DISTRIBUTIONS RELATED TO COMPARISON OF TWO MEANS AND TWO REGRESSION COEFFICIENTS

By Uttam Chand[1]

University of North Carolina

Summary. We consider here the relative merits of different statistics available for testing two means or two regression coefficients in relation to one-sided (asymmetric) and two-sided (symmetric) alternatives in case of unequal population variances. In so far as the Behrens-Fisher statistic is concerned we confine ourselves to the consideration of the behavior of its probability of Type I error in repeated sampling from populations with a fixed value of the unknown ratio of variances. In connection with the tests between two means, the present study takes its point of departure from the existing tests and investigates the question of utilizing an approximately determinate knowledge about the unknown ratio of variances. In connection with the comparison of two regression coefficients and also of two linear regression functions, we consider the effect of two concomitant sources of variation, viz., the unknown ratio of residual variances and the ratio of the sums of squares of the fixed variates, on the probability of Type I and Type II errors of certain well known statistics.

1. Introduction. Consider two independent samples $x_1 \cdots x_{n_1+1}$ and $x'_1 \cdots x'_{n_2+1}$ drawn from two normal populations with means m_1 and m_2, variances σ_1^2 and σ_2^2. Let $K = \sigma_1^2/\sigma_2^2$. If K is known and $m_1 = m_2$, the quantity

$$t_K = \frac{\bar{x} - \bar{x}'}{\left[\frac{S(x-\bar{x})^2 + KS'(x'-\bar{x}')^2}{n_1+n_2}, \left(\frac{1}{n_1+1} + \frac{1}{K(n_2+1)}\right)\right]^{\frac{1}{2}}}$$

(t_1 is Fisher's t) is distributed according to "Student's" distribution with $n_1 + n_2$ d.o.f.[2] and for the "Student's" hypothesis $H_0: m_1 = m_2$ provides a uniformly most powerful test against an asymmetric alternative $H_1: m_1 >$ (or $<$)m_2 and a type B_1 test against a symmetric alternative $H_2: m_1 \neq m_2$. If K is unknown certain approximate and exact tests have been suggested from time to time to meet this situation.

Welch [1], [2] using an approximation to the distribution of t_1 was the first to point out that if K is unknown and we assume it to be equal to unity, then the probability of Type I error of the t_1-test is subject to large variations as K varies from 0 to ∞. He also pointed out that the statistic

$$v = (\bar{x} - \bar{x}') \left[\frac{S(x-\bar{x})^2}{n_1(n_1+1)} + \frac{S'(x'-\bar{x}')^2}{n_2(n_2+1)}\right]^{-\frac{1}{2}}$$

[1] Now Assistant Professor of Mathematical Statistics at Boston University.

[2] Degrees of freedom.

which does not have "Student's" distribution for $K = 1$, has the advantage that its probability of Type I error is subject to less variation with respect to K. His approximate results were later confirmed by Hsu [3] who obtained the distribution of quantities $u_1(=t_1^2)$ and $u_2(=v^2)$ and also showed that these tests are unbiased in the sense of Neyman and Pearson. Hsu concluded on the basis of his investigations that when the sample sizes are equal and not very small, we may safely use $u_1(=u_2)$ as if K were unity. This also had been pointed out by Welch.

If on the basis of past experience some approximate value k of K were available, one would like to know if such a choice in some rough neighborhood of K would in any way improve the claim of $t_k(=t_K$ for $K = k)$ for the hypothesis $m_1 = m_2$. The distribution of this generic quantity $t_k\left(=t_1 \text{ for } k = 1; =v \text{ for } k = \dfrac{n_1(n_1+1)}{n_2(n_2+1)}\right)$ will be obtained in Section 2.1. It will be shown that variation in the probability of Type I error of t_k with respect to K for any k except when $t_k = v$, is essentially similar in character to that of t_1^2 [3] and is very sensitive in a neighborhood of K in which one would very often be interested (Section 2.4). This is also true of the behavior of the power function of t_k with respect to K. Consequently a t_k type of statistic will be unsuitable in general for utilizing an approximately determinate knowledge of K.

It is not possible to infer directly from Hsu's work on the relative merits of t_1 and v in relation to asymmetric aspects of "Student's" hypothesis. His basic conclusions as regards unbiasedness and the nature of variations in Type I error in the symmetric case also hold for the asymmetric case except that the Type I variations in t_1 and v are less for asymmetric than for symmetric comparisons (Section 2.5 and Table II). Furthermore it appears (Section 2.5 and Table III) that with respect to the variations of K both the asymmetric and symmetric power functions of t_1 are likely to be more sensitive than those of v. Since for equal d.o.f. both the asymmetric probability of Type I error and power function are insensitive to the vagaries of the 'nuisance' parameter K, there is an a fortiori reason for using $v(=t_1)$ as if K were unity.

Scheffé [4] considered the statistic

$$S = (\bar{x} - \bar{x}')\left(\sum_{i=1}^{n_1+1} \frac{(u_i - \bar{u})^2}{n_1(n_1+1)}\right)^{-\frac{1}{2}} \qquad (n_1 \leq n_2),$$

(equivalent to paired difference t when $n_1 = n_2$) where $u_i = x_i - \left(\dfrac{n_1+1}{n_2+1}\right)^{\frac{1}{2}} x'_i$ and where it is assumed that the variates in each sample have been randomized. This is essentially a "Student's" t comparison based on n_1 d.o.f. and as shown by Scheffé it is impossible to get a suitable statistic with the t-distribution with more than n_1 d.o.f. The statistic v has the t-distribution only when $K = \infty$ (n_1 d.o.f.), $K = 0$ (n_2 d.o.f.) and $K = \dfrac{n_1(n_1+1)}{n_2(n_2+1)}$ ($n_1 + n_2$ d.o.f.). For any given n_1, n_2, K and P we can solve $P = P(v \geq t_0 \mid H_0)$ for t_0 and thus indirectly obtain from the tabulated values of the t-distribution the number of 'effective' d.o.f. which will thus adjust v to any preassigned level of significance. We try to show in Section 2.6 that in situations where some approximate knowledge of K is available, the statistic v seems to have a decided advantage over any other statistic having the t-distribution. We show by actual computations that Welch's formula [2] provides a conservative estimate for the effective d.o.f. in the light of which this comparison will be considered.

