear that h must be a function of the individual variances s_i^2 and of P. If the abbreviation Pr. is used to mean 'the probability of the relation in the bracket following', our problem is to satisfy the equation

$$\text{Pr.}\,[(y - \eta) < h(s_1^2, s_2^2, \ldots, s_k^2, P)] = P. \tag{2}$$

In Gosset's case the solution was, of course, simply

$$\text{Pr.}\,[(\bar{x} - \alpha) < t_P s/\sqrt{n}] = P, \tag{3}$$

where t_P is the value, corresponding to the probability level P, in the 'Student' t-distribution with $f = (n-1)$ degrees of freedom.

In the next section the mathematical derivation of the exact solution of (2) is given. This is then followed by some consideration of its expression in numerical terms. First, a series solution in powers of $1/f_i$ is developed, which may be used for calculating tables. Then some comparisons are made with a non-series approximate solution which is based on a particular way of regarding the distribution of a quantity of the general form $z = (\Sigma a_i \chi_i^2)$.

Some brief discussion is then added which may serve to place the present contribution in its proper relationship to other papers which have been written on this topic.

Finally, it is shown how the inequality (2) may be adapted to provide an interval estimate for η.

2. *Mathematical derivation of solution.* Let $j(s_1^2, s_2^2, \ldots, s_k^2, P)$ denote the probability that $(y-\eta)$ is less than $h(s_1^2, s_2^2, \ldots, s_k^2, P)$, *given* s_i^2 $(i = 1, 2, \ldots, k)$. Then, since y is distributed quite independently of the estimated variances, we have

$$j(s_1^2, s_2^2, \ldots, s_k^2, P) = \int_{u=-\infty}^{\frac{h(s_1^2, s_2^2, \ldots, s_k^2, P)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2}\, du = I\left\{\frac{h(s_1^2, s_2^2, \ldots, s_k^2, P)}{\sqrt{(\Sigma\lambda_i\sigma_i^2)}}\right\}, \tag{4}$$

where I is used to denote the normal probability integral. The condition of equation (2) is then simply that, if $j(s_1^2, s_2^2, \ldots, s_k^2, P)$ is averaged over the probability distributions of s_i^2 as given by (1), the result will equal P. Thus

$$\int_{s^2} \cdots \int j(s_1^2, s_2^2, \ldots, s_k^2, P) \prod_i p(s_i^2)\, ds_i^2 = P. \tag{5}$$

Now we may expand $j(s_1^2, s_2^2, \ldots, s_k^2, P)$ about an origin $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$ in a Taylor expansion. Thus

$$j(s_1^2, s_2^2, \ldots, s_k^2, P) = \exp\left[\sum_i (s_i^2 - \sigma_i^2)\, \partial_i\right] j(w_1, w_2, \ldots, w_k, P), \tag{6}$$

it being understood that the exponential is to be expanded in a power series in ∂_i and that ∂_i is to be interpreted so that

$$\partial_i^r j(w_1, w_2, \ldots, w_k, P) = \left[\frac{\partial}{\partial w_i^r} j(w_1, w_2, \ldots, w_k, P)\right]_{w_j=\sigma_j^2} \tag{7}$$

On making the substitution of (6) into (5) our result may be written

$$\mathbf{\Theta} j(w_1, w_2, \ldots, w_k, P) = P, \tag{8}$$

where

$$\mathbf{\Theta} = \prod_i \int \exp\left[(s_i^2 - \sigma_i^2)\partial_i\right] p(s_i^2)\, ds_i^2. \tag{9}$$

Now, substituting into (9) from (1), the integral comes out in simple form, i.e.

