
DEPARTMENT OF STATISTICS

University of Wisconsin
Madison, Wisconsin 53706

TECHNICAL REPORT NO. 338

June 1973

ON TESTING EQUALITY OF MEANS OF TWO NORMAL POPULATIONS WITH UNEQUAL VARIANCES

by

Austin F. S. Lee and John Gurland
Boston University — University of Wisconsin

Typist: Bernice R. Weitzel

ON TESTING EQUALITY OF MEANS OF TWO NORMAL POPULATIONS WITH UNEQUAL VARIANCES

by

Austin F. S. Lee and John Gurland
Boston University University of Wisconsin

SUMMARY

A feasible method is developed for obtaining the size and power of a wide class of tests which includes solutions to the Behrens-Fisher problem proposed by various authors. A new test in this class is also developed which controls the size effectively and has high power. A comparison is made with other known tests as regards control of size. The power of the proposed test as well as that of Welch-Aspin is also obtained.

1. Introduction

Suppose $x_{11}, x_{12}, \ldots, x_{1n_1}$ and $x_{21}, x_{22}, \ldots, x_{2n_2}$ are the observed values of two random samples independently drawn from the normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Let $R = \sigma_1^2/\sigma_2^2$, $C = (\sigma_1^2/n_1)/(\sigma_1^2/n_1 + \sigma_2^2/n_2) = n_2R/(n_1 + n_2R)$, and for $i = 1, 2$, $\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$, $s_i^2 = \frac{1}{f_i}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2$, $f_i = n_i-1$. It is desired to test the null hypothesis H_0: $\mu_1 = \mu_2$ against alternatives $\mu_1 > \mu_2$. Various tests for this purpose have been developed by Fisher (1935, 1941), Welch (1947), Aspin (1948), Cochran and Cox (1950), Wald (1955). The critical regions corresponding to these tests have the following general form:

$$v > V(\hat{C}) \qquad (1)$$

where $v = (\bar{x}_1-\bar{x}_2)/\sqrt{s_1^2/n_1 + s_2^2/n_2}$, $\hat{C} = (s_1^2/n_1)/(s_1^2/n_1 + s_2^2/n_2)$,

and $V(\hat{c}) = h(C; n_1, n_2, \alpha)$ is a function of $\hat{C}$, n_1, n_2 and the preassigned nominal significance level α. Recently, Pagurova (1968) has also developed a test of form (1) but with $\hat{C}$ replaced by

$$\tilde{C} = \hat{C} - 2\hat{C}(1-\hat{C})(\frac{1-\hat{C}}{f_2} - \frac{\hat{C}}{f_1}).$$

For all of these tests, the computation of the exact size and power is quite involved due to the difficulty of evaluating certain integrals. Simulation studies have been conducted by Bennett and Hsu (1961), and Murphy (1967) to compare the size and power of some of the tests mentioned above. At best these provide only an indication of comparative behaviour, and leave much to be desired.

The purpose of the present paper is three fold. First, a test of form (1) is presented in which the function V has a comparatively simple form, the size of the test is extremely close to the nominal significance level, and at the same time the power is high. Second, the size of this proposed test is compared with that of others appearing in the literature. Third, the power of this test is obtained and compared with that of the Welch-Aspin test. For computing the size and power, a technique is presented here which yields a high degree of accuracy and is applicable to any test of form (1).

As pointed out by Scheffé (1970, footnote 4, p. 1504), there is a great need to verify the stability of the size of the test developed by Welch-Aspin. Due to the extreme mathematical complexity required for such verification, there does not appear to be any work in the literature except by Welch (1949) for the case $n_1 = n_2 = 7$ and $\alpha = .05$. More recently, Wang (1971) has provided an approximate verification for a few other cases.

As for the power, to our knowledge, no exact values have appeared anywhere in the literature. We believe the present article fills this gap and also verifies that the Welch-Aspin test behaves remarkably well.

2. Size and power for tests of form (1)

2.1 Notation

If the value of the parameter $C = n_2R/(n_1 + n_2R)$ were known, one could use a critical region

$$v > h(C; n_1, n_2, \alpha) \tag{2}$$

for testing H_o: $\mu_1 = \mu_2$ at level α. Define the non-centrality parameter $\delta = (\mu_1 - \mu_2)/\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ and write $V(C) = h(C; n_1, n_2, \alpha)$. We can now write the power of the test given by (2) as

$$Q_{V(C)}(\delta,C) = P[v > V(C)|\delta,C].$$

In particular, $Q_{V(C)}(0,C)$ is the size of the test.

In reality, however, the value of C is unknown, and we naturally resort to a critical region $v > V(\hat{C})$ of form (1). For such a region we denote the power by

$$Q_{V(\hat{C})}(\delta,C) = P[v > V(\hat{C})|\delta,C]$$

corresponding to the specific values of δ and C. In particular, the size of the test in (1) can be expressed as $Q_{V(\hat{C})}(0,C)$ corresponding to a specific value of C.

2.2 An expression for the power of the test $v > V(\hat{C})$

The power of this test can be written as

$$Q_{V(\hat{C})}(\delta, C) = P[v > V(\hat{C}) | \delta, C]$$

$$= P\left[Z > V(\hat{C}) \sqrt{\frac{s_1^2/n_1 + s_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \;\middle|\; \delta, C \right]$$

where $Z = (\bar{x}_1 - \bar{x}_2)/\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ is a normal random variable with mean δ and variance 1, and independent of s_1^2, s_2^2. Thus

$$Q_{V(\hat{C})}(\delta, C) = 1 - E\Phi\left(V(\hat{C}) \sqrt{\frac{s_1^2/n_1 + s_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}} - \delta \right)$$

where Φ is the distribution function of a standarized normal random variable. By utilizing the fact that $f_i s_i^2/\sigma_i^2$, $i = 1, 2$, are χ^2 random variables with f_i degrees of freedom, it follows that

$$Q_{V(\hat{C})}(\delta, C) = 1 - \int_0^\infty \int_0^\infty g(x,y) \int_{-\infty}^{V^*(x,y)-\delta} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, dt dx dy, \quad (3)$$

where

$$g(x,y) = \frac{1}{\Gamma(\frac{f_1}{2})\Gamma(\frac{f_2}{2})} x^{\frac{f_1}{2}-1} y^{\frac{f_2}{2}-1} e^{-x-y}, \quad x > 0,\ y > 0,$$

$$V^*(x,y) = V\{\hat{C}(x,y)\} \sqrt{\frac{2C}{f_1} x + \frac{2(1-C)}{f_2} y},$$

and

$$\hat{C}(x,y) = \frac{1}{1 + \frac{f_1}{f_2} \frac{1-C}{C} \frac{y}{x}}.$$

This result was obtained by Golhar (1964).

2.3 Evaluation of the integral in (3)

The third integral in the right-hand side of (3) involves evaluation of the standard normal distribution function

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

By expanding $e^{-t^2/2}$ in a power series and integrating term by term, we obtain the following series expansion for $\Phi(z)$, valid for all finite z:

$$\Phi(z) = \frac{1}{2} + \sum_{k=0}^{\infty} c_k z^{2k+1}, \qquad (4)$$

where

$$c_k = \frac{1}{\sqrt{2\pi}} \frac{(-1)^k}{k!} \frac{1}{(2k+1)2^k}$$

Convergence of the series in (4) can be accelerated by repeated application of a transformation described as follows.

Let $S = a_0 + a_1 + a_2 + \ldots$ be an infinite series and $S_n = a_0 + a_1 + \ldots + a_n$, its partial sum. Define

$$T_{n+1} = \frac{S_{n+1}S_{n-1} - S_n^2}{S_{n+1} + S_{n-1} - 2S_n}, \quad n = 1, 2, \ldots$$

The original series of partial sums $\{S_n\}_{n=0}^{\infty}$ is thereby transformed into a new series $\{T_n\}_{n=1}^{\infty}$. This transformation will be designated here as the T-transform. Lubkin (1952), Shanks (1955) and others have utilized this type of transformation to obtain series which converge to the same limit as the original series and do so faster. The T-transform can also be applied iteratively in attempting to further improve the rate of convergence. Thus, the $(j+1)$st iterate can be expressed as

$$T_{n+1}^{(j+1)} = \frac{T_{n+1}^{(j)} T_{n-1}^{(j)} - T_n^{(j)2}}{T_{n+1}^{(j)} + T_{n-1}^{(j)} - 2T_n^{(j)}}$$

$$n = 2j,\ 2j+1,\ 2j+2,\ \ldots\ ;$$

$$j = 0,\ 1,\ 2,\ \ldots$$

where

$$T_n^{(0)} = S_n = \frac{1}{2} + \sum_{k=0}^{n} c_k\ z^{2k+1}\ , \quad n = 0,\ 1,\ 2,\ \ldots$$

$T_n^{(j)}$ = the jth order transform,

$$n = 2j,\ 2j+1,\ \ldots\ ;$$

$$j = 1,\ 2,\ 3,\ \ldots$$

To approximate $\Phi(z)$ with a specified degree of accuracy, it is natural to take

$$\Phi(z) \doteq \frac{1}{2} + \sum_{k=0}^{M} c_k z^{2k+1}, \quad -z_0 \leq z \leq z_0 \tag{5}$$

with z_0 and M sufficiently large and with M to be reduced in the iterating process.

Now let us return to the integral in (3). By letting $u = x+y$,

$w = \frac{x}{x+y}$, and integrating out u, it can be seen that

$$Q_{V(\hat{C})}(\delta,C) \doteq \frac{1}{2} - \sum_{k=o}^{M} c_k \sum_{j=o}^{2k+1} b_{kj}(-\delta)^{2k+1-j} \, E[V^*(W)]^j \ , \quad (6)$$

where

$$b_{kj} = \binom{2k+1}{j} \frac{\Gamma(\frac{f_1+f_2+j}{2})}{\Gamma(\frac{f_1+f_2}{2})} \ , \ k = o, 1, \ldots, M; \quad j = o, 1, \ldots, 2k+1,$$

$$V^*(w) = V(\hat{C}(w))\sqrt{\frac{2C}{f_1} w + \frac{2(1-C)}{f_2} (1-w)} \ ,$$

$$\hat{C}(w) = \frac{1}{1 + \frac{f_1}{f_2} \frac{1-C}{C} \frac{1-w}{w}} \ ,$$

and W is a random variable which follows a Beta-distribution with parameters $f_1/2$ and $f_2/2$.

An alternative method of computing $Q_{V(\hat{C})}(\delta,C)$ is obtainable from (6) by interchanging the order of summation and of integration involved in the expectation. This yields

$$Q_{V(\hat{C})}(\delta,C) \doteq \frac{1}{2} - \frac{1}{B(\frac{f_1}{2},\frac{f_2}{2})} \int_o^1 w^{\frac{f_1}{2}-1} (1-w)^{\frac{f_2}{2}-1} \sum_{k=o}^{M(w)} c_k \sum_{j=o}^{2k+1} b_{kj}(-\delta)^{2k+1-j} [V^*(w)]^j dw \ , \quad (7)$$

Thus, only one numerical evaluation of integration is needed, making this method more attractive. The integer $M(w)$ can be determined according to the current value of w in the process of numerical integration. For example, to compute $Q_{V(\hat{C})}(\delta,C)$ with error less than 10^{-5}, and for $(n_1, n_2, \alpha) = (9, 9, .05)$, repeated use of the T-transform reduced values of $M(w)$ from 11 to 6. For $(n_1, n_2, \alpha) = (5, 9, .05)$, values of $M(w)$ were reduced from 30 to 7.