The Behrens-Fisher fiducial test employing the statistic d [5], [6], which has essentially the same structural form as v, has given rise to much controversy essentially because of inconsistencies arising from tests of significance based on the fiducial distribution of unknown parameters. We attempt to show in Section 2.7 that the fiducial test in general is 'conservative' in detecting significant results in repeated sampling from populations with a fixed value of the unknown ratio of variances.

In the case of comparison of two regression coefficients when the residual variances are unequal, we are faced with a similar type of problem. Consider two samples $y_\mu \mid x_\mu$ and $y'_\nu \mid x'_\nu$ ($\mu = 1, \cdots, n_1 + 1$; $\nu = 1, \cdots, n_2 + 1$), where x_μ and x'_ν are fixed and y_μ and y'_ν are normally and independently distributed according to $N(\alpha_1 + \beta_1(x_\mu - \bar{x}), \sigma_1^2)$ and $N(\alpha_2 + \beta_2(x'_\nu - \bar{x}'), \sigma_2^2)$ respectively. For the hypothesis $\beta_1 = \beta_2$ when the alternatives do not specify anything except $\beta_1 > \beta_2$ or $<\beta_2$; or $\beta_1 \neq \beta_2$ we shall consider the merits of statistics t^* and v^* which correspond to statistics t_1 and v for the two means. While the statistic t^* is sensitive to the variation of both $K = \sigma_1^2/\sigma_2^2$ and w, the ratio of the sums of squares of the fixed variates, the statistic v^* is insensitive to the variation of both. Barankin[3] has extended Scheffé's test to the comparison of two regression coefficients under the above assumptions. The statistic proposed by Barankin has Student's distribution with $n_1 - 1$ d.o.f. ($n_1 \leq n_2$) and provides the only exact unbiased test so far known. While Scheffé's test for the comparison of two means and Barankin's test for the comparison of two regression coefficients should not be used when K is known and were never intended to utilize any available approximate information about K, the question of investigating into the possibility of using v^* in the latter situation is not without interest (Section 3). In Section 4 we consider the hypothesis of equality of two linear regression functions viz., H_0: $\alpha_1 = \alpha_2, \beta_1 = \beta_2$ when the alternatives do not specify anything except $\alpha_1 \neq \alpha_2$ or $\beta_1 \neq \beta_2$.

In studying the behavior of the power function and the probability of Type I error of certain statistics under discussion we have made full use of Hsu's method and consequently only essential details have been given here.

2. Hypothesis of equality of two means when variances are unequal

2.1. *The distribution of t_k for any values of n_1 and n_2.* Consider the test function $t_k (= t_K$ for $K = k$; Section 1) where k is some inexact value of K. This can be

[3] E. W. Barankin, "Extension of the Romanovsky-Bartlett-Scheffe test" *Proc. Berkeley Symposium on Math. Stat. and Prob.*, University of California Press, 1949, pp. 433–449.

put in the form of $t_k = (\xi + \delta)(b\chi_1^2 + c\chi_2^2)^{-\frac{1}{2}}$ where ξ is $N(0, 1)$ and the χ^2's have independent χ^2-distribution with n_1 and n_2 d.o.f., and where

$$\begin{aligned}
\delta &= (m_1 - m_2)\left(\frac{\sigma_1^2}{n_1 + 1} + \frac{\sigma_2^2}{n_2 + 1}\right)^{-\frac{1}{2}},\\
b &= (K/k)(n_1 + n_2)^{-1}[k(n_2 + 1) + n_1 + 1]\,[K(n_2 + 1) + n_1 + 1]^{-1},\\
c &= (n_1 + n_2)^{-1}[k(n_2 + 1) + n_1 + 1]\,[K(n_2 + 1) + n_1 + 1]^{-1},\\
b/c &= K/k.
\end{aligned}$$

In what follows we shall omit the subscript k from t_k. The joint probability element of ξ, χ_1^2 and χ_2^2 is given by

$$dF(\xi_1, \chi_1^2, \chi_2^2) = \tfrac{1}{4}(2\pi)^{-\frac{1}{2}}[\Gamma(n_1/2)\Gamma(n_2/2)]^{-1}e^{-\frac{1}{2}(\xi_1^2+\chi_1^2+\chi_2^2)}(\chi_1^2/2)^{n_1/2-1} (\chi_2^2/2)^{n_2/2-1}\,d\xi d(\chi_1^2)\,d(\chi_2^2).$$

We transform to new variables t, r and θ by the relations

$$\begin{aligned}
\xi + \delta &= t(b\chi_1^2 + c\chi_2^2)^{\frac{1}{2}}, &&\\
b\chi_1^2 &= r^2\cos^2\theta && (0 \le \theta \le \pi/2),\\
c\chi_2^2 &= r^2\sin^2\theta && (-\infty \le r \le +\infty),
\end{aligned}$$

and integrate out r. To integrate out θ we put $z = \sin^2\theta$ if $b < c$ and $z = \cos^2\theta$ if $b > c$. This reduces the integration w.r.t. θ to a series of hypergeometric integrals. We finally have the following form for the frequency function of t_k:

$$g(t) = \frac{e^{-\delta^2/2}(b/c)^{n_2+1/2}c^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n_1+n_2}{2}\right)}\sum_{r=0}^{\infty}\frac{(\delta t)^r(2b)^{r/2}\Gamma\left(\frac{n_1+n_2+r+1}{2}\right)}{h!(1+bt^2)^{\frac{n_1+n_2+r+1}{2}}} \cdot F\left(\frac{n_1+n_2+r+1}{2}, \frac{n_2}{2}, \frac{n_1+n_2}{2}, \frac{1-b/c}{1+bt^2}\right), \tag{2.1.1}$$

where F denotes the hypergeometric function. As a check if we put $b = c = (n_1 + n_2)^{-1}$, we get the frequency function of non-central t for $n_1 + n_2$ d.o.f. For the case $b > c$ we have only to interchange b with c and n_1 with n_2.