$$\begin{aligned}
\mathbf{\Theta} &= \prod_i \left\{1 - \frac{2\sigma_i^2 \partial_i}{f_i}\right\}^{-\frac{1}{2}f_i} \exp\left[-\sigma_i^2 \partial_i\right] \\
&= \exp\left\{-\Sigma\sigma_i^2\partial_i - \tfrac{1}{2}\Sigma f_i \log\left(1 - \frac{2\sigma_i^2\partial_i}{f_i}\right)\right\} \\
&= \exp\left\{\Sigma\frac{\sigma_i^4\partial_i^2}{f_i} + \frac{4}{3}\Sigma\frac{\sigma_i^6\partial_i^3}{f_i^2} + 2\Sigma\frac{\sigma_i^8\partial_i^4}{f_i^3} + \text{etc.}\right\} \\
&= 1 + \Sigma\frac{\sigma_i^4\partial_i^2}{f_i} + \left\{\frac{4}{3}\Sigma\frac{\sigma_i^6\partial_i^3}{f_i^2} + \frac{1}{2}\left(\Sigma\frac{\sigma_i^4\partial_i^2}{f_i}\right)^2\right\} + \text{etc.}
\end{aligned} \tag{10}$$

Substituting (4) into (8) we have finally

$$\mathbf{\Theta} I\left\{\frac{h(w_1, w_2, \ldots, w_k, P)}{\sqrt{(\Sigma \lambda_i \sigma_i^2)}}\right\} = P. \tag{11}$$

This, in a very condensed form, is the solution to our problem.* The operator $\mathbf{\Theta}$ constitutes a direction to carry out the partial differentiations indicated by (10). w_j must then be equated to σ_j^2. The solution of the resulting equation will give $h(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2, P)$ and therefore the required $h(s_1^2, s_2^2, \ldots, s_k^2, P)$.

3. *The development of the series solution.* It will be convenient to write $h(w)$ for $h(w_1, w_2, \ldots, w_k, P)$ and ξ for the normal deviate such that $I(\xi) = P$. We may then expand $I\left\{\frac{h(w)}{\sqrt{(\Sigma \lambda_i \sigma_i^2)}}\right\}$ in a Taylor series about ξ as origin. Thus

$$I\left\{\frac{h(w)}{\sqrt{(\Sigma \lambda_i \sigma_i^2)}}\right\} = \exp\left[\left\{\frac{h(w)}{\sqrt{(\Sigma \lambda_i \sigma_i^2)}} - \xi\right\} D\right] I(v), \tag{12}$$

it being understood that the exponential is to be expanded in powers of D, and that these powers are to be interpreted so that

$$D^r I(v) = \left[\frac{d^r}{dv^r} I(v)\right]_{v=\xi}. \tag{13}$$

Equation (11) then becomes

$$\mathbf{\Theta} \exp\left[\left\{\frac{h(w)}{\sqrt{(\Sigma \lambda_i \sigma_i^2)}} - \xi\right\} D\right] I(v) = I(\xi). \tag{14}$$

This may now be solved by successive approximations.

The initial approximation is the large-sample normal approximation

$$h_0(w) = \xi \sqrt{(\Sigma \lambda_i w_i)}, \tag{15}$$

and we may write

$$h(w) = \xi \sqrt{(\Sigma \lambda_i w_i)} + h_1(w) + h_2(w) + \text{etc.}, \tag{16}$$

where $h_1(w)$ includes terms of order $1/f_i$, $h_2(w)$ terms of order $1/f_i^2$ and so on. For the moment we shall treat terms of the order $1/f_i^3$ as negligible. Then (14) gives

$$\mathbf{\Theta} \exp\left[\left\{\frac{\xi \sqrt{(\Sigma \lambda_i w_i)}}{\sqrt{(\Sigma \lambda_i \sigma_i^2)}} - \xi\right\} D\right] \exp\left[\left\{\frac{h_1(w) + h_2(w)}{\sqrt{(\Sigma \lambda_i \sigma_i^2)}}\right\} D\right] I(v) = I(\xi), \tag{17}$$

i.e.

