The Annals of Mathematical Statistics
1969, Vol. 40, No. 4, 1374–1385

A TEST OF EQUALITY OF TWO NORMAL POPULATION MEANS ASSUMING HOMOGENEOUS COEFFICIENTS OF VARIATION[1]

By R. K. Lohrding

University of California, Los Alamos Scientific Laboratory

1. Introduction and summary. In 1929 Behrens introduced the concept of testing equality of two population means without assuming the homogeneity of the two population variances. Since that time there have been extensive discussions of the validity of the test and of the interpretation of the results.

Fisher (1939) published a detailed paper defending Behrens' work and thus the work started by Behrens became known as the Behrens-Fisher problem. Fisher presented the test statistic (unequal variance t-test; Fryer, 1966) he felt was best for handling this situation. However, Fisher could only approximate the distribution of his test statistic. This approximation was tabled in Fisher and Yates (1948). Fisher's test statistic can also be approximated by the t-distribution, but the t-distribution approximation is not very good for small sample sizes (Cochran and Cox, 1957).

Box (1954, a and b) wrote two extensive papers concerned with the violation of the assumptions in the analysis of variance. Box states that the one-way analysis of variance with equal sample sizes (the t-test is the same as a two-sample one-way analysis of variance) is robust when the variances are heterogeneous. However, little is said about the power of the test when the variances are heterogeneous.

F. N. David and N. L. Johnson (1951, a and b) and Bozivich, Bancroft, and Hartley (1956) have written papers on the approximate theoretical power of the analysis of variance when the assumptions are violated. B. L. Welch (1937) and D. G. C. Gronow (1951) have written papers examining both the robustness and the power of the unequal variance t-test when the variances are heterogeneous.

This paper examines by simulation techniques both the power and the robustness of the t-test and several other tests when the variances are heterogeneous, and presents a new test statistic designed for the situation where the coefficients of variation are homogeneous. This new test is more powerful than the t-test for certain ranges of the coefficient of variation. The assumption of homogeneous coefficients of variation is a valid assumption in many types of agricultural, biological, and psychological experimentation, because many times the treatment that yields a larger mean also has a larger standard deviation.

Received 7 November 1968; revised 22 January 1969.

[1] Work performed in part under the auspices of the U. S. Atomic Energy Commission, and part under N.I.H. grant number HEW 0740. Parts of the results presented here are contained in a Ph.D. dissertation written by the author at Kansas State University.

2. A test of equality of two population means under the assumption of homogeneous coefficients of variation and normal populations. The testing procedure to be used in this paper will be the likelihood ratio test procedure.

THEOREM 2.1. *Let* X_{ij} *be independently distributed* $N(\mu_i, \beta^2\mu_i^2)$, $i = 1, 2$, *and* $j = 1, 2, \cdots, J$, *where* $\beta = \sigma_i/\mu_i$ *is the coefficient of variation. Let the non-restrictive parameter space be* Ω: $\{0 < \mu_1 < \infty, 0 < \mu_2 < \infty, 0 < \beta < \infty\}$.
Then the maximum likelihood estimators are

$$\hat{\mu}_1 = \bar{X}_1/2 + [\bar{X}_2/2][(\bar{X}_1^2 + 2S_1^2)/(\bar{X}_2^2 + 2S_2^2)]^{\frac{1}{2}} \tag{2.1}$$

$$\hat{\mu}_2 = \bar{X}_2/2 + [\bar{X}_1/2][(\bar{X}_2^2 + 2S_2^2)/(\bar{X}_1^2 + 2S_1^2)]^{\frac{1}{2}}, \tag{2.2}$$

and

$$\hat{\beta} = \frac{\left[\begin{array}{l} 2[(\bar{X}_1^2 + 2S_1^2)(\bar{X}_2^2 + 2S_2^2)]^{\frac{1}{2}} \\ \quad \cdot \{[(\bar{X}_1^2 + 2S_1^2)(\bar{X}_2^2 + 2S_2^2)]^{\frac{1}{2}} - \bar{X}_1\bar{X}_2\}\end{array}\right]^{\frac{1}{2}}}{\bar{X}_1[\bar{X}_2^2 + 2S_2^2]^{\frac{1}{2}} + \bar{X}_2[\bar{X}_1^2 + 2S_1^2]^{\frac{1}{2}}} \tag{2.3}$$

where $\bar{X}_i$ is the average of the ith sample and S_i^2 is the maximum likelihood estimate of the ith sample variance.