The null distribution of $t_k(\delta = 0)$ is an even function of t_k; consequently the forms of the single and two-equal-tailed probability of Type I error will be the same except for the constant $\frac{1}{2}$. If we let $\beta_1(\delta, K, k, n_1, n_2) = \int_{t_0}^{\infty} g(t)\,dt$ denote the single upper tail power function of t_k, from (2.1.1) we obtain

$$\beta_1(\delta, K, k, n_1, n_2) = \tfrac{1}{2}e^{-\delta^2/2}(K/k)^{n_2/2}\sum_{h=0}^{\infty}\sum_{r=0}^{\infty} \cdot\frac{(\delta^2/2)^{r/2}\Gamma\left(\frac{n_2}{2}+h\right)\left(1-\frac{K}{k}\right)^h}{\Gamma\left(\frac{n_2}{2}\right)h!\frac{r}{2}!} I_{x_0}\left(\frac{n_1+n_2}{2}+h, \frac{r+1}{2}\right), \tag{2.1.2}$$

where $x_0 = (1 + bt^2)^{-1}$ and $I_{x_0}(p, q)$ is the incomplete beta ratio. To obtain the two equal tailed power function $\beta_2(\delta, K, k, n_1, n_2)$ we need only change r into $2r$ and omit the factor $\frac{1}{2}$.

2.2. *Distribution of t_k for even values of n_1 and n_2*. (For notation refer to Section 2.1). When n_1 and n_2 are even, the method of characteristic functions yields a single infinite series for the distribution of t_k, and when $\delta = 0$ this series reduces to $\frac{n_1 + n_2}{2}$ terms. The characteristic function of $X = b\chi_1^2 + c\chi_2^2$ is given by $\phi(\tau) = (1 - 2bi\tau)^{-n_1/2}(1 - 2ci\tau)^{-n_2/2}$. To obtain the form of the frequency function of X we make use of the inversion theorem and integrate round a standard contour in the lower half of the complex plane. The distribution of t_k can then be obtained from the joint probability element of ξ and X. We obtain the following form for the single tailed power function of t_k:

$$\begin{aligned}\beta_1(\delta, K, k, n_1, n_2) = \tfrac{1}{2}e^{-\delta^2/2}\sum_{r=0}^{\infty}\frac{(\delta^2/2)^{r/2}}{\frac{r}{2}!}\Bigg[\left(\frac{K}{K-k}\right)^{n_2/2}\sum_{h=0}^{(n_1/2)-1}&\frac{(-1)^h\Gamma\left(\frac{n_2}{2}+h\right)}{\Gamma\left(\frac{n_2}{2}\right)h!}\left(\frac{k}{K-k}\right)^h I_{x_0}\left(\frac{n_1}{2}-h, \frac{r+1}{2}\right)\\ &+ (-1)^{n_1/2}\left(\frac{k}{K-k}\right)^{n_1/2}\sum_{h=0}^{(n_2/2)-1}\frac{\Gamma\left(\frac{n_1}{2}+h\right)}{\Gamma\left(\frac{n_1}{2}\right)h!}\\ &\cdot\left(\frac{K}{K-k}\right)^h I_{x_0'}\left(\frac{n_2}{2}-k, \frac{r+1}{2}\right)\Bigg] \quad (K \geq k)\end{aligned} \tag{2.2.1}$$

where x_0 has been defined in the previous section and $x_0' = (1 + ct_0^2)^{-1}$.

2.3. *Unbiasedness of a test based on t_k*. Since the single and two tailed forms of the power function of t_k (Section 2.1) are essentially the same functions of the standardised 'distance' δ, following Hsu [3] we can show that $\frac{\partial\beta_1}{\partial\delta} \geq 0$ and $\frac{\partial\beta_2}{\partial\delta} \geq 0$ for any fixed K and k; and consequently such a generic type of statistic provides an unbiased test both against symmetric and asymmetric alternatives.

2.4. *Variations in the power function and the probability of Type I error of t_k*. For the case $k = 1$, Hsu [3] has already shown that the probability of Type I error of the statistic t_1^2 is subject to large variations w.r.t. K. He also pointed out that the behavior of the derivative of its power function w.r.t. K for fixed δ was similar to that of its probability of Type I error w.r.t. K. We shall presently see that t_k also shares this property with t_1^2.

In the first place one would like to know if any choice of k in a small neighborhood of K would stabilize the variations in the Type I error of t_k to such an extent as to make it approximately insensitive to that difference between k and

K. With this end in view we shall examine the nature of variations in the probability of Type I error of t_k w.r.t. K for any fixed k.

From (2.1.2) by putting $\delta = 0$ we obtain

$$\begin{aligned} P = P(t_k \geq t_0) &= \tfrac{1}{2}(K/k)^{n_2/2} \sum_{h=0}^{\infty} \Gamma\left(\frac{n_2}{2} + h\right)(1 - K/k)^h \\ &\cdot \left(\Gamma\left(\frac{n_2}{2}\right)\Gamma(h+1)\right)^{-1} I_{x_0}\left(\frac{n_1+n_2}{2} + h, \tfrac{1}{2}\right). \end{aligned} \tag{2.4.1}$$

We now differentiate (2.4.1) and after simplification obtain

$$\frac{dP}{dK} < C_1(K/k)^{-1}[n_2(n_2+1) - n_1(n_1+1)/k][K(n_2+1) + n_1 + 1]^{-1} \quad (K < k).$$

Similarly

$$\frac{dP}{dK} > C_2[n_2(n_2+1) - n_1(n_1+1)/k][K(n_2+1) + n_1 + 1]^{-1} \qquad (K > k),$$

where C_1 and C_2 are certain positive constants independent of K and k.

If $k = \dfrac{n_1(n_1+1)}{n_2(n_2+1)}$ we have

$$\frac{dP}{dK} \lesseqgtr 0$$

for $K \lesseqgtr k$.

This is the case when t_k is identical with the statistic v defined in Section 1 and the probability of Type I error curve expressing P as a function of K has a minimum at this point: for $n_1 < n_2$ the minimum occurs for a value of $K < 1$ and vice versa. And since v is known to be insensitive to the variation of K [3], therefore t_k is insensitive to the variation of K for this value of k.