$$\mathbf{\Theta} \exp\left[\xi D\left\{\sqrt{\frac{\Sigma \lambda_i w_i}{\Sigma \lambda_i \sigma_i^2}} - 1\right\}\right]\left[1 + \frac{h_1(w)\, D}{\sqrt{(\Sigma \lambda_i \sigma_i^2)}} + \left\{\frac{h_2(w)\, D}{\sqrt{(\Sigma \lambda_i \sigma_i^2)}} + \frac{1}{2}\frac{h_1^2(w)\, D^2}{\Sigma \lambda_i \sigma_i^2}\right\} \ldots\right] I(v) = I(\xi). \tag{18}$$

Or, using (10),

$$\begin{aligned} &\left[\frac{h_1(\sigma^2)\, D}{\sqrt{(\Sigma \lambda_i \sigma_i^2)}} + \Sigma \frac{\sigma_i^4 \partial_i^2}{f_i} \exp\left(\xi D\left\{\sqrt{\frac{\Sigma \lambda_i w_i}{\Sigma \lambda_i \sigma_i^2}} - 1\right\}\right)\right] I(v) \\ &\quad + \left[\frac{h_2(\sigma^2)\, D}{\sqrt{(\Sigma \lambda_i \sigma_i^2)}} + \frac{1}{2}\frac{h_1^2(\sigma^2)\, D^2}{\Sigma \lambda_i \sigma_i^2} + \Sigma \frac{\sigma_i^4 \partial_i^2}{f_i} \exp\left(\xi D\left\{\sqrt{\frac{\Sigma \lambda_i w_i}{\Sigma \lambda_i \sigma_i^2}} - 1\right\}\right) \frac{h_1(w)}{\sqrt{(\Sigma \lambda_i \sigma_i^2)}} \right. \\ &\quad \left. + \left\{\frac{4}{3}\Sigma \frac{\sigma_i^6 \partial_i^3}{f_i^2} + \frac{1}{2}\left(\Sigma \frac{\sigma_i^4 \partial_i^2}{f_i}\right)^2\right\} \exp\left(\xi D\left\{\sqrt{\frac{\Sigma \lambda_i w_i}{\Sigma \lambda_i \sigma_i^2}} - 1\right\}\right)\right] I(v) = 0. \end{aligned} \tag{19}$$

The equation of the first order term to zero gives

$$h_1(\sigma^2) = \frac{\xi(1 + \xi^2)}{4} \frac{\left(\Sigma \frac{\lambda_i^2 \sigma_i^4}{f_i}\right)}{(\Sigma \lambda_i \sigma_i^2)^{\frac{3}{2}}}. \tag{20}$$

* Equation (11) can also be expressed as an integral equation and this form may be necessary for providing numerical values where the f_i are very small.

This can then be substituted in the second-order term which, when equated to zero, will give $h_2(\sigma^2)$. The process may obviously be extended to higher orders, although the expressions become so complex that a slightly different procedure has then been found to be preferable. To terms of order $1/f_i^2$ our solution is

$$h(s^2) = \xi\sqrt{(\Sigma\lambda_i s_i^2)}\left[1+\frac{(1+\xi^2)}{4}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)}{(\Sigma\lambda_i s_i^2)^2}-\frac{(1+\xi^2)}{2}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^2}\right.$$
$$\left.+\frac{(3+5\xi^2+\xi^4)}{3}\frac{\left(\Sigma\frac{\lambda_i^3 s_i^6}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^3}-\frac{(15+32\xi^2+9\xi^4)}{32}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)^2}{(\Sigma\lambda_i s_i^2)^4}\right]. \quad (21)$$

It may be noted that in the particular case $k = 1$, this reduces, as it should, to the already known expansion of the deviate of the straightforward 'Student' distribution (Fisher, 1941, p. 151), viz.

$$t_P = \xi\left[1+\frac{(1+\xi^2)}{4f}+\frac{(3+16\xi^2+5\xi^4)}{96f^2}+\text{etc.}\right]. \quad (22)$$

It is proposed in another communication to give tables of $h(s^2)$ based on the expansion (21) carried to some further terms.