PROOF. The likelihood function is

$$L(\Omega) = [2\pi]^{-J}\{\textstyle\prod_{i=1}^2 [\beta\mu_i]^{-J}\} \exp\{-\frac{1}{2}\sum_{i=1}^2 [\sum_{j=1}^J (X_{ij} - \mu_i)^2]/(\beta\mu_i)^2\}.$$

Taking derivatives of $\log L(\Omega)$ with respect to the parameters and setting them equal to zero gives the following set of equations:

$$\beta^2 = (1/2J)\textstyle\sum_{i=1}^2 \sum_{j=1}^J [(X_{ij} - \mu_i)/\mu_i]^2, \tag{2.4}$$

and

$$\beta^2\mu_k^2 + \bar{X}_k\mu_k - (\bar{X}_k^2 + S_k^2) = 0, \qquad k = 1, 2. \tag{2.5}$$

Thus, in (2.4) and (2.5) we have three equations and three unknowns which do not yield simple solutions due to the relationship between β and the μ_k's. (2.5) has two real roots, one positive and one negative, thus μ_k may always be chosen positive (Peterson, 1947).

To solve this set of equations, let us start by solving (2.5) for β^2, thus

$$\beta^2 = (\bar{X}_k^2 + S_k^2 - \bar{X}_k\mu_k)/\mu_k^2, \qquad k = 1, 2. \tag{2.6}$$

The two equations in (2.6) are added together to give

$$\beta^2 = [(\bar{X}_1^2 + S_1^2 - \bar{X}_1\mu_1)/2\mu_1^2] + (\bar{X}_2^2 + S_2^2 - \bar{X}_2\mu_2)/2\mu_2^2. \tag{2.7}$$

Therefore, subtracting (2.7) from (2.4) gives

$$[(\mu_1^2 - \bar{X}_1\mu_1)/2\mu_1^2] + (\mu_2^2 - \bar{X}_2\mu_2)/2\mu_2^2 = 0.$$

Solving for μ_1 in terms of μ_2 gives $\mu_1 = \bar{X}_1\mu_2/(2\mu_2 - \bar{X}_2)$. The two equations in (2.6) are set equal, giving

$$(\bar{X}_1^2 + S_1^2 - \bar{X}_1\mu_1)/\mu_1^2 = (\bar{X}_2^2 + S_2^2 - \bar{X}_2\mu_2)/\mu_2^2. \tag{2.8}$$

The value of μ_1 in terms of μ_2 is substituted into (2.8), giving an equation in μ_2 of the form

$$\mu_2^2 - \bar{X}_2\mu_2 + (\bar{X}_2^2S_1^2 - \bar{X}_1^2S_2^2)/2(\bar{X}_1^2 + 2S_1^2) = 0.$$

The quadratic formula gives a solution which simplifies to

$$\hat{\mu}_2 = \bar{X}_2/2 \pm [\bar{X}_1/2][(\bar{X}_2^2 + 2S_2^2)/(\bar{X}_1^2 + 2S_1^2)]^{\frac{1}{2}}. \tag{2.9}$$

Using the root with the minus sign one gets $L(\Omega) = 0$ under the null hypothesis, and this gives a nonsense result for the likelihood ratio test of Theorem 2.3. The root using the plus sign may give a negative estimator for μ, although under the assumptions of Theorem 2.1 the probability of $\mu_k < 0$ tends to zero as J tends to infinity.