As a check on the validity of (7), the size of the critical region for various combinations of sample sizes n_1, n_2 satisfying $3 \leq n_1, n_2 \leq 7$ and $\alpha = .05$ were computed and then compared with corresponding values obtained by Mehta and Srinivasan (1970) who used a different method involving a rational approximation of $\Phi(z)$ appearing in Abramowitz and Stegun (1965), and numerical integration of the double integral in (3) based on Simpson's rule.

The values listed by Mehta and Srinivasan, all given to four decimal places, agree exactly with the corresponding values obtained from (7). Moreover, to check the accuracy of our values given to five decimal places, we evaluated the double integral in (3) for various values of $V(\hat{C})$ and δ by the method of Gaussian quadrature and found that the differences in the values thus obtained did not exceed 10^{-5}.

As an additional check on the accuracy of (7), we considered $V(\hat{C})$ = constant and n_1, n_2 both odd, because for this case the size can be expressed in closed form as a finite series (McCullough, Gurland and Rosenberg (1960), Gurland (1962)). In all instances, the values of size obtained by both methods coincided up to the 6th decimal place.

3. A proposed test

Our purpose here is to find a function $V(\hat{C}) = h(\hat{C}; n_1, n_2, \alpha)$ so that

$$P[v > V(\hat{C})|0,C] = \alpha, \quad 0 \leq C \leq 1 . \tag{8}$$

It is generally believed that there exists no smooth and simple function V such that (8) holds exactly (cf. Wilks (1940), Bartlett (1956), Wallace (1958), Linnik (1964, 1966)). Nevertheless, it is of some practical as well as theoretical interest to seek an approximate solution $V_0(\hat{C})$, say, for which (8) holds with only negligible error. In attempting to develop a solution which is practicable, the search here has been concentrated upon functions which are comparatively simple in form. The power of the resulting tests has also been considered.

3.1 Criteria in seeking a critical point $V(\hat{C})$

Since the different forms of $V(\hat{C})$ considered in our study depend on sets of constants, we find it convenient for our purpose to indicate their presence through the use of the symbol $V(\hat{C};\, \underline{a})$, whatever be the form of $V(\hat{C})$. Thus, equation (8) will now be written as

$$Q_{V(\hat{C})}(0,C) = P[v > V(\hat{C};\, \underline{a})\,|\,0,C] = \alpha$$

where $V(\hat{C}) = V(\hat{C};\, \underline{a})$ and $\underline{a}$ is the vector of constants involved in the function V.

We shall seek a function V of $\hat{C}$ satisfying the following criteria.

(1) The function V should have a comparatively simple form so that, in practice, the necessary calculation of critical points may be carried out easily.

(2) The size of the test with critical region $v > V(\hat{C})$ should be well controlled in the whole interval $0 \leq C \leq 1$.

(3) The power of the test should be reasonably high.

To perform the above non-linear minimization problem, a subroutine available on UNIVAC 1108 at the University of Wisconsin Computing Center was used. This subroutine applies the Marquardt algorithm, a summary of which appears in Box and Jenkins (1970, p. 504). At each step of the minimizing process, the size $Q_{V(\hat{C};\ \underline{a})}(0,C)$ corresponding to the current values of $\underline{a}$ was calculated by using (7) and the T-transform introduced above.

3.2 Proposed critical point $V_0(\hat{C})$

After extensive trials and comparisons involving several functional forms of $V(\hat{C};\ \underline{a})$ it was found that the ratio of two polynomials

$$V(C;\ \underline{a}) = \frac{a_1 + a_2\hat{C} + a_3\hat{C}^2}{1 + a_4\hat{C} + a_5\hat{C}^2} \qquad (9)$$

provides the most satisfactory result. In attempting to satisfy criterion (2), the values of the a_i's were chosen so that the error sum of squares

$$\sum_{j=1}^{M} [Q_{V(\hat{C};\ a)}(0,C_j) - \alpha]^2 \qquad (10)$$

was minimized, subject to the conditions

$$Q_{V(\hat{C};\ \underline{a})}(0,0) = \alpha; \qquad Q_{V(\hat{C};\ \underline{a})}(0,1) = \alpha.$$

The imposition of these conditions is equivalent to setting

$$V(0;\ \underline{a}) = t_{f_2;\alpha} \qquad V(1;\ \underline{a}) = t_{f_1;\alpha}, \qquad (11)$$

where $t_{f;\alpha}$ denotes the upper 100α% probability point of the t-distribution with f degrees of freedom. The quantities C_j, $j = 1, \ldots, m$ in (10) are m prespecified points of C in the interval $0 < C < 1$. We designate by

$V_0(C)$ the solution (9) with the values of a determined by the optimization process given by (10) and (11).

By way of illustration, the values of the a_i's in $V_0(\hat{C})$ for $n_1 = n_2 = 9$ and $\alpha = .05, .025, .01, .005$ are shown in Table 1. The size at log R = 0, 1, 2, 3, 4, is also shown. For the values of α and log R appearing in the table, the actual size lies between $\alpha \pm .0001$.

As an example for the case of unequal sample sizes, $V_0(\hat{C})$ for $(n_1, n_2, \alpha) = (5, 9, .05)$ is considered in Table 2. In addition to the size of the corresponding test, the power at $\delta = 1, 2, 3, 4$ is listed for log R = $-\infty$, -4 (1) 4, and ∞. For finite values of log R, the power was calculated using (7). For log R = $-\infty (C = 0)$, $V_0(\hat{C}) = t_{f_2;\alpha}$ with probability 1 and the distribution of v is that of $t_{f_2}(\delta)$, a non-central t-variate with f_2 degrees of freedom and non-centrality parameter δ. Hence the power at log R = $-\infty$ is equal to $P[t_{f_2}(\delta) > t_{f_2;\alpha}]$. Similarly, the power at log R = $\infty (C = 1)$ is equal to $P[t_{f_1}(\delta) > t_{f_1;\alpha}]$.

The size listed in Table 2 lies between .05 ± .0003. In appraising the power of the test using $V_0(\hat{C})$, a comparison is given in Table 2 with that of the optimal test when R is known, namely,

$$u_R > t_{f_1+f_2;\alpha}$$

where

$$u_R = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{f_1 s_1^2 + f_2 s_2^2 R}{f_1 + f_2}\left(\frac{1}{n_1} + \frac{1}{n_2 R}\right)}} .$$

The test based on u_R, when R is known, is optimal in the sense that it is uniformly most powerful unbiased for testing $\mu_1 = \mu_2$ against alternatives

Table 1

Constants in $V_o(\hat{C}; \underline{a})$
and size of Corresponding Test
for $n_1 = n_2 = 9$

α	a_1	a_3	a_5	Size				
				log R=0	1	2	3	4
.05	1.8595	3.139172	1.509582	.0500	.0501	.0500	.0500	.0500
.025	2.3060	4.676735	1.819262	.0249	.0251	.0250	.0249	.0249
.01	2.8965	6.969084	2.168606	.0099	.0101	.0100	.0099	.0099
.005	3.3554	8.989510	2.433362	.0049	.0051	.0050	.0049	.0049

Note: When $n_1 = n_2$, we have $a_2 = -a_3$, and the size is symmetric about $\log R = 0$.

Table 2

Size and Power $Q_{V_o(\hat{C})}(\delta, C)$

for $(n_1, n_2, \alpha) = (5, 9, .05)$*

δ \ log R =	$-\infty$	-4	-3	-2	-1	0	1	2	3	4	∞	
R =	0	.02	.05	.14	.37	1	2.7	7.4	20.1	54.6	∞	u_R
C =	0	.03	.08	.20	.40	.64	.83	.93	.97	.99	1	
0 (Size)	.0500	.0500	.0501	.0501	.0498	.0500	.0503	.0499	.0497	.0498	.0500	.0500
1	.2332	.2346	.2368	.2403	.2402	.2326	.2213	.2135	.2107	.2100	.2098	.2417
	96%	97%	98%	99%	99%	96%	92%	88%	87%	87%	87%	
2	.5725	.5759	.5813	.5907	.5919	.5676	.5343	.5150	.5078	.5055	.5043	.5951
	96%	97%	98%	99%	99%	95%	90%	87%	85%	85%	85%	
3	.8618	.8647	.8690	.8770	.8784	.8541	.8211	.8025	.7952	.7926	.7911	.8810
	98%	98%	99%	100%	100%	97%	93%	91%	90%	90%	90%	
4	.9767	.9776	.9790	.9815	.9820	.9726	.9587	.9502	.9466	.9452	.9444	.9828
	99%	99%	100%	100%	100%	99%	98%	97%	96%	96%	96%	

$$* \; V_o(\hat{C}) = \frac{1.8595 - 4.055449\,\hat{C} + 3.641717\,\hat{C}^2}{1 - 1.992212\,\hat{C} + 1.670403\,\hat{C}^2}$$

$\mu_1 > \mu_2$. Its power is listed in the last column of Table 2. For convenience in making comparisons, the following ratio is included among the entries of the table:

$$\frac{\text{Power of } V_0(\hat{C}) \text{ test}}{\text{Power of } u_R \text{ test with } R \text{ known}} .$$

From these ratios, which are indicated as percentage values in the table, we can see that the power of the $V_0(\hat{C})$ test is for the most part more than 90% of that of the u_R test, even though for the former no information about R is assumed. For further comparison, the power of the $V_0(\hat{C})$ test for $C = f_1/(f_1 + f_2) = 1/3$ and $C = 1$ is plotted in Figure 1 along with that of the u_R test. For $C = 1/3$, the $V_0(\hat{C})$ test is virtually as powerful as the u_R test. For other values of C, the power of the $V_0(\hat{C})$ test does not fall below that for $C = 1$, as is evident from Table 2.

The above results are illustrations for particular sample size combinations, but a thorough study has also been made of other cases and the conclusions agree with those above. The solution $V(\hat{C}) = V_0(\hat{C})$ obtained through the optimization procedure described above fulfills the three criteria stated in Section 3.1.

In Table 3, the constants a_i, $i = 1, 2, 3, 4, 5$, of $V_0(\hat{C})$ are listed for sample sizes $5 \le n_1 \le n_2 \le 21$ and $\alpha = .05$. A more complete table of the constants a_i for sample sizes $5 \le n_1 \le n_2 \le 31$ and $\alpha = .05$ appears in Lee (1972).

The extent of deviation of the actual size from the nominal significance level $\alpha = .05$ that results in using the critical region $v > V_0(\hat{C})$ may be summarized as shown in Figure 2. For example, for $n_1 = 5$ and $5 \le n_2 \le 9$, the actual size lies between $.05 \pm .0003$. For $n_1 = 6$, $n_2 = 8, 9$, it lies between $.05 + .0002$ and $.05 - .0001$. For $10 \le n_1 \le n_2 \le 31$, the size is .0500. In fact, if the values are rounded off to the fifth decimal place, then for $n_1, n_2 \ge 20$, the actual size = .05000.