For any other assumed value of k the curve either starts decreasing from $K = \infty$ or from $K = 0$ to the point where $K = k$ depending upon the values of n_1 and n_2. In each case the ordinate of the curve continues to decrease for some distance; it may decrease to a minimum and then start increasing or else decrease indefinitely. For fixed δ the power function of t_k also has a minimum when $K = k = \dfrac{n_1(n_1+1)}{n_2(n_2+1)}$; and for any other k the behavior of its power function is similar to that of its probability of Type I error. For the case $k = 1$ numerical values of the single and two-tailed values of the probability of Type I error and power function for different values of n_1 and n_2 and K are given in Tables II and III (Section 2.5).

In certain practical situations it may happen for example that on the basis of past experience one can determine k so that $\frac{1}{2} \leqq |k - K| \leq 2$. The question arises: how much is t_k sensitive to such a neighborhood for any k, K, n_1 and n_2? That it is hard to provide a practically useful answer to this question will be apparent from the nature of the distribution of t_k, which depends both on K and k and not merely on their ratio. The following Table I will indicate how in such a small neighborhood $P(t_k \geq t_0)$ can be in serious error in two different directions.

2.5. *Statistics t_1 and v in relation to asymmetric and symmetric aspects of "Student's" hypothesis.* Statistics t_1 and v are special cases of t_k and the behavior of their probability of Type I error and power function has already been discussed (Sections 2.3 and 2.4). In this section we compare the single-tailed and two tailed values of the probability of Type I error and power function in the light of several particular examples. In all these calculations e.g. in $P(t \geq t_0)$ and

TABLE I

Variations in $P(t_k \geq t_0)$ with respect to k for fixed K

($K = 5$; $n_1 = 2$; $n_2 = 4$; $t_0 = 2.447$)

$k =$	1	2	3	4	5	6	7
	.1129	.0936	.0749	.0607	.05	.0418	.0355

TABLE II

Variations in the symmetric and asymmetric probability of Type I error of v and t_1 in relation to the unknown ratio of variances K

	K	0	.125	.5	1	2	4	8	16	∞	% point of tabulated t_1
	$n_1 = n_2 = 3$	.074	.0633	.0504	.05	.0504	.0568	.0633	.0691	.074	single tailed 5%
$v = t$	"	.092	.0681	.0525	.05	.0525	.0597	.0681	.0770	.092	two-tailed 5%
	"	.034	.0181	.0110	.01	.0110	.0138	.0181	.0227	.034	two-tailed 1%
	$n_1 = 4$, $n_2 = 16$	.0112	.0129†	.0142	.0195	.0227	.0265	.0293	.0305	.0324	single tailed 1%
v	"	.012	.0161†	.0197	.0238	.0294	.0359	.0407	.0433	.0465	two-tailed 1%
	$n_1 = 8$, $n_2 = 4$	.075	.0687	.0598	.0543	.0541	.0511‡	.0521	.0531	.056	single tailed 5%
	$n_1 = 4$, $n_2 = 16$	.00011	.00043	.00310	.01	.0221	.0483	.0793	.0864	.133	single tailed 1%
t	"	.00007	.00031	.00244	.01	.0310	.0592	.1169	.1544	.222	two-tailed 1%
	$n_1 = 8$, $n_2 = 4$	.1342	.1056	.0710	.05	.0368	.0287	.0246	.0224	.0204	single tailed 5%

† $P = .01$ when $K = .074$
‡ $P = .05$ when $K = 3.6$

$P(|t| \geq t_0')$, t_0 refers to the single and t_0' to the two tailed values of Fisher's t for the appropriate number of d.o.f. Tables II and III give the approximate values for the probability of Type I error and the power function respectively both against symmetric and asymmetric alternatives.

For equal sample sizes ($v = t_1$) the Type I error and power function curves, representing probability of Type I error and power function as a function of K, have a minimum when K is unity and a maximum occurs when K is either zero or infinity. Maximum values of the probability of Type I error for several equal sample sizes are given in Table IV. It appears that for equal sample sizes the probability of Type I error and the power function are likely to be insensitive to the variation of K. We also notice in this connection that while the single tailed values of the probability of Type I error are less than those of the two tailed values, the values of the two tailed power function for $\delta = 1$ are less than the corresponding single tailed values. This appears to be true also for the statistic v when $n_1 \neq n_2$. For unequal sample sizes also the probability of Type I error and the power function of t_1 are likely to be more sensitive to the variation of K than those of v. It may be pointed out in the sequel that while it is recognized that for unequal d.o.f. a fair comparison of the probability of Type I error and the power function of v with those of t_1 ought to adjust v and t_1 to the same level of significance, namely the same maximum (for all K) probability of Type I error, this would not alter our conclusions about the sensitive nature of t_1.

TABLE III[4]

Variations in the asymmetric and symmetric power function of t_1 and v corresponding to the 5% point of tabulated $t_1(\delta = 1)$

$K =$	0	.5	1	2	∞	
$n_1 = n_2 = 3$	.189	.141	.137	.141	.189	symmetric
$v = t_1$	.269	.229	.225[5]	.229	.269	asymmetric
$n_1 = 8, n_2 = 4$	.354	.262	.152	.112	.063	symmetric
t_1	.428	.294	.242[5]	.194	.122	asymmetric
$n_1 = 8, n_2 = 4$	.208	.196	.162	.156†	.168	symmetric
v	.286	.259	.247	.244‡	.255	asymmetric

† minimum of .152 is reached for $K = 3.6$.
‡ minimum of .242 is reached for $K = 3.6$.

TABLE IV

Maximum probability of Type I error of $v(= t_1)$ for equal degrees of freedom

	Symmetric		Asymmetric	
$n_1 + 1 = n_2 + 1$	5%	1%	5%	1%
7	.0721	.0224	.0625	.0182
9	.0668	.0193	.0595	.0162
11	.0635	.0173	.0576	.0150
15	.0598	.0152	.0555	.0136
21	.0569	.0137	.0538	.0125

2.6. *Statistic v, Scheffé's test and paired difference t.* If K is known, v or Scheffé's statistic S should not be used. If K is unknown, S is an ingenious device for getting a Student's t with $\min(n_1, n_2)$ d.o.f. and provides the only exact unbiased test so far known. In such a situation since nothing is known about K, a fair comparison of the power function of S with v ought to adjust v to the same maximum probability of Type I error for all K (maximum will occur for $K = 0$ or $K = \infty$ according as $n_1 \gtrless n_2$); and at such a maximum significance level it is

[4] The author acknowledges with pleasure the help given in the preparation of this table by Miss Elizabeth Shuhany of the Statistical Laboratory, Boston University.