Due to symmetry of the likelihood equations in the μ's, the estimate of μ_1 is

$$\hat{\mu}_1 = \bar{X}_1/2 + [\bar{X}_2/2][(\bar{X}_1^2 + 2S_1^2)/(\bar{X}_2^2 + 2S_2^2)]^{\frac{1}{2}}. \tag{2.10}$$

The estimate of β is found by using the values for $\hat{\mu}_1$ and $\hat{\mu}_2$ in (2.9) and (2.10) and substituting them into (2.4), giving (2.3).

This set of estimators solves the set of likelihood equations. Now a check must be made to see if the joint estimators $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\beta}$ define a maximum.

If the matrix

$$A = \left[\frac{\partial \log (\Omega)}{\partial\theta_r\partial\theta_s}\right]_{\hat{\theta}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \tag{2.11}$$

where $\theta_1 = \mu_1$, $\theta_2 = \mu_2$, and $\theta_3 = \beta$, is negative definite, then the joint estimators $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\beta}$ define a maximum (Kendall and Stuart, 1961), where

$$A = \begin{bmatrix} \dfrac{J\hat{\beta}^2\hat{\mu}_1^2 + 2J\bar{X}_1\hat{\mu}_1 - 3\sum_{j=1}^{J} X_{1j}^2}{\hat{\beta}^2\hat{\mu}_1^4} & 0 & \dfrac{2J\bar{X}_1\hat{\mu}_1 - 2\sum_{j=1}^{J} X_{1j}^2}{\hat{\beta}^3\hat{\mu}_1^3} \\ 0 & \dfrac{J\hat{\beta}^2\hat{\mu}_2 + 2J\bar{X}_2\hat{\mu}_2 - 3\sum_{j=1}^{J} X_{2j}^2}{\hat{\beta}^2\hat{\mu}_2^4} & \dfrac{2J\bar{X}_2\hat{\mu}_2 - 2\sum_{j=1}^{J} X_{2j}^2}{\hat{\beta}^3\hat{\mu}_2^3} \\ \dfrac{2J\bar{X}_1\hat{\mu}_1 - 2\sum_{j=1}^{J} X_{1j}^2}{\hat{\beta}^3\hat{\mu}_1^3} & \dfrac{2J\bar{X}_2\hat{\mu}_2 - 2\sum_{j=1}^{J} X_{2j}^2}{\hat{\beta}^3\hat{\mu}_2^3} & \dfrac{2J\hat{\beta}^2 - 3\sum_{i=1}^{2}\sum_{j=1}^{J}\hat{\mu}_i^{-2}(X_{ij} - \hat{\mu}_i)^2}{\hat{\beta}^4} \end{bmatrix}$$

To show A negative definite involves a theorem in Hohn (1965, page 349) on negative definiteness and considerable manipulation of inequalities. The details are contained in Lohrding (1969). This completes the proof of Theorem 2.1.

We have shown that the estimators $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\beta}$ are joint maximum likelihood estimators in the non-restricted parameter space; now the estimators will be found for the restricted parameter space.

THEOREM 2.2. *Let X_{ij} be independently distributed, $N(\mu, \mu^2\beta^2)$, $i = 1, 2$, and $j = 1\ 2, \cdots, J$, where $\beta = \sigma/\mu$ is the coefficient of variation. Let the restricted parameter space be ω: $\{0 < \mu < \infty, 0 < \beta < \infty\}$. Then the maximum likelihood estimators are*

$$\hat{\mu} = \bar{X} \tag{2.12}$$

and

$$\hat{\beta} = S/\bar{X} \tag{2.13}$$

where $\bar{X}$ is the grand average over the two samples and S is the maximum likelihood estimate of the standard deviation over both samples.