Fig. 1

Comparison of Power of the $V_o(\hat{C})$ Test and the u_R-Test

$n_1=5$, $n_2=9$, $\alpha=.05$

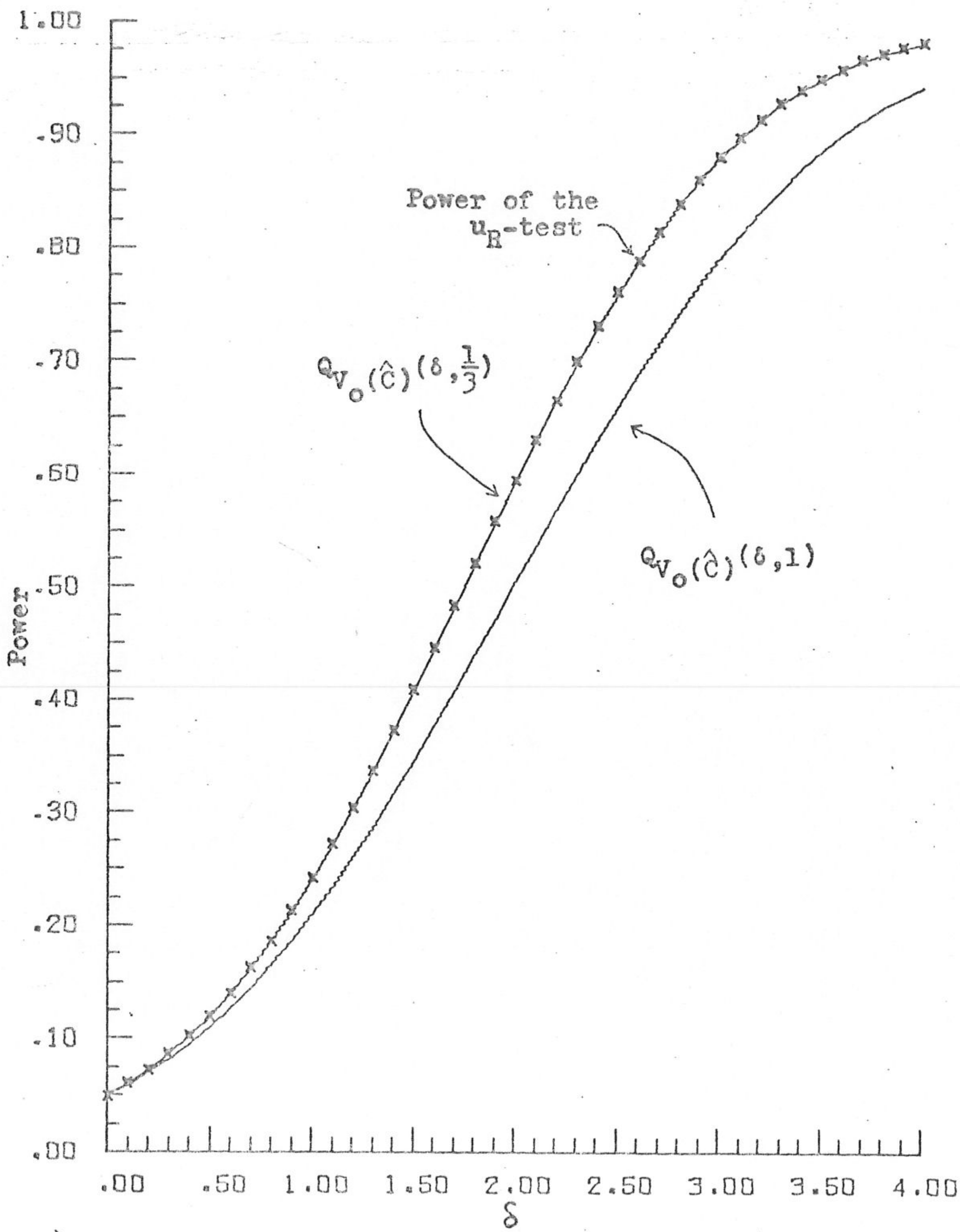


Table 3* Constants in $V_o(\hat{C}\,;\,\underline{a})$ for $5 \le n_1 \le n_2 \le 21$ and $\alpha = .05$

$n_2 \backslash n_1$	5	6	7	8	9	10	11	12	13
	2.1318								
	-6.4441								
5	6.4441								
	-2.7949								
	2.7949								
	2.0150	2.0150							
	-5.4628	-5.1822							
6	5.2231	5.1822							
	-2.4821	-2.3487							
	2.3148	2.3487							
	1.9432	1.9432	1.9432						
	-4.8321	-4.5038	-4.3100						
7	4.4825	4.3523	4.3100						
	-2.2694	-2.1036	-2.0099						
	2.0169	1.9928	2.0099						
	1.8946	1.8946	1.8946	1.8946					
	-4.4600	-4.0611	-3.8216	-3.6155					
8	4.0271	3.8258	3.7118	3.6155					
	-2.1563	-1.9440	-1.8213	-1.7141					
	1.8420	1.7675	1.7397	1.7141					
	1.8595	1.8595	1.8595	1.8595	1.8595				
	-4.0554	-3.7666	-3.5339	-3.3059	-3.1392				
9	3.6417	3.4733	3.3492	3.2242	3.1392				
	-1.9922	-1.8424	-1.7203	-1.5983	-1.5096				
	1.6704	1.6197	1.5822	1.5367	1.5096				
	1.8331	1.8331	1.8331	1.8331	1.8331	1.8331			
	-3.7828	-3.5882	-3.3014	-3.0936	-2.8897	-2.7566			
10	3.3837	3.2191	3.0746	2.9560	2.8290	2.7566			
	-1.8859	-1.7913	-1.6342	-1.5219	-1.4105	-1.3386			
	1.5585	1.5179	1.4608	1.4169	1.3637	1.3386			
	1.8125	1.8125	1.8125	1.8125	1.8125	1.8125	1.8125		
	-3.6257	-3.3347	-3.1124	-2.9100	-2.7334	-2.5612	-2.4486		
11	3.2009	3.0044	2.8599	2.7349	2.6286	2.5163	2.4486		
	-1.8371	-1.6814	-1.5620	-1.4516	-1.3546	-1.2593	-1.1977		
	1.4881	1.4170	1.3648	1.3158	1.2729	1.2235	1.1977		
	1.7959	1.7959	1.7959	1.7959	1.7959	1.7959	1.7959	1.7959	
	-3.4335	-3.1682	-2.9491	-2.7677	-2.6002	-2.4376	-2.3020	-2.2131	
12	3.0376	2.8300	2.6826	2.5645	2.4625	2.3553	2.2621	2.2131	
	-1.7572	-1.6160	-1.4971	-1.3980	-1.3052	-1.2145	-1.1391	-1.0900	
	1.4139	1.3394	1.2842	1.2386	1.1969	1.1493	1.1079	1.0900	
	1.7823	1.7823	1.7823	1.7823	1.7823	1.7823	1.7823	1.7823	1.7823
	-3.3642	-3.0420	-2.8437	-2.7268	-2.4889	-2.3356	-2.1969	-2.0923	-2.0101
13	2.9409	2.6993	2.5567	2.4796	2.3265	2.2252	2.1281	2.0603	2.0101
	-1.7449	-1.5679	-1.4604	-1.3983	-1.2632	-1.1773	-1.0994	-1.0409	-0.9950
	1.3823	1.2824	1.2299	1.2086	1.1344	1.0893	1.0448	1.0155	0.9950

* Five numbers in each (n_1, n_2)-cell are consecutively a_1, a_2, a_3, a_4 and a_5.

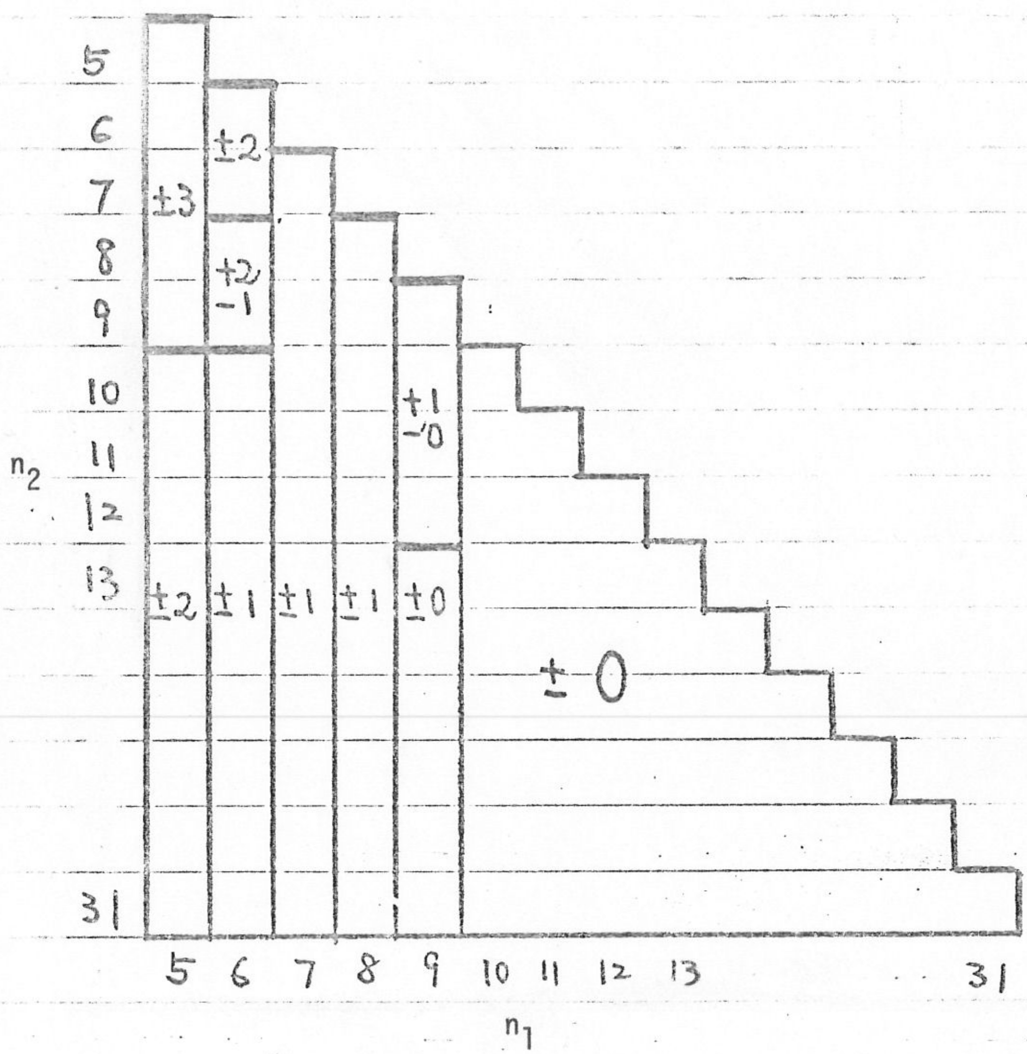
Table 3 (continued)