4. *Discussion of a non-series approximation.* It will be recalled that in Gosset's original approach to the single sample problem ('Student', 1908) his initial step was to note that the first four moments of the distribution of s^2 were consistent with the assumption that the distribution could be represented by a Pearson Type III curve. He was fortunate in this way to rediscover a distribution which had already been found by Helmert, as this permitted him to go on to the derivation of the t-distribution. In our present case, as in many others arising naturally in statistical work, we are led to consider, instead of s^2, a linear function $\Sigma\lambda_i s_i^2$ of several s_i^2. If this linear function were distributed in a Pearson Type III distribution a whole range of new problems could be dealt with by well-established theory. However, in general, we do not have this good fortune. For $\Sigma\lambda_i s_i^2$ is of the form $\Sigma a_i \chi_i^2$, where $a_i = \lambda_i\sigma_i^2/f_i$, and the distribution of this quantity is only of Type III if all the a_i, except one, are zero, or if all the a_i happen to be equal.

Nevertheless, for practical purposes an *approximation* to the distribution of $\Sigma\lambda_i s_i^2$, using a Type III curve with start, mean and variance suitably adjusted, can still be useful. In two previous papers (Welch, 1936, 1938) I have employed this method to obtain numerical comparisons of the merits of different statistical procedures, where full calculations with the true distributions would have been unduly laborious. The method of determining the constants in the approximation was given for the case $k = 2$ in the first of these papers and is as follows.

If $z = (a\chi_1^2 + b\chi_2^2)$, and the approximate distribution curve is written in the form

$$p(z)\,dz = \frac{1}{\Gamma(\frac{1}{2}f)}e^{-\frac{1}{2}(z/g)}\left(\frac{z}{2g}\right)^{\frac{1}{2}f-1}d\left(\frac{z}{2g}\right), \quad (23)$$

then making the first two moments of (23) agree with the true moments of z, we find

$$f = \frac{(af_1+bf_2)^2}{a^2f_1+b^2f_2}, \quad g = \frac{a^2f_1+b^2f_2}{af_1+bf_2}. \quad (24)$$

Phrasing the matter rather differently, we can say that z/g is approximately distributed as

χ^2 with degrees of freedom f. Of course f, given by (24), will in general be fractional, but the letter used to designate this quantity was chosen, and the term 'effective degrees of freedom' has been used, because by doing so we can appeal immediately to a considerable body of further theoretical results.

In particular we can say that the criterion

$$v = \frac{(y-\eta)}{\sqrt{(\lambda_1 s_1^2 + \lambda_2 s_2^2)}} \tag{25}$$

follows approximately the 'Student' t-distribution with degrees of freedom

$$f = \frac{(\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2)^2}{\dfrac{\lambda_1^2\sigma_1^4}{f_1} + \dfrac{\lambda_2^2\sigma_2^4}{f_2}}. \tag{26}$$

More generally, when k is not restricted to 2, the same line of argument leads us to say that the criterion

$$v = \frac{(y-\eta)}{\sqrt{(\Sigma\lambda_i s_i^2)}} \tag{27}$$

is approximately distributed as 'Student's' t with degrees of freedom

$$f = \frac{(\Sigma\lambda_i\sigma_i^2)^2}{\Sigma\dfrac{\lambda_i^2\sigma_i^4}{f_i}}. \tag{28}$$

Not knowing the σ_i's in (28), there are several ways in which we may now proceed, depending on what weight we may be willing to attach to any vague *a priori* notions we may possess of their *relative* magnitudes (cf. Welch, 1938). If we are not willing to assume anything, perhaps the best choice is

$$f = \frac{(\Sigma\lambda_i s_i^2)^2 - 2\left(\Sigma\dfrac{\lambda_i^2 s_i^4}{f_i+2}\right)}{\left(\Sigma\dfrac{\lambda_i^2 s_i^4}{f_i+2}\right)}. \tag{29}$$

It may be shown that the numerator of (29) has, in repeated samples, an average value $(\Sigma\lambda_i\sigma_i^2)^2$, and the denominator has average value $\Sigma\lambda_i^2\sigma_i^4/f_i$. In a certain sense, therefore, (29) is a fair estimate of (28).