[5] Values taken from [7].

recognized that v cannot be uniformly better than S. For samples of equal size n the use of the paired difference t with $n - 1$ d.o.f. (equivalent to S when $n_1 = n_2$; Section 1) provides a suitable test for two reasons: (i) it is exact and (ii) as shown by Walsh [8] has a high power efficiency.

If any approximate a priori information about K is available, v appears to be the only suitable statistic to utilize such information. While S was not intended to cope with such a situation, t_k (Section 2.4) has been shown to be unsuitable. Since v is insensitive to the variation of K, we shall not be far wrong in using 'effective' d.o.f. based upon an assumed value k of K satisfying some such relation as $\frac{1}{2} \leq | k - K | \leq 2$. The effective d.o.f. of v as given by Welch [1] and as given by $P = P(v \geq t_0)$ or by $P = P(| v | \geq t_0')$ for fixed P (listed in Table V as calculated d.o.f.) are identical for $K = 0, 1$, an ∞ $(n_1 = n_2)$ and (ii) $K = 0, \dfrac{n_1(n_1 + 1)}{n_2(n_2 + 1)}$, and ∞ $(n_1 \neq n_2)$. For other values of K it appears from Table V that Welch's formula errs on the conservative side. The effective number of d.o.f. vary between $n_1 + n_2$ and $\min(n_1, n_2)$ (cf. d.o.f. for S). Consequently in the absence of any

TABLE V

Adjusted power function of v in the light of 'effective' degrees of freedom

Sample Size	Adjusted asymmetric power function of v for probability of Type I error of .05								Effective d.o.f.							
	$\delta = 1$				$\delta = 2$				Calculated				Welch's formula			
	$K = 0$	.125	4	∞	$K = 0$	.125	4	∞	$K = 0$	.125	4	∞	$K = 0$	.125	4	∞
$n_1 + 1 = n_2 + 1 = 3$	.174	.204	.204	.174	.384	.476	.476	.384	2	3.36	3.36	2	2	2.94	2.94	2
$n_1 + 1 = n_2 + 1 = 7$	.225	.236	.236	.225	.550	.581	.581	.550	6	9.14	9.14	6	6	8.82	8.82	6
$n_1 + 1 = 9$; $n_2 + 1 = 5$	.210	.227	.242	.233	.504	.556	.594	.572	4	6.50	11.90	8	4	5.14	11.90	8

best unbiased test and in the light of any approximate information about K it would appear that v has a decided advantage over any other statistic.

2.7. *The Behrens-Fisher test in repeated sampling.* Consider the statistic

$$d = (\bar{x} - \bar{x}')\,(s_1^2 + s_2^2)^{-\frac{1}{2}} = t_1 \sin \theta - t_2 \cos \theta,$$

where s_1^2 and s_2^2 are the unbiased estimates of the variances of the means $\bar{x}$ and $\bar{x}'$ respectively, t_1 and t_2 have independent "Student's" distributions with n_1 and n_2 d.o.f. respectively, and $\tan \theta = s_1/s_2$. On the basis of the "fiducial" distribution of σ_1^2 and σ_2^2 Fisher [6] regards d as a "mixture" of t_1 and t_2 with constant coefficients. It is to be noted that if s_1 and s_2 are fixed in the classical sense t_1 and t_2 have independent normal conditional distributions with zero means and variances σ_1^2/s_1^2 and σ_2^2/s_2^2 respectively; and if s_1 and s_2 vary in their own distribution d is identical with v (Section 1).

Neyman [9] considered the integral of the joint probability law of $\bar{x}$, $\bar{x}'$, s_1^2, s_2^2 over the set $\dfrac{| \bar{x} - \bar{x}' |}{\sqrt{s_1^2 + s_2^2}} \leq t_1 \sin \theta - t_2 \cos \theta$ where the quantity on the right also depends upon s_1 and s_2 and is the quantity d tabulated by Sukhatme [10], [11].

Neyman showed in particular that if pairs of normal populations with different K are sampled ($n_1 + 1 = 13$, $n_2 + 1 = 7$), then the relative frequency of correct statements about $m_1 - m_2$ based on the 5% points of d will not be equal to the expected .95 and will vary with K.

We consider here the following similar type of question: what is the nature of discrepancies that will arise in the probability of Type I error by the repeated use of the Behrens-Fisher test in sampling from two normal populations? We observe that since d and v have the same structural form, the appropriate probability of Type I error in such a situation will be given by the probability integral of v (Sections 2.2 and 2.5).

TABLE VI

Minimum and maximum† values of $P(\mid v \mid \geq d_0)$ for different values of K

K	0	.05	1	2	∞	d_0
$n_1 + 1 = n_2 + 1 = 7$	.05	.0321	.0307	.0321	.05	2.447
	.0508	**.0329**	**.0313**	**.0329**	**.0508**	2.435
$n_1 + 1 = n_2 + 1 = 9$	.05	.0362	.0346	.0362	.05	2.306
	.0512	**.0367**	**.0358**	**.0367**	**.0512**	2.292
$n_1 + 1 = n_2 + 1 = 13$	.05	.0405	.0396	.0405	.05	2.179
	.0507	**.0434**	**.0403**	**.0434**	**.0507**	2.170
$n_1 + 1 = 7, n_2 + 1 = 13$	.0307	.0281	.0317	.0393	.05	2.447
	.05	**.0460**	**.0516**	**.0597**	**.0720**	2.179
$n_1 = n_2 = \infty$	.05	.05	.05	.05	.05	1.960

† maximum values have been indicated in bold type.