$n_2 \backslash n_1$	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	1.7709	1.7709	1.7709	1.7709	1.7709	1.7709	1.7709	1.7709	1.7709	1.7709							
	-3.2394	-2.9324	-2.7256	-2.5614	-2.4042	-2.2491	-2.1246	-2.0182	-1.9224	-1.8565							
14	2.8230	2.5900	2.4343	2.3177	2.2183	2.1137	2.0281	1.9573	1.8924	1.8565							
	-1.6944	-1.5252	-1.4120	-1.3215	-1.2331	-1.1455	-1.0754	-1.0155	-0.9613	-0.9243							
	1.3298	1.2341	1.1735	1.1276	1.0855	1.0377	0.9992	0.9676	0.9381	0.9243							
	1.7613	1.7613	1.7613	1.7613	1.7613	1.7613	1.7613	1.7613	1.7613	1.7613	1.7613						
	-3.0972	-2.8367	-2.6437	-2.4473	-2.3103	-2.1679	-2.0560	-1.9477	-1.8425	-1.7735	-1.7078						
15	2.7312	2.4995	2.3458	2.2084	2.1141	2.0184	1.9409	1.8665	1.7935	1.7522	1.7078						
	-1.6281	-1.4868	-1.3813	-1.2709	-1.1942	-1.1136	-1.0506	-0.9891	-0.9291	-0.8901	-0.8529						
	1.2826	1.1935	1.1344	1.0744	1.0359	0.9929	0.9588	0.9246	0.8899	0.8727	0.8529						
	1.7531	1.7531	1.7531	1.7531	1.7531	1.7531	1.7531	1.7531	1.7531	1.7531	1.7531	1.7531					
	-3.0057	-2.7405	-2.5616	-2.3792	-2.2182	-2.1032	-1.9690	-1.8817	-1.7790	-1.7034	-1.6268	-1.5605					
16	2.6589	2.4179	2.2650	2.1325	2.0215	1.9409	1.8453	1.7882	1.7158	1.6675	1.6117	1.5605					
	-1.5898	-1.4450	-1.3476	-1.2450	-1.1537	-1.0890	-1.0124	-0.9630	-0.9043	-0.8611	-0.8174	-0.7795					
	1.2495	1.1549	1.0972	1.0401	0.9907	0.9568	0.9114	0.8871	0.8524	0.8308	0.8042	0.7795					
	1.7459	1.7459	1.7459	1.7459	1.7459	1.7459	1.7459	1.7459	1.7459	1.7459	1.7459	1.7459	1.7459				
	-2.9372	-2.6833	-2.4970	-2.3232	-2.1765	-2.0479	-1.9226	-1.8699	-1.7361	-1.6649	-1.5886	-1.5075	-1.4688				
17	2.5967	2.3485	2.1957	2.0656	1.9624	1.8724	1.7827	1.8237	1.6550	1.6101	1.5546	1.4887	1.4688				
	-1.5637	-1.4258	-1.3227	-1.2247	-1.1416	-1.0684	-0.9967	-1.0261	-0.8901	-0.8494	-0.8056	-0.7590	-0.7370				
	1.2229	1.1261	1.0661	1.0103	0.9653	0.9251	0.8828	0.9168	0.8242	0.8044	0.7776	0.7441	0.7370				
	1.7396	1.7396	1.7396	1.7396	1.7396	1.7396	1.7396	1.7396	1.7396	1.7396	1.7396	1.7396	1.7396	1.7396			
	-2.8821	-2.6299	-2.4333	-2.2750	-2.1127	-1.9955	-1.8695	-1.7803	-1.6991	-1.6258	-1.5447	-1.4741	-1.4179	-1.3829			
18	2.5483	2.2972	2.1323	2.0088	1.8922	1.8100	1.7189	1.6585	1.6021	1.5546	1.4958	1.4394	1.4019	1.3829			
	-1.5435	-1.4061	-1.2965	-1.2075	-1.1147	-1.0479	-0.9755	-0.9246	-0.8781	-0.8360	-0.7892	-0.7485	-0.7162	-0.6962			
	1.2030	1.1043	1.0369	0.9851	0.9316	0.8957	0.8523	0.8254	0.7997	0.7781	0.7491	0.7210	0.7034	0.6962			
	1.7341	1.7341	1.7341	1.7341	1.7341	1.7341	1.7341	1.7341	1.7341	1.7341	1.7341	1.7341	1.7341	1.7341	1.7341		
	-2.8166	-2.5934	-2.3784	-2.2161	-2.0558	-1.9383	-1.8245	-1.7251	-1.6416	-1.5774	-1.4966	-1.4229	-1.3747	-1.3275	-1.2889		
19	2.4978	2.2561	2.0805	1.9519	1.8351	1.7516	1.6683	1.5990	1.5427	1.4985	1.4395	1.3806	1.3503	1.3189	1.2889		
	-1.5152	-1.3950	-1.2738	-1.1821	-1.0902	-1.0230	-0.9577	-0.9005	-0.8524	-0.8156	-0.7689	-0.7263	-0.6985	-0.6713	-0.6491		
	1.1791	1.0883	1.0129	0.9579	0.9041	0.8671	0.8283	0.7958	0.7699	0.7503	0.7210	0.6913	0.6778	0.6632	0.6491		
	1.7291	1.7291	1.7291	1.7291	1.7291	1.7291	1.7291	1.7291	1.7291	1.7291	1.7291	1.7291	1.7291	1.7291	1.7291	1.7291	
	-2.7654	-2.5214	-2.3396	-2.1731	-2.0143	-1.9197	-1.7951	-1.7036	-1.6200	-1.5598	-1.4832	-1.4088	-1.3614	-1.3174	-1.2675	-1.2560	
20	2.4546	2.2017	2.0356	1.9036	1.7847	1.7157	1.6250	1.5610	1.5025	1.4620	1.4080	1.3482	1.3184	1.2909	1.2501	1.2560	
	-1.4946	-1.3614	-1.2601	-1.1655	-1.0744	-1.0205	-0.9486	-0.8959	-0.8476	-0.8131	-0.7686	-0.7254	-0.6981	-0.6727	-0.6438	-0.6374	
	1.1599	1.0609	0.9935	0.9359	0.8808	0.8525	0.8087	0.7793	0.7518	0.7342	0.7076	0.6772	0.6638	0.6514	0.6309	0.6374	
	1.7247	1.7247	1.7247	1.7247	1.7247	1.7247	1.7247	1.7247	1.7247	1.7247	1.7247	1.7247	1.7247	1.7247	1.7247	1.7247	1.7247
	-2.7085	-2.4757	-2.2947	-2.1258	-1.9977	-1.8727	-1.7575	-1.6671	-1.5818	-1.5301	-1.4485	-1.3747	-1.3265	-1.2874	-1.2329	-1.2193	-1.1950
21	2.4164	2.1623	1.9934	1.8575	1.7606	1.6679	1.5826	1.5186	1.4584	1.4248	1.3656	1.3075	1.2762	1.2512	1.2070	1.2108	1.1950
	-1.4690	-1.3423	-1.2414	-1.1452	-1.0719	-0.9999	-0.9333	-0.8811	-0.8317	-0.8022	-0.7546	-0.7117	-0.6838	-0.6612	-0.6295	-0.6219	-0.6080
	1.1410	1.0427	0.9739	0.9139	0.8719	0.8290	0.7883	0.7588	0.7302	0.7166	0.6867	0.6571	0.6428	0.6318	0.6091	0.6144	0.6080

Figure 2

Error Limits of the Actual Size of the $V_o(\hat{C})$ Test*

(α = .05)



* The numbers in the entries are limits of (actual size - .05) x 10^4.

4. Comparison of various solutions to the Behrens-Fisher Problem

In the preceeding section we have obtained a critical region $v > V_o(\hat{C})$ such that V_o has a rather simple form, the size of the test is well controlled and the power is, for the most part, only slightly less than that of the optimal test based on u_R when R is known. It is of interest to compare critical points, size and power of the $V_o(\hat{C})$ test with that of other tests which have appeared in the literature on the Behrens-Fisher problem. Tests considered for this comparison here are the following

1) McCullough-Banerjee. Banerjee (1960, 1961) suggested a critical region

$$v > \left\{ t^2_{f_1;\alpha} \hat{C} + t^2_{f_2;\alpha}(1-\hat{C}) \right\}^{1/2} . \qquad (12)$$

The size of this test, as shown by Banerjee (1960), is never larger than the nominal significance level, and is more conservative than that of Cochran and Cox. Indeed, we shall see that Banerjee's test is only "slightly" more conservative than that of Cochran and Cox. McCullough, Gurland and Rosenberg (1960) derived a test which employs a 'bilateral statistic' with the critical point unity. That is, H_o is rejected when

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{t^2_{f_1;\alpha}(s_1^2/n_1) + t^2_{f_2;\alpha}(s_2^2/n_2)}} > 1. \qquad (13)$$

The rationale leading to this solution is that the size of the test is equal to the preassigned significance level at $C = 0$ and 1. It can easily be seen that this test is equivalent to Banerjee's; for from (13),

$$\bar{x}_1 - \bar{x}_2 > \sqrt{t^2_{f_1;\alpha} \frac{s_1^2}{n_1} + t^2_{f_2;\alpha} \frac{s_2^2}{n_2}} .$$

Dividing both sides by $\sqrt{s_1^2/n_1 + s_2^2/n_2}$ yields (12). The critical point due to McCullough-Banerjee will be denoted as $V_{mb}(\hat{C})$.

2) Cochran and Cox. Cochran and Cox (1950) suggested a solution to the problem by using the v-statistic with critical value as a weighted mean of $t_{f_1;\alpha}$ and $t_{f_2;\alpha}$ with weights s_1^2/n_1 and s_2^2/n_2 divided by $s_1^2/n_1 + s_1^2/n_2$. For convenience here, we designate this random critical value as $V\ (\hat{C})$

It is given by

$$V_{cc}(\hat{C}) = \frac{t_{f_1;\alpha}\dfrac{s_1^2}{n_1} + t_{f_2;\alpha}\dfrac{s_2^2}{n_2}}{s_1^2/n_1 + s_2^2/n_2} = t_{f_1;\alpha}\hat{C} + t_{f_2;\alpha}(1-\hat{C}). \qquad (14)$$

The main advantage of this solution is its simplicity. Perhaps this is the reason that makes it one of the most widely used tests in practice, even though the size of the test may differ substantially from the nominal significance level, as shown by Cochran (1964). For the most part, the size of this test errs on the conservative side.

3) Fisher. Fisher (1935) applied fiducial probability to obtain as the distribution of the v-statistic a weighted difference of two independent t-variables. Fisher (1941) showed that this test can be regarded as being based on $v > V_f(\hat{C})$ where $V_f(\hat{C})$ can be represented by an asymptotic series, the first few terms of which are as follows.

$$\begin{aligned} V_f(\hat{C}) = \xi \Bigg\{ 1 &+ \frac{1+\xi^2}{4}\left[\frac{\hat{C}^2}{f_1} + \frac{(1-\hat{C})^2}{f_2}\right] + \hat{C}(1-\hat{C})\left(\frac{1}{f_1} + \frac{1}{f_2}\right) \\ &+ \frac{3+16\xi^2+5\xi^4}{96}\left[\frac{\hat{C}^4}{f_1^2} + \frac{(1-\hat{C})^4}{f_2^2}\right] + \frac{3-\xi^2+4\xi^4}{12}\left[\frac{\hat{C}^3(1-\hat{C})}{f_1^2} + \frac{\hat{C}(1-\hat{C})^3}{f_2^2}\right] \\ &+ \frac{1}{2}(4\xi^2-1)\hat{C}^2(1-\hat{C})^2\left(\frac{1}{f_1^2} + \frac{1}{f_2^2}\right) + 2\left[\frac{\hat{C}(1-\hat{C})^3}{f_1^2} + \frac{\hat{C}^3(1-\hat{C})}{f_2^2}\right] \\ &+ \frac{5-3\xi^2}{4f_1f_2}\left[\hat{C}(1-\hat{C})^3 + \hat{C}^3(1-\hat{C})\right] - \frac{47-48\xi^2+9\xi^4}{16f_1f_2}\hat{C}^2(1-\hat{C})^2 \Bigg\} \end{aligned} \qquad (15)$$

where $\zeta = \xi_\alpha$, the upper α-point of the standard normal distribution (i.e., the probability of exceeding ξ_α is α).