To sum up, then, the interpretation of y as an estimate of η, using the present type of approximation involves only the reference of the criterion (27) to tables of the 'Student' distribution, entered with degrees of freedom given by (29).

Some further light is now thrown on this procedure by the expansion for the exact solution of our problem derived in the preceding section. For the implications of referring v to the 'Student' distribution may be seen by substituting f from (29) into the expansion (22) of the 'Student' deviate. On doing this and then expanding in powers of $1/f_i$ it is found that, in effect, our approximation corresponds to assuming that

$$h(s^2) = \xi\sqrt{(\Sigma\lambda_i s_i^2)}\left[1 + \frac{(1+\xi^2)}{4}\frac{\left(\Sigma\dfrac{\lambda_i^2 s_i^4}{f_i}\right)}{(\Sigma\lambda_i s_i^2)^2} - \frac{(1+\xi^2)}{2}\frac{\left(\Sigma\dfrac{\lambda_i^2 s_i^4}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^2} + \frac{(51+64\xi^2+5\xi^4)}{96}\frac{\left(\Sigma\dfrac{\lambda_i^2 s_i^4}{f_i}\right)^2}{(\Sigma\lambda_i s_i^2)^4} + \dots\right], \tag{30}$$

whereas, in fact, the true solution is given by (21). Comparison shows that we have exact agreement to terms of order $1/f_i$ and in the first of the quadratic terms. To the second order the difference between the expressions in square brackets in equations (21) and (30) is

$$\frac{(3+5\xi^2+\xi^4)}{3}\left\{\frac{\left(\Sigma\frac{\lambda_i^3 s_i^6}{f_i^2}\right)}{(\Sigma\lambda_i s_i^2)^3}-\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)^2}{(\Sigma\lambda_i s_i^2)^4}\right\}. \tag{31}$$

This difference vanishes if any one of the s_i^2 is overwhelmingly larger than all the others, or if s_i^2 is proportional to f_i/λ_i. It appears that, in general, the difference is not likely to be large. We have, therefore, found some justification for using the Type III approximation in the present case.

The above comparison has been made on the basis of the series developments, but it should be borne in mind that approximations based on positive frequency functions, such as those falling under the Pearson system, usually provide a higher degree of accuracy than might appear from any consideration of expansions. Furthermore, they are apt to give an insight into the nature of the situation which may sometimes be lost in working out the details of exact solutions. In the present case I feel that the comparison of this section serves to give added confidence in the exact solution,* which I have put forward in the previous two sections, quite as much as it demonstrates the value of the approximate method.

5. *Further discussion.* In comparing the present contribution with other work on the subject, the essential point to notice is the averaging process involved in equation (5). We are not trying here to make probability statements valid for *fixed* s_i^2, but are averaging over the joint probability distribution of the s_i^2, taking into account, therefore, the different values which can arise by chance in sampling from populations with fixed σ_i^2.

This averaging over the joint distribution of the s_i^2 is parallel to the step taken in Section III of Gosset's original memoir (1908) where, in effect, he starts with the distribution of t for samples with *fixed* s and then averages over the distribution of s which he has already derived earlier. He thus arrives at the unrestricted distribution of t (or, more strictly, of a quantity z, which is equal to t multiplied by a constant). This distribution forms the basis of the significance tests which he illustrates in his Section IX and of the method of deriving interval estimates for the population mean which he outlines in his Section VIII.

In the present paper the parallelism with Gosset's work may be obscured to some extent by the fact that we do not from the outset seek the probability distribution of some pivotal quantity like t, explicitly expressed. It so happens that we are able to proceed to a method of deriving an expansion for the required probability level without making explicit reference to such a quantity. Nevertheless there remains the important resemblance with Gosset's development, in that we do not confine ourselves to samples with fixed s_i^2.