We observe that $P(\mid v \mid \geq x)$ is a monotone decreasing function of x for any fixed K, n_1 and n_2. Furthermore for fixed x, n_1 and n_2 we have $\frac{dP}{dK} \gtreqless 0$ for (i) $K \gtreqless 1$, $n_1 = n_2$ and (ii) $K \gtreqless \frac{n_1(n_1+1)}{n_2(n_2+1)}$, $n_1 \neq n_2$. Table VI gives the minimum and maximum values of $P(\mid v \mid \geq d_0)$ for different values of K where d_0 corresponds to the highest and lowest value of tabulated d. It appears that for equal sample sizes the minimum probability of Type I error is less than .05 and will converge to .05 when K is either infinity or zero. The maximum probability of Type I error converges to a value slightly higher than .05. This probability also converges to .05 with increasing size of equal samples for every K. For unequal sample sizes e.g. $n_1 < n_2$, the minimum values converge to .05 when $K = \infty$ and if $n_1 > n_2$, this convergence takes place when $K = 0$. The maximum values are both greater and less than .05.

3. Hypothesis of equality of regression coefficients when residual variances are unequal.

3.1. *Unbiasedness of tests based on statistics t^* and v^*.* Consider

$$t^* = (b_1 - b_2)\left[\frac{S(y-Y)^2 + S'(y'-Y')^2}{n_1+n_2-2}\left(\frac{1}{M_1}+\frac{1}{M_2}\right)\right]^{-\frac{1}{2}}$$

and

$$v^* = (b_1 - b_2)\left[\frac{S(y - Y)^2}{M_1(n_1 - 1)} + \frac{S'(y' - Y')^2}{M_2(n_2 - 1)}\right]^{-\frac{1}{2}},$$

where b_1 and b_2 are regression coefficients calculated from independent samples; Y and Y' are the sample regression functions; $M_1 = S(x - \bar{x})^2$ and $M_2 = S'(x' - \bar{x}')^2$. Under the assumptions of Section 1 these two quantities are distributed as

$$t^* = (\xi + \Delta)\ (\mu_1\chi^2_{1,n_1-1} + \mu_2\chi^2_{2,n_2-1})^{-\frac{1}{2}},$$
$$v^* = (\xi + \Delta)\ (\lambda_1\chi^2_{1,n_1-1} + \lambda_2\chi^2_{2,n_2-1})^{-\frac{1}{2}},$$

respectively, where ξ is $N(0, 1)$ and the χ^2's have independent χ^2-distribution with d.o.f. indicated in the second subscripts, and where

$$\begin{aligned}
M_1/M_2 &= w,\\
\mu_1 &= K(w + 1)\ (K + w)^{-1}\ (n_1 + n_2 - 2)^{-1},\\
\mu_2 &= (w + 1)\ (K + w)^{-1}\ (n_1 + n_2 - 2)^{-1},\\
\frac{\mu_1}{\mu_2} &= K,\\
\Delta &= (\beta_1 - \beta_2)\left(\frac{\sigma_1^2}{M_1} + \frac{\sigma_2^2}{M_2}\right)^{-1},\\
\lambda_1 &= K(K + w)^{-1}\ (n_1 - 1)^{-1},\\
\lambda_2 &= w(K + w)^{-1}\ (n_2 - 1)^{-1},\\
\frac{\lambda_1}{\lambda_2} &= (K/w)\,\frac{n_2 - 1}{n_1 - 1}.
\end{aligned}$$

Consequently these two statistics have the same basic distribution as obtained previously for t_k (Section 2.1) and their power functions are monotone increasing functions of the standardized 'distance' Δ for fixed values of K, w, n_1 and n_2. While the statistic t^* has "Student's" distribution with $n_1 + n_2 - 2$ d.o.f. whenever $K = 1$, the statistic v^* is only so distributed when $K = w(n_1 - 1)(n_2 - 1)^{-1}$.

3.2. *Variations in the probability of Type I error and power function of t^* and v^*.* The behavior of the partial derivatives of the probability of Type I error and the power function of t^* and v^* w.r.t. K and also in relation to w is essentially the same. For purposes of illustration we shall only consider the behavior of the probability of Type I error. We shall presently see that for the hypothesis $\beta_1 = \beta_2$ (cf. "Student's" hypothesis $m_1 = m_2$) while t^* is sensitive to the variation of K and w, v^* is insensitive to both.

3.2.1. *Variations w.r.t. K for fixed w.* Remembering that the χ^2's in the denominator of t^* have respectively $n_1 - 1$ and $n_2 - 1$ d.o.f., we can write down $P(t^* \geq t_0)$ from the corresponding form for t_1 (Section 2.3). After simplification we obtain

$$\text{(3.2.1.1)} \qquad \frac{\partial P}{\partial K} < L_1[(n_2 - 1) - w(n_1 - 1)]\ (K + w)^{-1}/K \qquad (K < 1),$$

where $z_0 = (1 + \mu_1 t_0^2)^{-1}$. If we make use of the relation $P(n_1, n_2, M_1, M_2, K) = P(n_2, n_1, M_2, M_1, K^{-1})$ in (3.2.1.1) we obtain

$$\text{(3.2.1.2)} \qquad \frac{\partial P}{\partial K} > L_2(K + w)^{-1}[(n_2 - 1) - w(n_1 - 1)] \qquad (K > 1),$$

where L_1 and L_2 are certain positive constants independent of M_1, M_2 and K.

Similarly for the statistic v^* we have

$$\text{(3.2.1.3)} \qquad \frac{\partial P}{\partial K} < D_1(K\phi)^{-1}[(n_2 - 1) - w(n_1 - 1)\phi]/(K + w) \qquad (K\phi < 1)$$

and

$$\text{(3.2.1.4)} \qquad \frac{\partial P}{\partial K} > D_2[(n_2 - 1) - w(n_1 - 1)\phi]/(K + w) \qquad (K\phi > 1),$$

where D_1 and D_2 are certain positive constants independent of K, M_1 and M_2 and where $\phi = \dfrac{n_2 - 1}{w(n_1 - 1)}$. We notice that if (i) $n_1 = n_2$ and $w = 1$ or (ii) $w = \dfrac{n_2 - 1}{n_1 - 1}$, we have $t^* = v^*$ and both from (3.2.1.1), (3.2.1.2) and from (3.2.1.3), (3.2.1.4) we obtain $\dfrac{\partial P}{\partial K} \lesseqgtr 0$ for $K \lesseqgtr 1$. In the case (i) the maximum probability of Type I error occurs at $K = \infty$ and $K = 0$. In case (ii) the maximum will sometimes occur for $K = 0$ and sometimes for $K = \infty$, depending on the relative magnitude of n_1 and n_2.