4) Welch (approximate t solution). Welch (1947) approximated v by means of a Student t with a random number of degrees of freedom. This is often referred to as the "approximate degrees of freedom" or "approximate t" solution.

According to this method, the v-statistic is used with a critical value read from the Student t-table and with degrees of freedom approximated by $\hat{f}$ (Welch (1949)), where

$$1/\hat{f} = \hat{C}^2/f_1 + (1-\hat{C})^2/f_2 .$$

This is, H_0 is rejected for

$$v > t_{\hat{f};\alpha} .$$

Scheffé (1970) considers this the most practical solution to the Behrens-Fisher problem.

The following explicit expression for $t_{\hat{f};\alpha} = V_{wt}(\hat{C})$, say, can be obtained from the series expansion of a Student-t deviate (Fisher 1941; Abramowitz and Stegun 1965):

$$V_{wt}(\hat{C}) = \xi \left\{ 1 + \frac{1+\xi^2}{4\hat{f}} + \frac{3+16\xi^2+5\xi^4}{96\hat{f}^2} + \frac{-15+17\xi^2+19\xi^4+3\xi^6}{384\hat{f}^3} + \frac{-945-1920\xi^2+1482\xi^4+776\xi^6+79\xi^8}{92160\hat{f}^4} \right\} \quad (16)$$

5) Wald (in case of equal sample sizes) and Pagurova. For the case of equal sample sizes, Wald (1955) obtained a solution which uses the v-statistic with a critical point given by

$$V_{wd}(\hat{C}) = t_{f;\alpha} - 4(t_{f;\alpha} - t_{2f;\alpha})\hat{C}(1-\hat{C}) \quad (17)$$

where

$$f = f_1 = f_2 .$$

This solution is tantamount to imposing the restrictions that $V_{wd}(C) = \alpha$ at three points $C = 0, \frac{1}{2}$ and 1. Pagurova (1968) extended his result to the case of unequal sample sizes to obtain as random critical point $V_p(\tilde{C})$ for the v-statistic:

$$V_p(\tilde{C}) = t_{f_2;\alpha} \frac{(\theta-\tilde{C})^2(1-\tilde{C})}{\theta^2}$$

$$+ t_{f_1+f_2;\alpha} \frac{[\theta(1-\theta) + (2\theta-1)(\tilde{C}-\theta)]\tilde{C}(1-\tilde{C})}{\theta^2(1-\theta)^2} + t_{f_1;\alpha} \frac{(\theta-\tilde{C})^2\tilde{C}}{(1-\theta)^2} \quad (18)$$

where

$$\tilde{C} = \hat{C} - 2\hat{C}(1-\hat{C})(\frac{1-\hat{C}}{f_2} - \frac{\hat{C}}{f_1})$$

and

$$\theta = \frac{f_1}{f_1 + f_2} \quad .$$

Pagurova arrived at this solution by imposing the restrictions that $V_p(C) = \alpha$ at $C = 0,\ f_1/(f_1+f_2),\ 1$, and $V_p'(C) = 0$ at $C = f_1/(f_1+f_2)$. She also investigated the possible maximal deviation of the actual size of the test from the nominal significance level.

6) Welch and Aspin. Welch (1947) obtained an asymptotic solution $V(\hat{C}) = V_w(\hat{C})$, say, up to terms of order f_i^{-2}, $i = 1, 2$, by directly solving the equation (8). Aspin (1948) extended it to terms of order f_i^{-4}. The latter is usually called the Welch-Aspin test and we shall denote its critical point by $V_{wa}(\hat{C})$.

Let

$$V_{ru} = \frac{\hat{C}^r}{f_1^u} + \frac{(1-\hat{C})^r}{f_2^u} \quad .$$

Then

$$V_w(\hat{C}) = \xi\{1 + V_1 + V_2\} \tag{19}$$

and

$$V_{wa}(\hat{C}) = \xi\{1 + V_1 + V_2 + V_3 + V_4\} \tag{20}$$

where

$$V_1 = \frac{1+\xi^2}{4} V_{21};$$

$$V_2 = -\frac{1+\xi^2}{2} V_{22} + \frac{3+5\xi^2+\xi^4}{3} V_{32} - \frac{15+32\xi^2+9\xi^4}{32} V_{21}^2;$$

$$V_3 = (1+\xi^2)V_{23} - 2(3+5\xi^2+\xi^4)V_{33} + \frac{15+32\xi^2+9\xi^4}{8} V_{22}V_{21} + \frac{75+173\xi^2+63\xi^4+5\xi^6}{8} V_{43}$$
$$- \frac{105+298\xi^2+140\xi^4+15\xi^6}{12} V_{32}V_{21} + \frac{945+3169\xi^2+1811\xi^4+243\xi^6}{384} V_{21}^3$$

$$V_4 = -2(1+\xi^2)V_{24} + \frac{28(3+5\xi^2+\xi^4)}{3} V_{34} - \frac{15+32\xi^2+9\xi^4}{4} (V_{23}V_{21} + \tfrac{1}{2}V_{22}^2)$$

$$- \frac{3(75+173\xi^2+63\xi^4+5\xi^6)}{2} V_{44} + \frac{105+298\xi^2+140\xi^4+15\xi^6}{2} (\tfrac{1}{3}V_{22}V_{32} + V_{21}V_{33})$$

$$+ \frac{15+33\xi^2+11\xi^4+\xi^6}{4} V_{44} + \frac{735+2170\xi^2+1126\xi^4+168\xi^6+7\xi^8}{5} V_{54}$$

$$- \frac{945+3169\xi^2+1811\xi^4+243\xi^6}{64} V_{22}V_{21}^2 \, {}^{*} - \frac{945+3354\xi^2+2166\xi^4+425\xi^6+25\xi^8}{18} V_{32}^2$$

$$- \frac{4725+16586\xi^2+10514\xi^4+1974\xi^6+105\xi^8}{32} V_{21}V_{43}$$

$$+ \frac{10395+42429\xi^2+31938\xi^4+7335\xi^6+495\xi^8}{96} V_{32}V_{21}^2$$

$$- \frac{135135+626144\xi^2+542026\xi^4+145320\xi^6+11583\xi^8}{6144} V_{21}^4 .$$

Trickett and Welch (1954) computed the size of the Welch-Aspin test by means of the following integral:

$$Q_{V(\hat{C})}(0,C) = \int_0^1 T_{f_1+f_2}\left\{ \sqrt{f_1+f_2} \; \sqrt{\frac{C}{f_1} w + \frac{1-C}{f_2}(1-w)} \; V(\hat{C}(w)) \right\} p(w)\, dw \qquad (21)$$

where $T_{f_1+f_2}(x)$ is the upper tail of the Student-t integral with f_1+f_2 degrees of freedom and $p(w)$ is the probability density function of a Beta variate with parameters $f_1/2$ and $f_2/2$. Pagurova (1968) and Wang (1971) have obtained the same expression for $Q_{V(\hat{C})}(0,C)$, apparently unaware of the above result.

For $\alpha = .05$, Trickett and Welch (1954, p. 373) concluded (using our notation) "... direct calculation shows that the values of $P[v < V_{wa}(\hat{C}) | 0,C]$

---

* In Aspin (1948), this term appears with $V_{42}V_{21}^2$, but after checking the expansion, we find it should be corrected to read $V_{22}V_{21}^2$ as above. Perhaps it is a misprint in the article, but we have been unable to locate any other source for the expression $V_{wa}(\hat{C})$.

obtained at $f_1 = f_2 = 6$ by using the two-decimal $V_{wa}(\hat{C})$, never differ from 0.9500 by more than two units in the fourth place, whatever C." That is, the actual size of the Welch-Aspin test lies between .0498 and .0502 for $n_1 = n_2 = 7$. Welch (1949, p. 294) has given values at C = .1, .2, .3, .4 and .5 to the fourth decimal point and, according to Welch, "... the last decimal place given ... cannot ... be fully guaranteed". These values are listed here in Table 5. By fitting $V_{wa}(\hat{C})$ with a fourth-degree polynomial based on 11 critical values available in Aspin's tables (1948), Wang (1971, Table 3) computed the size of the Welch-Aspin test with "... maximum total error in the computation ... about ... .0002 ...". (Wang 1971, p. 606). By using (7), we have computed the actual size of the Welch-Aspin test and the values given below in Table 4 are correct to five decimal places. For comparison, the values obtained by Wang are also included in the table.

Table 4
Size of the Welch-Aspin Test
$n_1 = n_2 = 7, \alpha = .05$

C	Computed by		
	Welch (1949)	Wang (1971)	Formula (7)
.1	.0501	.0494	.05003
.2	.0500	.0495	.05010
.3	.0500	.0493	.05006
.4	.0498	.0492	.04998
.5	.0498	.0491	.04994

From these results, we can see that the values obtained by using (7) support Welch's conclusion mentioned above. Wang's values, on the other hand, show some discrepancy with this conclusion, perhaps due to the inadequacy of the fourth degree polynomial in the calculation.

We now turn to a comparison of size and power of the various tests mentioned above.

4.1 Comparison of size

The sizes of the various tests are listed in Table 5 for $(n_1, n_2, \alpha) = (5, 9, .05)$ and certain values of C. Plots are also given in Figure 3. As might have been anticipated, the tests corresponding to V_{mb} and V_{cc} are the most conservative, followed by that corresponding to V_f. The size of the test based on V_p lies mostly below $\alpha = .05$. Welch's approximate t test has size .0521 at C = .9 and manifests a wider variation than his series solution $V_w(\hat{C})$. The size of the Welch-Aspin test lies between .0500 and .0502, while that based on $V_o(\hat{C})$ varies from .0497 to .0503. Although Trickett and Welch (1954) were very cautious in using $V_{wa}(\hat{C})$ for n_1 or n_2 less than 7, Table 5 and Figure 3 show the fine performance of this test in controlling size. Its complexity, howeve limits its use in practice. On the other hand, the test based on $V_o(\hat{C})$ is much simpler and behaves virtually as well. In Table 6, where the size for $n_1 = n_2 = 7$ and $n_1 = n_2 = 5$ are given, the $V_o(\hat{C})$ test exhibits as effective a control of size as does the Welch-Aspin test for $n_1 = n_2 = 7$, and slightly better for $n_1 = n_2 = 5$.