PROOF. The likelihood function is

$$L(\omega) = [2\pi]^{-J}[\beta\mu]^{-2J} \exp\left[-(2\beta^2\mu^2)^{-1} \sum_{i=1}^{2} \sum_{j=1}^{J} (X_{ij} - \mu)^2\right].$$

Taking the derivative of $\log L(\omega)$ with respect to β and setting the derivative to zero, we have

$$\beta^2 = (2J)^{-1} \sum_{i=1}^{2} \sum_{j=1}^{J} [(X_{ij} - \mu)/\mu]^2. \tag{2.14}$$

Taking the derivative with respect to μ and setting the derivative equal to zero gives the following solution for β^2, where

$$\beta^2 = (2J\mu^2)^{-1} \sum_{i=1}^{2} \sum_{j=1}^{J} (X_{ij}^2 - X_{ij}\mu). \tag{2.15}$$

If we subtract (2.14) from (2.15), we have

$$\sum_{i=1}^{2} \sum_{j=1}^{J} (X_{ij}^2 - 2X_{ij}\mu + \mu^2 - X_{ij}^2 + X_{ij}\mu) = 0,$$

which implies $\hat{\mu} = \bar{X}$. The value for $\hat{\mu}$ is inserted into (2.14), giving $\hat{\beta} = S/\bar{X}$.

It is easily shown that the estimators of μ and β do actually maximize the likelihood function. To do this the matrix

$$A = \left[\frac{\partial^2 \log L(\omega)}{\partial\theta_r \partial\theta_s}\right]_{\hat{\theta}}, \qquad \text{where } r, s = 1, 2; \quad \text{and} \quad \theta_1 = \mu, \theta_2 = \beta,$$

must be shown to be negative definite as was explained in the proof of Theorem 2.1. This completes the proof.

THEOREM 2.3. *Let X_{ij} be independently distributed, $N(\mu_i, \mu_i^2\beta^2)$, $i = 1, 2$, and $j = 1, 2, \cdots, J$. Assume homogeneous coefficients of variation $\beta = \sigma_1/\mu_1 = \sigma_2/\mu_2$. If the hypothesis $H_0(\mu_1 = \mu_2)$ is tested against the alternative $H_a(\mu_1 \neq \mu_2)$, then the likelihood ratio is*

$$\lambda_1 = L(\hat{\omega})/L(\hat{\Omega}) = [\{[(\bar{X}_1^2 + 2S_1^2)(\bar{X}_2^2 + 2S_2^2)]^{\frac{1}{2}} - \bar{X}_1\bar{X}_2\}/(2S^2)]^J, \tag{2.16}$$

where $\bar{X}_i$ and S_i^2 are defined in Theorem 2.1 and S^2 is defined in Theorem 2.2.

PROOF. The estimators $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\beta}$ from Theorem 2.1 are inserted into $L(\Omega)$, and the estmators $\hat{\mu}$ and $\hat{\beta}$ from Theorem 2.2 are inserted into $L(\omega)$. The result is (2.16). This completes the proof of Theorem 2.3.

3. Determining the large sample distribution of the likelihood ratio. The next problem is to determine the distribution of λ_1; however, this is a very difficult task. The difficulty arises from the fact that the distribution of a product of functions of random variables is generally difficult to find, and these particular functions are quite complicated. Therefore, the approach will be to investigate the asymptotic distribution of λ_1 for sample sizes larger than 40 and to simulate the distribution for samples less than or equal to 40.

A theorem will now be stated concerning the large sample distribution of the likelihood ratio.

Let X_{ij}, $i = 1, 2, \cdots, I$ *and* $j = 1, 2, \cdots, J$ *be a random sample from the cumulative distribution function* $F_i(x; \theta)$, *where* θ *is r-dimensional and* $F_i(x; \theta)$ *is regular in all of its second partial* θ*-derivatives, for* θ *in the non-restricted parameter space. Then if the hypothesis* $H_0(\mu_1 = \mu_2)$ *is true,* $-2 \log \lambda$ *converges in probability to a random variable having the chi-square distribution with* $r - r'$ *degrees of freedom, where r is the number of parameters in the* Ω *and* r' *is the number of parameters in* ω (Wilks, 1963).

The cumulative distribution function, $F_i(x; \theta)$, is regular in all of its first and second partial θ-derivatives in the non-restrictive parameter space.