For other situations t^* and v^* exhibit a type of behavior essentially similar to that of t_1 and v (Section 2.5). We notice that the (P, K) curve for v^* has a minimum when $K = \dfrac{w(n_1 - 1)}{n_2 - 1}$. If $n_1 = n_2$, the minimum point is given by $K = w$. Therefore with an approximate knowledge of K, a useful practical hint to remember is to so adjust M_1 and M_2 as to have w approximately equal to K. If $n_1 \neq n_2$ any information about σ_1^2 being greater or less than σ_1^2 can be used with decided advantage to adjust M_1, M_2, n_1 and n_2 so as to reduce considerably the risk of the first kind and thus work in a region of the (P, K) curve where there is not much danger of bias in the probability of Type I error. This will also reduce the fluctuations of the power function of v about its minimum which also occurs for $K = \dfrac{w(n_1 - 1)}{n_2 - 1}$.

3.2.2. *Variations in relation to w for fixed K.* The partial derivative of $P(t^* \geq t_0)$ with respect to w is given by

$$\text{(3.2.2.1)} \qquad \frac{\partial P}{\partial w} = \tfrac{1}{2}(1 - K)K^{n_2 - 1/2}(K + w)^{-1} \sum_{h=0}^{\infty} (1 - K)^h \cdot \frac{\Gamma\left(\frac{n_2 - 1}{2} + h\right) z_0^{(n_1+n_2-2)/2+h}(1 - z_0)^{\frac{1}{2}}}{h!\,\Gamma\left(\frac{n_2 - 1}{2}\right) B\left(\frac{n_1 + n_2 - 2}{2} + h, \frac{1}{2}\right)} \qquad (K < 1).$$

Therefore

$$\frac{\partial P}{\partial w} > 0$$

for $K < 1$.
Similarly

$$\frac{\partial P}{\partial w} \leq 0$$

for $K \geq 1$.
To justify the differentiation of the series in (3.2.2.1) we make use of the result

$$I_{z_0}\left(\frac{n_1 + n_2 - 2}{2} + h, \tfrac{1}{2}\right) - I_{z_0}\left(\frac{n_1 + n_2 - 2}{2} + h + 1, \tfrac{1}{2}\right)$$
$$= \frac{z_0^{(n_1+n_2-2)/2+h}(1 - z_0)^{\frac{1}{2}}}{\left(\frac{n_1 + n_2 - 2}{2} + h\right) \mathrm{B}\left(\frac{n_1 + n_2 - 2}{2} + h, \tfrac{1}{2}\right)},$$

and consequently the series under consideration may be shown to be dominated by an absolutely and uniformly convergent series for $0 < K < 1$.

For the statistic v^* consider

$$\begin{aligned} P(v^* \geq t_0) &= \tfrac{1}{2}(K\phi)^{(n_2-1)/2} \sum_{h=0}^{\infty} (1 - K\phi)^h \, \Gamma\left(\frac{n_2 - 1}{2} + h\right) \\ &\cdot \left[\Gamma(h + 1)\Gamma\left(\frac{n_2 - 1}{2}\right)\right]^{-1} I_{y_0}\left(\frac{n_1 + n_2 - 2}{2} + h, \tfrac{1}{2}\right) \quad (K\phi < 1) \end{aligned} \tag{3.2.2.2}$$

where $y_0 = (1 + \lambda_1 t_0^2)^{-1}$. We notice from (3.2.2.2) and from the form of quantities λ_1 and λ_2 (Section 3.1) that $P(v^* \geq t_0)$ depends on K and w only through the product of K and $1/w$. Consequently variations of P w.r.t. $1/w$ for fixed K are the same as those of P w.r.t. K for fixed w. Thus we may directly infer that $P(v^* \geq t_0)$ will be insensitive to the variations of w. The following Table VII will illustrate the nature of variations in the probability of Type I error in the tests based on t^* and v^* in relation to w.

TABLE VII

Variations in the probability of Type I error of t^* and v^*

$(K = 2;\ n_1 = n_2 = 7;\ t_0 = 1.782)$

w	0	.25	.5	1	2	∞
$P(t^* \geq t_0)$	.0259	.0358	.0427	.0512	.0594	.0866
$P(v^* \geq t_0)$	.0625	.0570	.0539	.0512	.05	.0625

It would appear that on the analogy of statistics t_1 and v for the comparison of two means one could guess about the sensitive nature of t^* in relation to the variations of the 'nuisance' parameter K. The additional drawback in t^* which stems from the monotone nature of its variations with respect to w is a further warning against the use of a t^* type statistic for the hypothesis $\beta_1 = \beta_2$ when $\sigma_1^2 \neq \sigma_2^2$.

4. Hypothesis of equality of two linear regression functions when variances are unequal.

4.1. *The statistic* Z. (For notation refer to Sections 2.1 and 3.1). Consider the model given in Sections 1 and 3 for the comparison of two regression coefficients. If the variances are equal, the statistic based on the likelihood ratio criterion for the composite hypothesis $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$ is given by

$$Z = \frac{(\bar{y}_1 - \bar{y}_2)^2(n_1 + 1)(n_2 + 1)(n_1 + n_2 + 2)^{-1} + (b_1 - b_2)^2 M_1 M_2 (M_1 + M_2)^{-1}}{S(y - Y)^2 + S'(y' - Y')^2}.$$

The quantity Z is distributed like the ratio of two independently distributed χ^2's and consequently its distribution is precisely determined under the hypothesis. If $\sigma_1^2 \neq \sigma_2^2$, Z can be put in the form of

$$Z = (a_1\chi_{1,1}^2 + a_2\chi_{2,1}^2)\,(K\chi_{3,n_1-1}^2 + \chi_{4,n_2-1}^2)^{-1},$$

which is now distributed as the ratio of 'mixtures' of independently distributed χ^2's with d.o.f. indicated in the second subscripts and where

$$a_1 = [n_1 + 1 + K(n_2 + 1)]\,(n_1 + n_2 + 2)^{-1},$$
$$a_2 = (K + w)\,(1 + w)^{-1}.$$