4.2 Comparison of power

It is not meaningful to compare the power of tests when their sizes differ noticeably. Among the tests considered here, the size of the $V_{wa}(\hat{C})$ and $V_o(\hat{C})$ tests are very close and a comparison of their power appears in Table 7 for $\alpha = .05$ and $(n_1, n_2) = (7, 7)$. To our knowledge, no values of the exact power of the Welch-Aspin test have appeared anywhere in the literature. The percentage values in the table represent the ratio:

$$\frac{\text{Power of the } V(\hat{C}) \text{ test}}{\text{Power of the } u_R \text{ test when } R \text{ is known}}$$

where $V = V_o$ or V_{wa}. The actual value of the power of the u_R test is given in the last row of the table. From this table, one can see that the power of both $V_{wa}(\hat{C})$ and $V_o(\hat{C})$ tests is very high and also very close to each other.

Table 5

Comparison of Size $Q_{V(\hat{C})}(0,C)$

$\alpha = .05$

$n_1=5,\ n_2=9$

C \ V	V_{mb}	V_{co}	V_f	V_{wt}	V_w	V_p^*	V_{wa}	V_o
.001	.0499	.0499	.0501	.0500	.0502	.0500	.0500	.0500
.010	.0492	.0492	.0496	.0501	.0502	.0500	.0500	.0500
.050	.0464	.0465	.0476	.0502	.0503	.0499	.0500	.0501
.100	.0437	.0439	.0454	.0499	.0502	.0495	.0500	.0502
.150	.0416	.0418	.0435	.0495	.0501	.0491	.0501	.0502
.200	.0400	.0402	.0421	.0490	.0499	.0486	.0501	.0501
.250	.0387	.0390	.0410	.0486	.0496	.0482	.0501	.0500
.300	.0378	.0380	.0401	.0484	.0494	.0478	.0500	.0499
.350	.0371	.0374	.0395	.0482	.0493	.0475	.0500	.0498
.400	.0367	.0370	.0392	.0481	.0492	.0473	.0500	.0498
.450	.0365	.0368	.0390	.0482	.0492	.0472	.0500	.0497
.500	.0365	.0368	.0391	.0483	.0493	.0472	.0500	.0498
.550	.0367	.0370	.0393	.0486	.0495	.0473	.0500	.0498
.600	.0371	.0374	.0398	.0490	.0498	.0475	.0501	.0499
.650	.0377	.0380	.0404	.0495	.0500	.0478	.0501	.0501
.700	.0386	.0388	.0412	.0500	.0504	.0481	.0502	.0502
.750	.0396	.0399	.0423	.0506	.0507	.0484	.0502	.0503
.800	.0410	.0412	.0436	.0513	.0511	.0488	.0502	.0503
.850	.0426	.0428	.0451	.0518	.0514	.0492	.0502	.0502
.900	.0445	.0447	.0469	.0521	.0515	.0495	.0501	.0500
.950	.0470	.0471	.0489	.0519	.0514	.0497	.0500	.0498
.990	.0493	.0493	.0506	.0507	.0510	.0498	.0500	.0498
.999	.0499	.0499	.0510	.0501	.0510	.0500	.0500	.0500

* For V_p, $V = V_p(\tilde{C})$.

Fig. 3

Comparison of Size $Q_{V(\hat{C})}(0,C)$

$n_1 = 5$, $n_2 = 9$, $\alpha = .05$

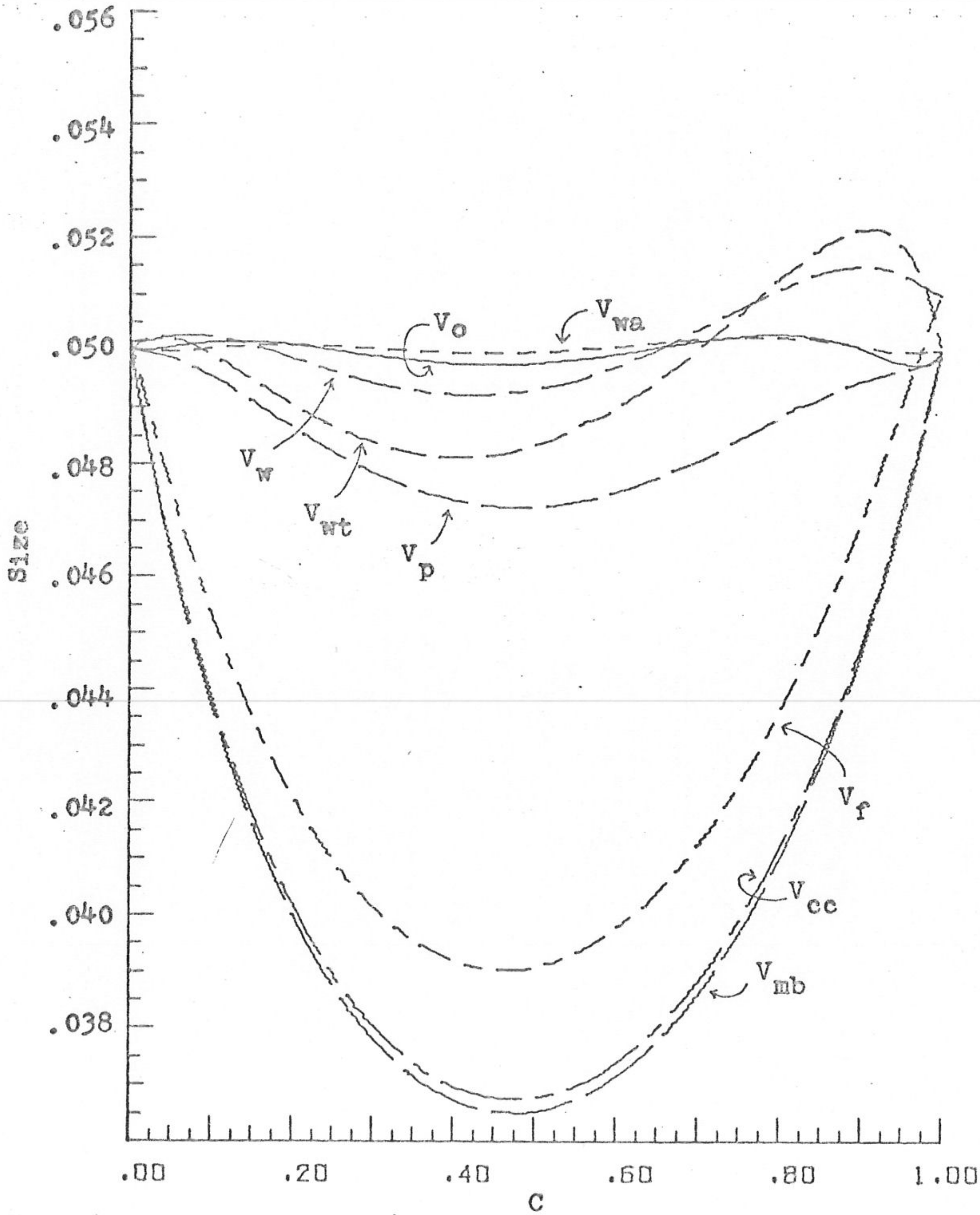


Table 6

Comparison of Size $Q_{V(\hat{C})}(0,C)$

$\alpha = .05$

C \ V	$V_{mb}=V_{oc}$	V_f	V_{wt}	V_w	V_{wd}	V_p^*	V_{wa}	V_o
				$n_1 = n_2 = 7$				
.001	.0499	.0503	.0500	.0503	.0500	.0500	.0500	.0500
.010	.0494	.0500	.0503	.0504	.0502	.0500	.0500	.0499
.050	.0472	.0484	.0507	.0505	.0505	.0498	.0500	.0499
.100	.0449	.0465	.0506	.0505	.0504	.0496	.0500	.0500
.150	.0431	.0449	.0503	.0504	.0501	.0493	.0501	.0501
.200	.0416	.0436	.0498	.0502	.0497	.0490	.0501	.0501
.250	.0404	.0425	.0494	.0500	.0492	.0486	.0501	.0501
.300	.0395	.0416	.0490	.0498	.0489	.0483	.0501	.0501
.350	.0388	.0410	.0487	.0496	.0485	.0480	.0500	.0500
.400	.0383	.0405	.0484	.0494	.0483	.0479	.0500	.0500
.450	.0380	.0402	.0483	.0493	.0482	.0477	.0499	.0499
.500	.0379	.0401	.0482	.0493	.0481	.0477	.0499	.0499
				$n_1 = n_2 = 5$				
.001	.0499	.0509	.0501	.0510	.0501	.0500	.0500	.0500
.010	.0489	.0503	.0506	.0511	.0505	.0499	.0500	.0498
.050	.0454	.0475	.0514	.0514	.0510	.0496	.0500	.0498
.100	.0421	.0445	.0510	.0514	.0506	.0490	.0503	.0501
.150	.0395	.0421	.0502	.0510	.0498	.0483	.0504	.0503
.200	.0375	.0402	.0493	.0504	.0490	.0475	.0505	.0503
.250	.0359	.0386	.0484	.0498	.0481	.0468	.0504	.0502
.300	.0348	.0375	.0477	.0493	.0474	.0462	.0503	.0501
.350	.0339	.0366	.0472	.0489	.0469	.0457	.0502	.0499
.400	.0333	.0360	.0467	.0486	.0465	.0453	.0501	.0498
.450	.0329	.0356	.0465	.0484	.0462	.0451	.0500	.0497
.500	.0328	.0355	.0464	.0483	.0462	.0450	.0500	.0497

* For V_p, $V = V_p(\tilde{C})$.

Table 7

Comparison of Power $Q_{V(\hat{C})}(\delta, C)$

for $n_1=7$, $n_2=7$, $\alpha=.05$

C	$\delta=1$ V_{wa}	$\delta=1$ V_o	$\delta=2$ V_{wa}	$\delta=2$ V_o	$\delta=3$ V_{wa}	$\delta=3$ V_o	$\delta=4$ V_{wa}	$\delta=4$ V_o
.001	.2251	.2251	.5498	.5498	.8405	.8405	.9685	.9685
	93%	93%	92%	92%	95%	95%	99%	99%
.010	.2256	.2254	.5510	.5508	.8416	.8414	.9689	.9689
	93%	93%	93%	93%	96%	96%	99%	99%
.050	.2275	.2272	.5563	.5557	.8463	.8460	.9707	.9706
	94%	94%	93%	93%	96%	96%	99%	99%
.100	.2301	.2299	.5628	.5659	.8521	.8518	.9729	.9727
	95%	95%	95%	95%	97%	97%	99%	99%
.150	.2326	.2326	.5692	.5691	.8578	.8576	.9749	.9748
	96%	96%	96%	96%	97%	97%	99%	99%
.200	.2349	.2349	.5753	.5754	.8631	.8631	.9767	.9767
	97%	97%	97%	97%	98%	98%	99%	99%
.250	.2368	.2369	.5808	.5809	.8680	.8680	.9784	.9784
	98%	98%	98%	98%	99%	99%	100%	100%
.300	.2385	.2385	.5855	.5856	.8722	.8723	.9799	.9799
	99%	99%	98%	98%	99%	99%	100%	100%
.350	.2397	.2397	.5893	.5894	.8757	.8757	.9810	.9811
	99%	99%	99%	99%	99%	99%	100%	100%
.400	.2406	.2406	.5920	.5920	.8782	.8782	.9819	.9819
	100%	100%	99%	99%	100%	100%	100%	100%
.450	.2411	.2411	.5937	.5937	.8798	.8798	.9825	.9825
	100%	100%	100%	100%	100%	100%	100%	100%
.500	.2413	.2413	.5942	.5942	.8803	.8803	.9826	.9826
	100%	100%	100%	100%	100%	100%	100%	100%
u_R	.2417		.5951		.8810		.9828	

4.2 Comparison of size

The sizes of the various tests are listed in Table 8 for $(n_1, n_2, \alpha) = (5, 9, .05)$ and certain values of C. Plots are also given in Figure 4 based on 53 values of C as mentioned above. As might have been anticipated, the tests corresponding to V_{mb} and V_{cc} are the most conservative, followed by that corresponding to V_f. The size of the test based on V_p lies mostly below $\alpha = .05$. Welch's approximate t test has size $.0521$ at $C = .9$ and manifests a wider variation than his series solution $V_w(\hat{C})$. The size of the Welch-Aspin test lies between $.0500$ and $.0502$, while that based on $V_o(\hat{C})$ varies from $.0497$ to $.0503$. Although Trickett and Welch (1954) were very cautious in using $V_{wa}(\hat{C})$ for n_1 or n_2 less than 7, Table 8 and Figure 4 show the fine performance of this test in controlling size. Its complexity, however, limits its use in practice. On the other hand, the test based on $V_o(\hat{C})$ is much simpler and behaves virtually as well. In Table 9, where the size for $n_1 = n_2 = 7$ and $n_1 = n_2 = 5$ are given, the $V_o(\hat{C})$ test exhibits as effective a control of size as does the Welch-Aspin test for $n_1 = n_2 = 7$, and slightly better for $n_1 = n_2 = 5$.