This procedure stands in sharp contrast to the formulation of the problem of comparing two means, favoured by R. A. Fisher (e.g. 1941) and H. Jeffreys (1940). These writers prefer a solution which they ascribe initially to W. U. Behrens (1929). Looked at from one point of view, Behrens's paper appears to contain some gross algebraical errors. Fisher and Jeffreys, however, develop lines of argument by means of which they claim that Behrens's solution is quite justified. It seems to me difficult to say how far (if at all) any of these arguments may have been in Behrens's mind when he wrote his paper and I shall not attempt to elucidate this question here. We may, however, permit ourselves one observation about the developments according to Fisher and Jeffreys.

* Exact in the sense that it is independent of the irrelevant population parameters $\sigma_i{}^2$.

Both these writers, at some stage, limit the field of their probability inferences to a subset in which the s_i^2 are regarded as *fixed*. In order to solve the problem on these lines Jeffreys introduces an *a priori* distribution function for the unknown σ_i, following his general philosophy for dealing with such questions. Fisher, on the other hand, arrives at the same answer by a special utilization of what he terms the *fiducial* distribution of σ_i.

Jeffreys's approach here does not raise any new issues to those who are familiar with the general body of his researches on statistical inference. Fisher's justification of Behrens's solution is perhaps of more immediate interest as it raises controversial points which are important more specifically in relation to our present topic of discussion. For although Fisher's approach has been very much criticized by a number of writers, starting with M. S. Bartlett (1936), the critics have not wished to throw doubt on the whole body of results which Fisher includes under the heading of fiducial inference. The criticism has been for the most part selective, directed mainly at the way in which so-called *simultaneous* fiducial distributions of several parameters have been defined and manipulated.

I have, myself, quite definite views on these questions (particularly on the usage of the word 'fiducial') but do not feel that I need express them at any great length here. I disagree with Fisher, but this divergence of opinion must already have become apparent in the way I have defined the field within which I make my probability inferences about η. It appears to me to be quite artificial to restrict our view to one which, even in a limited sense, fixes s_i^2. It is true that, in the two-sample problem, we have to draw our inferences from the unique pair of samples observed, or, more precisely, from the statistics $\bar{x}_1$, $\bar{x}_2$, s_1^2 and s_2^2 which they provide. These statistics are our only *data* for the purpose of making inferences, but we add something to these data in the *interpretation* when we regard the samples as being drawn randomly from hypothetical normal populations. Once having embarked on this method of interpretation, we should stick to it consistently throughout. The sampling variations of s_i^2 should be taken into account only by a direct use of the probability distributions as given by our equation (1) and not by any inversion such as is involved in Fisher's conception of the fiducial distribution of σ_i^2. As we have seen, it is quite possible to make probability statements about the difference between the population means without making any reference whatever either to inverse probability or to fiducial distributions.

The distinction between the procedure which Fisher advocates and one which averages over the s^2 distributions has, of course, been stressed by most of the writers who have contributed papers on the subject, from whatever viewpoint (e.g. Bartlett, 1936, p. 566, and Yates, 1939.) What has been lacking hitherto, however, is a solution, analogous to Gosset's single sample solution, which makes complete use of the information contained in the data provided. Bartlett indicated one particular way in which probability inferences about the difference between two population means might be made, but was careful to point out that the problem of making the best possible inferences (in the theoretical sense of utilizing all the information in the data to its full extent) was still an open one. There has indeed been some doubt expressed whether a fully satisfactory solution from this point of view existed at all. I believe, however, that the one I advance above in equation (11), and develop in equation (21), meets all the requirements that one can reasonably expect.