This implies that $-2 \log \lambda_1$ converges in probability to a random variable having a chi-square distribution with one degree of freedom, where

$$-2 \log \lambda_1 = 2J \log (2S^2) - 2J \log ([(\bar{X}_1^2 + 2S_1^2)(\bar{X}_2^2 + 2S_2^2)]^{\frac{1}{2}} - \bar{X}_1\bar{X}_2).$$

The fact that $F_i(x; \theta)$ is regular in all its second partial derivatives implies that the likelihood ratio test λ_1 is consistent; that is,

$$\lim_{n\to\infty} P(-2 \log \lambda_1 > \chi_\alpha^2 \mid \theta \in \Omega - \omega) = 1,$$

where χ_α^2 is the $(1 - \alpha)$th percentile of the chi-square distribution with one degree of freedom (Wilks, 1963).

4. Determining the small sample distribution of the likelihood ratio. We have shown that $-2 \log \lambda_1$ converges in probability to a random variable having a chi-square distribution with one degree of freedom, $(\chi^2(1))$, but how good is this asymptotic result for small sample sizes?

To investigate small sample sizes, 10,000 test statistics were computed from random samples, assuming the null hypothesis, H_0 $(\mu_1 = \mu_2)$, to be true. Each of the 10,000 test statistics was computed by generating the desired sample size using the Randu random number generator from I.B.M. scientific subroutines. Randu generates random rectangular $(0, 1)$ variates, which are then transformed into $N(\mu, \sigma^2)$ variates. These sample values are used to compute the test statistics.

The Kolmogorov test, explained in Conover (1956b), has a critical value of .0122 at $\alpha = .10$, $n = 10{,}000$, so that the distribution function we obtained is everywhere within .0122 of the true unknown distribution function, with probability .90.

Now we order the 10,000 values of $-2 \log \lambda_1$ and graph the cumulative frequency curve. We see in Figure 1, for a sample size of 20, that $-2 \log \lambda_1$ is closely approximated by the chi-square distribution with one degree of freedom. The distribution of $-2 \log \lambda_1$ for a sample of size 6 has the same general shape as the chi-square distribution with one degree of freedom; but by looking at Figure 1, we can see that the value 2.7, which corresponds to the 90th quantile of the chi-square distribution with one degree of freedom, would correspond to approximately the 87th quantile of the $-2 \log \lambda_1$ graph with a sample size of 6 and the 89th quantile for sample size of 20, so the approximation is not conservative.

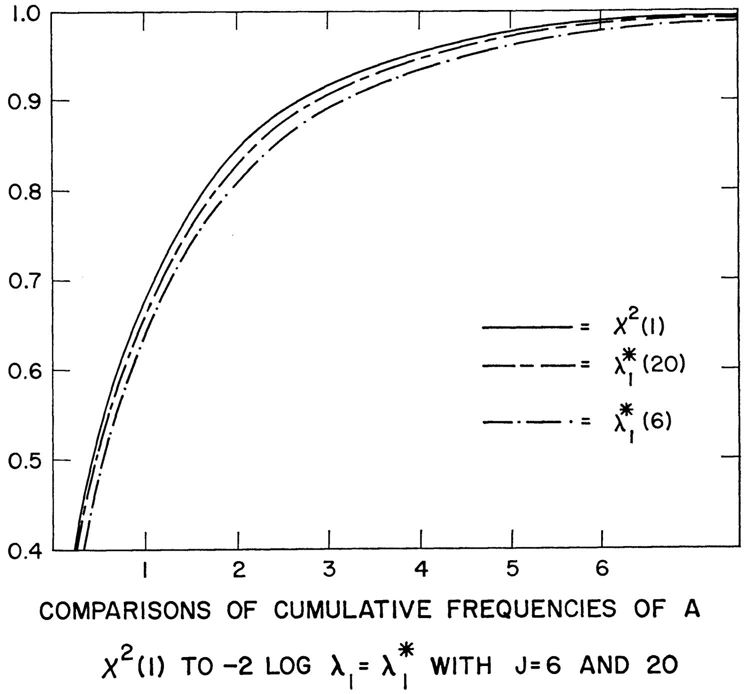

FIG. 1

Although the distribution of $-2 \log \lambda_1$ for a sample size of 20 is closely approximated by the chi-square distribution with one degree of freedom, it was decided to simulate 10,000 test statistics to give the rejection regions for $\alpha = .10$, .05, .025, and .01, for selected sample sizes less than or equal to 40. The results are presented in Table 1.