4.3 Comparison of power

It is not meaningful to compare the power of tests when their sizes differ noticeably. Among the tests considered here, the size of the $V_{wa}(\hat{C})$ and $V_o(\hat{C})$ tests are very close and a comparison of their power appears in Tables 10, 11 and 12 for $\alpha = .05$ and $(n_1, n_2) = (5, 9), (7, 7), (5, 5)$, respectively. To our knowledge, no values of the exact power of the Welch-Aspin test have appeared anywhere in the literature. The percentage values in these tables represent the ratio:

$$\frac{\text{Power of the } V(\hat{C}) \text{ test}}{\text{Power of the } u_R \text{ test when } R \text{ is known}}$$

where $V = V_o$ or V_{wa}. The actual value of the power of the u_R test is given in the last row of the tables. From these tables, one can see that the power of both $V_{wa}(\hat{C})$ and $V_o(\hat{C})$ tests is very high and also very close to each other.

Table 8

Comparison of Size $Q_{V(\hat{C})}(0,C)$

$\alpha = .05$

$n_1=5,\ n_2=9$

C \ V	V_{mb}	V_{co}	V_f	V_{wt}	V_w	V_p^*	V_{wa}	V_o
.001	.0499	.0499	.0501	.0500	.0502	.0500	.0500	.0500
.010	.0492	.0492	.0496	.0501	.0502	.0500	.0500	.0500
.050	.0464	.0465	.0476	.0502	.0503	.0499	.0500	.0501
.100	.0437	.0439	.0454	.0499	.0502	.0495	.0500	.0502
.150	.0416	.0418	.0435	.0495	.0501	.0491	.0501	.0502
.200	.0400	.0402	.0421	.0490	.0499	.0486	.0501	.0501
.250	.0387	.0390	.0410	.0486	.0496	.0482	.0501	.0500
.300	.0378	.0380	.0401	.0484	.0494	.0478	.0500	.0499
.350	.0371	.0374	.0395	.0482	.0493	.0475	.0500	.0498
.400	.0367	.0370	.0392	.0481	.0492	.0473	.0500	.0498
.450	.0365	.0368	.0390	.0482	.0492	.0472	.0500	.0497
.500	.0365	.0368	.0391	.0483	.0493	.0472	.0500	.0498
.550	.0367	.0370	.0393	.0486	.0495	.0473	.0500	.0498
.600	.0371	.0374	.0398	.0490	.0498	.0475	.0501	.0499
.650	.0377	.0380	.0404	.0495	.0500	.0478	.0501	.0501
.700	.0386	.0388	.0412	.0500	.0504	.0481	.0502	.0502
.750	.0396	.0399	.0423	.0506	.0507	.0484	.0502	.0503
.800	.0410	.0412	.0436	.0513	.0511	.0488	.0502	.0503
.850	.0426	.0428	.0451	.0518	.0514	.0492	.0502	.0502
.900	.0445	.0447	.0469	.0521	.0515	.0495	.0501	.0500
.950	.0470	.0471	.0489	.0519	.0514	.0497	.0500	.0498
.990	.0493	.0493	.0506	.0507	.0510	.0498	.0500	.0498
.999	.0499	.0499	.0510	.0501	.0510	.0500	.0500	.0500

* For V_p, $V = V_p(\tilde{C})$.

Fig. 4

Comparison of Size $Q_{V(\hat{C})}(0,C)$

$n_1 = 5$, $n_2 = 9$, $\alpha = .05$

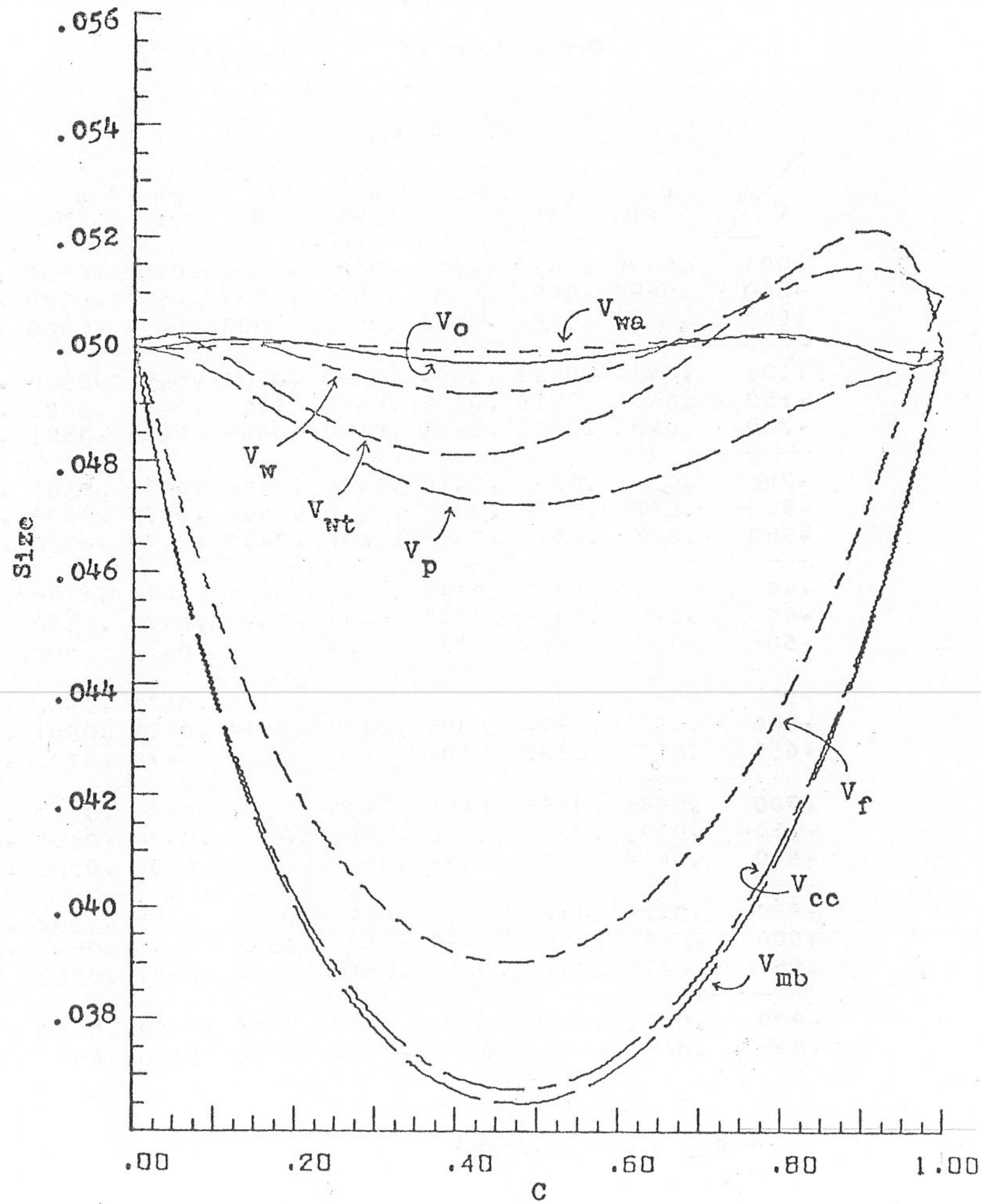


Table 9

Comparison of Size $Q_{V(\hat{C})}(0,C)$

$\alpha = .05$

$n_1 = n_2 = 7$

C \ V	$V_{mb}=V_{oc}$	V_f	V_{wt}	V_w	V_{wd}	V_p^*	V_{wa}	V_o
.001	.0499	.0503	.0500	.0503	.0500	.0500	.0500	.0500
.010	.0494	.0500	.0503	.0504	.0502	.0500	.0500	.0499
.050	.0472	.0484	.0507	.0505	.0505	.0498	.0500	.0499
.100	.0449	.0465	.0506	.0505	.0504	.0496	.0500	.0500
.150	.0431	.0449	.0503	.0504	.0501	.0493	.0501	.0501
.200	.0416	.0436	.0498	.0502	.0497	.0490	.0501	.0501
.250	.0404	.0425	.0494	.0500	.0492	.0486	.0501	.0501
.300	.0395	.0416	.0490	.0498	.0489	.0483	.0501	.0501
.350	.0388	.0410	.0487	.0496	.0485	.0480	.0500	.0500
.400	.0383	.0405	.0484	.0494	.0483	.0479	.0500	.0500
.450	.0380	.0402	.0483	.0493	.0482	.0477	.0499	.0499
.500	.0379	.0401	.0482	.0493	.0481	.0477	.0499	.0499

$n_1 = n_2 = 5$

C \ V	$V_{mb}=V_{oc}$	V_f	V_{wt}	V_w	V_{wd}	V_p^*	V_{wa}	V_o
.001	.0499	.0509	.0501	.0510	.0501	.0500	.0500	.0500
.010	.0489	.0503	.0506	.0511	.0505	.0499	.0500	.0498
.050	.0454	.0475	.0514	.0514	.0510	.0496	.0500	.0498
.100	.0421	.0445	.0510	.0514	.0506	.0490	.0503	.0501
.150	.0395	.0421	.0502	.0510	.0498	.0483	.0504	.0503
.200	.0375	.0402	.0493	.0504	.0490	.0475	.0505	.0503
.250	.0359	.0386	.0484	.0498	.0481	.0468	.0504	.0502
.300	.0348	.0375	.0477	.0493	.0474	.0462	.0503	.0501
.350	.0339	.0366	.0472	.0489	.0469	.0457	.0502	.0499
.400	.0333	.0360	.0467	.0486	.0465	.0453	.0501	.0498
.450	.0329	.0356	.0465	.0484	.0462	.0451	.0500	.0497
.500	.0328	.0355	.0464	.0483	.0462	.0450	.0500	.0497

* For V_p, $V = V_p(\tilde{C})$.