In the non-null case when $\alpha_1 \neq \alpha_2$, $\beta_1 \neq \beta_2$ the numerator of Z is a mixture of non-central squares. If we let $\beta(K, w, \delta, \Delta, n_1, n_2)$ denote the power function of Z, following Robbins and Pittman [12] we obtain

$$\beta(K, w, \delta, \Delta, n_1, n_2) = \sum_{j=0}^{\infty}\sum_{h=0}^{\infty}\sum_{k=0}^{\infty} c_j\, d_h p_k I_\zeta\left(\frac{n_1 + n_2}{2} + h - 1, k + j + 1\right) \qquad \left(K > 1,\ w < \frac{n_1 + 1}{n_2 + 1}\right), \tag{4.1.1}$$

where

$$c_j = \frac{(a_1/a_2)^{\frac{1}{2}}\Gamma(j + \frac{1}{2})}{\Gamma(\frac{1}{2})j!}(1 - a_1/a_2)^j,$$

$$d_h = \frac{K^{-(n_1-1)/2}\Gamma\left(\frac{n_1 - 1}{2} + h\right)\left(1 - \frac{1}{K}\right)^h}{\Gamma\left(\frac{n_1 - 1}{2}\right)h!},$$

$$p_k = e^{-\frac{1}{2}D^2}(\tfrac{1}{2}D^2)^k/k! \qquad (D^2 = \delta^2 + \Delta^2),$$
$$\zeta = (1 + Z_0/a_1)^{-1}.$$

4.2. *Variations in the probability of Type I error and the power function of Z.* Corresponding to (4.1.1) we obtain the expression for the probability of Type I error $P(Z \geq Z_0)$ by putting $D = 0$ and $k = 0$. It has not been possible to establish any definite law concerning the behavior of the probability of Type I error and the power function w.r.t. the 'nuisance' parameter K. However we shall presently establish their monotone dependence on the variable parameter w.

We differentiate $P(Z \geq Z_0)$ with respect to w and after simplification obtain

$$\frac{\partial P}{\partial w} = (K - 1)(a_1/a_2)^{\frac{1}{2}}\Sigma\Sigma \frac{d_h\Gamma(j + \frac{1}{2})}{j!\Gamma(\frac{1}{2})}\left[\frac{1}{2}\left(1 - \frac{a_1}{a_2}\right)^j - \frac{a_1}{a_2}j\left(1 - \frac{a_1}{a_2}\right)^{j-1}\right]$$

$$\cdot I_\xi\left(\frac{n_1 + n_2}{2} + h - 1, j + 1\right) = \frac{(K - 1)(a_1/a_2)^{\frac{1}{2}}}{(K + w)(1 + w)} \Sigma\Sigma\, d_h \frac{\Gamma(j + 3/2)}{j!\Gamma(\frac{1}{2})}$$

$$\cdot\left[I_\xi\left(\frac{n_1 + n_2}{2} + h - 1, j + 1\right) - I_\xi\left(\frac{n_1 + n_2}{2} + h - 1, j + 2\right)\right] < 0$$

for $K > 1$, $w < \dfrac{n_1 + 1}{n_2 + 1}$. Similarly by utilizing an appropriate expression for $P(Z \geq Z_0)$ for $K > 1$, $w > \dfrac{n_1 + 1}{n_2 + 1}$ we can show that $\dfrac{\partial P}{\partial w} < 0$. For the case $K < 1$ it can be shown that $P(Z \geq Z_0)$ is a monotone increasing function of w. This is also true of the dependence of the power function of Z on w.

4.3. *Unbiasedness of Z.* We differentiate (4.1.1) w.r.t. δ and Δ and after simplification obtain $\dfrac{\partial\beta}{\partial\delta} \geq 0, \dfrac{\partial\beta}{\partial\Delta} \geq 0$. Thus the power function of Z has a relative minimum at $\delta = 0$, $\Delta = 0$.

The author is greatly indebted to Professors Harold Hotelling and William G. Madow for guidance in this research and to the referees for many useful suggestions and criticisms.

REFERENCES

[1] B. L. WELCH, "The significance of the difference between two means when the population variances are unequal", *Biometrika*, Vol. 29 (1938), pp. 350–361.

[2] M. G. KENDALL, *The Advanced Theory of Statistics*, Vol. 2, C. Griffin and Co., 1946, pp. 112–115.

[3] P. L. HSU, "Contribution to the theory of "Student's" t-test as applied to the problem of two samples", *Stat. Res. Memoirs*, Vol. 2 (1938), pp. 1–24.

[4] H. SCHEFFÉ, "On the solutions of the Behrens-Fisher problem based on the t-distribution", *Annals of Math. Stat.*, Vol. 14 (1943), pp. 35–44.

[5] W. V. BEHRENS, "Ein Beitrag zur fehlerberechnung bei wenigen Beobachtungen", *Landw. Jb.*, Vol. 68 (1929), pp. 807–837.

[6] R. A. FISHER, "The fiducial argument in statistical inference", *Annals of Eugenics*, Vol. 6 (1935), pp. 391–398.

[7] J. NEYMAN, "Statistical problems in agricultural experiments", *Jour. Roy. Stat. Soc., Supp.*, Vol. 2 (1935), pp. 107–180.

[8] J. E. WALSH, "On the power efficiency of a t-test formed by pairing sample values" *Annals of Math. Stat.*, Vol. 38 (1947), pp. 601-604.
[9] J. NEYMAN, "Fiducial argument and the theory of confidence intervals", *Biometrika*, Vol. 32 (1941), pp. 128, ff.
[10] P. V. SUKHATME, "On Fisher and Behrens' test of significance for the difference in means of two normal samples", *Sankhyā*, Vol. 4 (1938), pp. 39-48.
[11] R. A. FISHER AND F. YATES, *Statistical Tables*, Oliver and Boyd, 1943.
[12] H. ROBBINS AND E. J. G. PITMAN, "Application of the method of mixtures to quadratic forms in normal variates", *Annals of Math. Stat.*, Vol. 20 (1949), pp. 552-560.