We can see that due to sampling error, not all the values of Table 1 are monotonic (see α-level .01 and sample sizes 10, 14, and 20). Thus, we shall construct confidence intervals on some of these values in order to have some feel for the amount of sampling error to be expected.

The procedure for constructing confidence intervals for quantiles is explained in Conover's notes (1965b). For example, the 95 per cent confidence interval for the 90th quantile for sample size 20 can be written in a probability statement as

$$P[2.7527 \leqq x_{.90} \leqq 2.9329] = .95,$$

and for a sample size of 30 it is

$$P[2.6333 \leqq x_{.90} \leqq 2.8479] = .95,$$

where $x_{.90}$ represents the 90th quantile. The 95 per cent confidence interval

TABLE 1

Raw values for the approximate critical values of the test statistic $-2 \log \lambda_1$ computed from 10,000 simulations assuming the null hypothesis to be true

Sample Size	α-Level			
	.10	.05	.025	.01
2	5.5840	7.9238	10.5834	13.7958
4	3.6086	5.1035	6.8053	9.0439
6	3.2181	4.5613	6.1457	8.2337
8	3.1655	4.4783	5.9372	7.9485
10	3.0233	4.2754	5.5516	7.1822
14	2.9674	4.1498	5.4429	7.2077
20	2.8519	3.9535	5.2697	7.0768
30	2.7357	3.9614	5.0946	6.4878
40	2.7377	3.8036	5.0252	6.4295
∞	2.7055	3.8415	5.0239	6.6349

TABLE 2

Smoothed approximate critical values for the test statistic $-2 \log \lambda_1$

Sample Size	α-Level			
	.10	.05	.025	.01
2	5.58	7.9	10.6	13.8
4	3.60	5.1	6.8	9.1
6	3.21	4.6	6.2	8.2
8	3.17	4.5	6.0	7.9
10	3.02	4.3	5.6	7.4
14	2.96	4.2	5.4	7.2
20	2.85	4.0	5.2	7.0
30	2.73	3.9	5.1	6.7
40	2.71	3.85	5.03	6.64
∞	2.70	3.841	5.024	6.635

on the 99th quantile for a sample size of 20 can be written in a probability statement as

$$P[6.7353 \leqq x_{.99} \leqq 7.4599] = .95;$$

and for a sample size of 30 we have

$$P[6.2830 \leqq x_{.99} \leqq 6.9312] = .95.$$

Thus, we can see that there could be enough sampling error to cause a loss of monotonicity in some of the α-levels. With this in mind, an intuitive smoothing of Table 1 will give a more usable table.

Thus, we have looked at the distribution of $-2 \log \lambda_1$ under the null hypothesis. We shall now look at the power of the test statistic $-2 \log \lambda_1$ in Table 3.

5. Power studies of the likelihood ratio test statistic. In Table 3 we shall let $\lambda_1^* = -2 \log \lambda_1$. Bartlett's test is the well-known test for homogeneity of variances (Bartlett, 1937). Con. Slip. is Conover's distribution-free slippage test, which uses as a test statistic the number of observations in the sample with the largest extreme value that exceeds the largest value in the other sample (Conover, 1968).

Conover's slippage test was used in the comparisons of power because it should have good power if the coefficients of variation are homogeneous and also because it does not assume homogeneous variances. The standard two-sample t-test is represented by t, and t' is the unequal variance t-test (Fryer, 1966). The coefficient of variation for population i is cv_i.

In Table 3, all of the sections are computed from 100 random samples; therefore, these can be used only as indicators of what the power may be for the stated populations.

Sections 1 through 5 have homogeneous coefficients of variation. In section 1, with a .25 coefficient of variation, we see that there is little difference in the power of the tests λ_1^*, t, and t', but as the coefficient of variation increases to .50 in section 2, the test λ_1^* appears to be the most powerful. As the coefficient of variation increases, the power of λ_1^* also increases; also, it can be observed that the power of Bartlett's test is increasing. In section 5, where the coefficients of variation are 2, we can see that Bartlett's test and λ_1^* have approximately the same power.