Whatever conclusion the reader may come to on these matters, however, he will probably wish to know how, in the numerical details, this solution will differ from that of Behrens. This will be more easily seen when some tables become available, but fortunately certain comparisons can already be made. For Fisher (1941, p. 155) has provided a series expansion of the Behrens solution. In our notation, and with $k=2$, this may be written, to order $1/f_i$, as follows:

$$h(s^2) = \xi\sqrt{(\lambda_1 s_1^2 + \lambda_2 s_2^2)}\left[1 + \frac{(1+\xi^2)}{4}\frac{\left(\frac{\lambda_1^2 s_1^4}{f_1} + \frac{\lambda_2^2 s_2^4}{f_2}\right)}{(\lambda_1 s_1^2 + \lambda_2 s_2^2)^2} + \left(\frac{1}{f_1} + \frac{1}{f_2}\right)\frac{\lambda_1\lambda_2 s_1^2 s_2^2}{(\lambda_1 s_1^2 + \lambda_2 s_2^2)^2}\right]. \tag{32}$$

Even to this order, this differs from our equation (21) in the inclusion of an extra term. In other words, although the two solutions are the same when samples are large enough to adopt the large-sample normal approximation, they differ immediately we take into account the first corrective term, i.e. they differ as soon as we begin to attach any importance to 'Studentization'.

6. *An interval estimate for* η. We have shown in §§ 2 and 3 how to calculate a value $h(s_1^2, s_2^2, \ldots, s_k^2, P)$, depending on the observed variances $s_1^2, s_2^2, \ldots, s_k^2$, such that the probability is P that $(y-\eta) < h(s_1^2, s_2^2, \ldots, s_k^2, P)$. This provides a method of testing the consistency of an observed y with a prescribed value η.

When the question is not whether any particular given η is contradicted by the data, but rather one of estimating η and at the same time of providing a measure of the uncertainty of the estimate, the further step required is immediate. For, as in the case of a single sample, the order of the words in our probability statement can be changed so that it becomes—the probability is P that η is greater than $\{y - h(s_1^2, s_2^2, s_k^2, P)\}$. An interval estimate for η is then obtained by taking two levels P_1 and P_2 for P. Thus the probability is $(P_1 - P_2)$ that η lies between $\{y - h(s_1^2, s_2^2, \ldots, s_k^2, P_1)\}$ and $\{y - h(s_1^2, s_2^2, \ldots, s_k^2, P_2)\}$.

If $P_2 = (1-P_1)$ the range will be symmetrically placed about y. Thus, for example, if $P_1 = 0{\cdot}95$ and $P_2 = 0{\cdot}05$, the chance will be 90 % that η lies within the range

$$y \pm 1{\cdot}6449\sqrt{(\Sigma\lambda_i s_i^2)}\left[1 + \frac{1+(1{\cdot}6449)^2}{4}\frac{\left(\Sigma\frac{\lambda_i^2 s_i^4}{f_i}\right)}{(\Sigma\lambda_i s_i^2)^2} + \text{etc.}\right]. \tag{33}$$

It may be noted, incidentally, that this range is always narrower than similar ranges calculated from Behrens's solution.

REFERENCES

BARTLETT, M. S. (1936). The information available in small samples. *Proc. Camb. Phil. Soc.* **32**, 560–6.

BEHRENS, W. U. (1929). Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen. *Landw. Jb.* **68**, 807–37.

FISHER, R. A. (1941). The asymptotic approach to Behrens's integral, with further tables for the d test of significance. *Ann. Eugen., Lond.*, **11**, 141–72.

JEFFREYS, H. (1940). Note on the Behrens-Fisher formula. *Ann. Eugen., Lond.*, **10**, 48–51.

'STUDENT' (1908). The probable error of a mean. *Biometrika*, **6**, 1–25.

WELCH, B. L. (1936). Specification of rules for rejecting too variable a product, with particular reference to an electric lamp problem. *J. Roy. Statist. Soc. Suppl.* **3**, 29–48.

WELCH, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, **29**, 350–62.

YATES, F. (1939). An apparent inconsistency arising from tests of significance based on fiducial distributions of unknown parameters. *Proc. Camb. Phil. Soc.* **35**, 579–91.