Table 10

Comparison of Power $Q_{V(\hat{C})}(\delta,C)$

for $n_1=5$, $n_2=9$, $\alpha=.05$

C	δ=1 V_{wa}	δ=1 V_o	δ=2 V_{wa}	δ=2 V_o	δ=3 V_{wa}	δ=3 V_o	δ=4 V_{wa}	δ=4 V_o
.001	.2333 97%	.2333 97%	.5726 96%	.5726 96%	.8619 98%	.8619 98%	.9767 99%	.9767 99%
.010	.2336 97%	.2336 97%	.5736 96%	.5736 96%	.8627 98%	.8627 98%	.9769 99%	.9769 99%
.100	.2372 98%	.2376 98%	.5826 98%	.5831 98%	.8702 99%	.8705 99%	.9794 100%	.9794 100%
.200	.2403 99%	.2404 99%	.5907 99%	.5909 99%	.8770 100%	.8772 100%	.9815 100%	.9816 100%
.300	.2415 100%	.2411 100%	.5944 100%	.5940 100%	.8803 100%	.8801 100%	.9826 100%	.9826 100%
.400	.2408 100%	.2402 99%	.5926 100%	.5919 99%	.8787 100%	.8783 100%	.9820 100%	.9820 100%
.500	.2385 99%	.2379 98%	.5855 98%	.5849 98%	.8718 99%	.8716 99%	.9795 100%	.9795 100%
.600	.2347 97%	.2345 97%	.5737 96%	.5736 96%	.8601 98%	.8602 98%	.9750 99%	.9750 99%
.700	.2297 95%	.2297 95%	.5581 94%	.5584 94%	.8447 96%	.8449 96%	.9688 99%	.9688 99%
.800	.2233 92%	.2235 92%	.5401 91%	.5402 91%	.8271 94%	.8269 94%	.9614 98%	.9612 98%
.900	.2162 89%	.2159 89%	.5216 88%	.5206 87%	.8090 92%	.8079 92%	.9532 97%	.9527 97%
.990	.2105 87%	.2100 87%	.5061 85%	.5055 85%	.7930 90%	.7926 90%	.9453 96%	.9452 96%
.999	.2099 87%	.2098 87%	.5046 85%	.5045 85%	.7913 90%	.7912 90%	.9445 96%	.9445 96%
u_R	.2417		.5951		.8810		.9828	

Table 11

Comparison of Power $Q_{V(\hat{C})}(\delta,C)$

for $n_1=7$, $n_2=7$, $\alpha=.05$

C	δ=1 V_{wa}	δ=1 V_o	δ=2 V_{wa}	δ=2 V_o	δ=3 V_{wa}	δ=3 V_o	δ=4 V_{wa}	δ=4 V_o
.001	.2251 93%	.2251 93%	.5498 92%	.5498 92%	.8405 95%	.8405 95%	.9685 99%	.9685 99%
.010	.2256 93%	.2254 93%	.5510 93%	.5508 93%	.8416 96%	.8414 96%	.9689 99%	.9689 99%
.050	.2275 94%	.2272 94%	.5563 93%	.5557 93%	.8463 96%	.8460 96%	.9707 99%	.9706 99%
.100	.2301 95%	.2299 95%	.5628 95%	.5659 95%	.8521 97%	.8518 97%	.9729 99%	.9727 99%
.150	.2326 96%	.2326 96%	.5692 96%	.5691 96%	.8578 97%	.8576 97%	.9749 99%	.9748 99%
.200	.2349 97%	.2349 97%	.5753 97%	.5754 97%	.8631 98%	.8631 98%	.9767 99%	.9767 99%
.250	.2368 98%	.2369 98%	.5808 98%	.5809 98%	.8680 99%	.8680 99%	.9784 100%	.9784 100%
.300	.2385 99%	.2385 99%	.5855 98%	.5856 98%	.8722 99%	.8723 99%	.9799 100%	.9799 100%
.350	.2397 99%	.2397 99%	.5893 99%	.5894 99%	.8757 99%	.8757 99%	.9810 100%	.9811 100%
.400	.2406 100%	.2406 100%	.5920 99%	.5920 99%	.8782 100%	.8782 100%	.9819 100%	.9819 100%
.450	.2411 100%	.2411 100%	.5937 100%	.5937 100%	.8798 100%	.8798 100%	.9825 100%	.9825 100%
.500	.2413 100%	.2413 100%	.5942 100%	.5942 100%	.8803 100%	.8803 100%	.9826 100%	.9826 100%
u_R	.2417		.5951		.8810		.9828	

Table 12

Comparison of Power $Q_{V(\hat{C})}(\delta,C)$

for $n_1=5$, $n_2=5$, $\alpha=.05$

C	δ=1		δ=2		δ=3		δ=4	
	V_{wa}	V_o	V_{wa}	V_o	V_{wa}	V_o	V_{wa}	V_o
.001	.2099	.2098	.5046	.5044	.7913	.7912	.9445	.9445
	90%	90%	88%	88%	92%	92%	97%	97%
.010	.2105	.2099	.5061	.5054	.7930	.7925	.9453	.9452
	90%	90%	88%	88%	92%	92%	97%	97%
.050	.2128	.2122	.5126	.5112	.7999	.7988	.9489	.9484
	91%	91%	90%	89%	93%	93%	97%	97%
.100	.2165	.2160	.5215	.5205	.8087	.8078	.9531	.9526
	93%	93%	91%	91%	94%	94%	98%	98%
.150	.2202	.2199	.5310	.5304	.8179	.8173	.9573	.9569
	94%	94%	93%	93%	95%	95%	98%	98%
.200	.2236	.2232	.5403	.5397	.8271	.8267	.9613	.9611
	96%	96%	94%	94%	96%	96%	98%	98%
.250	.2264	.2258	.5487	.5481	.8357	.8354	.9651	.9650
	97%	97%	96%	96%	97%	97%	99%	99%
.300	.2285	.2278	.5558	.5550	.8433	.8430	.9685	.9684
	98%	98%	97%	97%	98%	98%	99%	99%
.350	.2301	.2293	.5615	.5605	.8496	.8491	.9713	.9712
	99%	98%	98%	98%	99%	99%	99%	99%
.400	.2312	.2303	.5656	.5645	.8543	.8537	.9734	.9733
	99%	99%	99%	99%	99%	99%	100%	100%
.450	.2319	.2308	.5681	.5668	.8571	.8565	.9747	.9746
	99%	99%	99%	99%	99%	99%	100%	100%
.500	.2321	.2310	.5689	.5676	.8581	.8574	.9751	.9750
	100%	99%	99%	99%	100%	99%	100%	100%
u_R	.2332		.5725		.8618		.9767	

REFERENCES

1. Abramowitz, M. and Stegun, I.A. (1965). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, National Bureau of Standards, Applied Mathematics Series, 55, 3rd printing.

2. Aspin, A. A. (1948). "An examination and further development of a formula arising in the problem of comparing two mean values," Biometrika 35, 88-96.

3. Banerjee, S. K. (1960). "Approximate confidence interval for linear functions of means of k populations when the population variances are not equal," Sankhyā 22, 357-358.

4. Banerjee, S. K. (1961). "On confidence interval for two-means problem based on separate estimates of variances and tabulated values of t-table," Sankhyā, A, 23, 359-378.

5. Bartlett, M.S. (1956). "Comment on Sir Ronald Fisher's paper "On a test of significance in Pearson's Biometrika Tables (No. 11)'," J. Royal Stat. Soc., B, 18, 295-296.

6. Bennett, B. M. and Hsu, P. (1961). "Sampling studies on the Behrens-Fisher problem," Metrika, 4, 89-104.

7. Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis, Forecasting and control, Holden-Day, San Francisco.

8. Cochran, W. G. (1964). "Approximate significance levels of the Behrens-Fisher test," Bilmetrics 20, 191-195.

9. Cochran, W. G. and Cox, G. M. (1950). Experimental Designs, John Wiley, New York.

10. Fisher, R. A. (1935). "The fiducial argument in statistical inference," Ann. Eugen., 6, 391-398.

11. Fisher, R. A. (1941). "The asymptotic approach to Behrens' integral, with further tables for the d test of significance," Ann. Eugen., 11, 141-172.

12. Golhar, M. B. (1964). "On the comparison of two means from normal populations with unknown variances," Indian Society of Agriculture Statistics, Journal, 16, 62-71.

13. Gurland, J. (1962). "Note on a paper by Ray and Pitman," J. Royal Stat. Soc., B, 24, 537-538.

14. Lee, A. F. S. (1972). Inference Concerning Means of Two Normal Populations, Ph.D. Thesis, Department of Statistics, University of Wisconsin, Madison.

15. Linnik, Yu. V. (1964). "On A. Wald's test comparing two normal samples," Theory of Probability and Its Applications, 9, 14-27.

16. Linnik, Yu. V. (1966). "Latest investigations on Behrens-Fisher problem," Sankhyā, A, 28, 15-24.

17. Lubkin, S. (1952). "A method of summing infinite series," Journal of Research, National Bureau of Standards, U.S. 48, 228-54.

18. McCullough, R. S., Gurland, J. and Rosenberg, L. (1960). "Small sample behaviour of certain tests of the hypothesis of equal means under variance heterogeneity," Biometrika 47, 345-353.

19. Mehta, J. S. and Srinivasan, R. (1970). "On the Behrens-Fisher problem," Biometrika 57, 649-655.

20. Murphy, B. P. (1967). "Some two-sample tests when the variances are unequal: a simulation study," Biometrika, 54, 679-683.

21. Owen, D. B. (1956). "Tables for computing bivariate normal probabilities," Ann. Math. Stat., 27, 1075-1090.

22. Owen, D. B. (1968). "A survey of properties and applications of the noncentral t-distribution," Technometrics 10, 445-478.

23. Pagurova, V. I. (1968). "On a comparison of means of two normal samples," Theory of Probability and Its Applications, 13, 527-534.

24. Scheffé, H. (1970). "Practical solutions of the Behrens-Fisher problem," J. Amer. Stat. Assoc., 65, 1501-1508.

25. Shanks, D. (1955). "Non-linear transformations of divergent and slowly convergent sequences," J. Math. Phys., 34, 1-42.

26. Trickett, W. H. and Welch, B. L. (1954). "On the comparison of two means: furhter discussion of iterative methods for calculating tables," Biometrika 41, 361-74.!

27. Wald, A. (1955). "Testing the difference between the means of two normal populations with unknown standard deviations," Selected Papers in Statistics and Probability by Abraham Wald, McGraw Hill, New York, 669-695.

28. Wallace, D. L. (1958). "Asymptotic approximations to distributions," Ann. Math. Stat., 29, 635-654.

29. Wang, Y. Y. (1971). "Probabilities of the type I errors of the Welch tests for the Behrens-Fisher problem," J. Amer. Stat. Assoc., 66, 605-608.

30. Welch, B. L. (1947). "A generalization of 'Students' problem," Biometrika 34, 28-35.

31. Welch, B. L. (1949). "Further note on Mrs. Aspin's Tables and on certain approximations to the tables function," Biometrika 36, 293-6.

32. Wilks, S. S. (1940). "On the problem of two samples from normal populations with unequal variances," Ann. Math. Stat., 11, 475-476.