Sections 1 through 5 seem to indicate that if the coefficients of variation are homogeneous and between the values of .15 and 1.0, the most powerful test of mean differences is λ_1^*. For coefficients of variation larger than 1.0 we run into some danger of rejecting the null hypothesis $H_0(\mu_1 = \mu_2)$ due to variance differences, when indeed $H_0(\mu_1 = \mu_2)$ is true as in section 6. For these cases of coefficients of variation greater than 1, we must be certain of our assumption of homogeneous coefficients of variation, because if the assumption is true, section 5 indicates that λ_1^* is the most powerful test for detecting mean differences.

Section 6 emphasizes the danger when the coefficients of variation are greater than one, by rejecting far too many samples when the null hypothesis is true, although the assumption of homogeneous coefficients of variation is violated. Section 6 supports Box's (1954) position on the robustness of the t-test.

Section 7 indicates that, with two samples with small heterogeneous coefficients of variation and homogeneous standard deviations, there is little difference in the powers of λ_1^*, t, and t'.

In section 8 we have $\mu_1 < \mu_2$ with $\sigma_1 > \sigma_2$. This violates the assumptions of λ_1^* and causes an extreme loss in the power of λ_1^*.

TABLE 3

Total number of rejection out of 100 random samples simulated with the given population parameters for 2 samples of size 10

	1			2			3		
	$\mu_1 = 8.0$, $\sigma_1 = 2.0$, $cv_1 = .25$		$\mu_2 = 10.0$, $\sigma_2 = 2.5$, $cv_2 = .25$	$\mu_1 = 2.0$, $\sigma_1 = 1.0$, $cv_1 = .50$		$\mu_2 = 6.0$, $\sigma_2 = 3.0$, $cv_2 = .50$	$\mu_1 = 2.67$, $\sigma_1 = 2.00$, $cv_1 = .75$		$\mu_2 = 4.67$, $\sigma_2 = 3.50$, $cv_2 = .75$
Test statistic	α-Level			α-Level			α-Level		
	.01	.05	.10	.01	.05	.10	.01	.05	.10
Bartlett's	4	15	22	6	13	25	12	35	47
λ_1*	24	43	57	28	49	63	39	70	82
t	22	48	55	16	36	46	22	41	57
t'	26	49	58	20	38	50	25	46	59
Con. slip.	13	35	44	9	31	48	10	43	64

	4			5			6		
	$\mu_1 = 2.0$, $\sigma_1 = 2.0$, $cv_1 = 1.0$		$\mu_2 = 4.0$, $\sigma_2 = 4.0$, $cv_2 = 1.0$	$\mu_1 = 1.0$, $\sigma_1 = 2.0$, $cv_1 = 2.0$		$\mu_2 = 3.0$, $\sigma_2 = 6.0$, $cv_2 = 2.0$	$\mu_1 = 1.0$, $\sigma_1 = 2.0$, $cv_1 = 2.0$		$\mu_2 = 1.0$, $\sigma_2 = 4.0$, $cv_2 = 4.0$
Test statistic	α-Level			α-Level			α-Level		
	.01	.05	.10	.01	.05	.10	.01	.05	.10
Barlettt's	20	47	65	71	83	89	31	56	66
λ_1*	42	72	83	78	91	93	26	46	64
t	10	31	42	3	14	29	1	5	10
t'	11	33	45	6	16	34	1	6	13
Con. slip.	6	35	61	8	41	64	0	12	26

	7			8			9		
	$\mu_1 = 8.0$, $\sigma_1 = 2.0$, $cv_1 = .25$		$\mu_2 = 10.0$, $\sigma_2 = 2.0$, $cv_2 = .20$	$\mu_1 = 1.0$, $\sigma_1 = 2.0$, $cv_1 = 2.0$		$\mu_2 = 2.0$, $\sigma_2 = 1.0$, $cv_2 = .5$	$\mu_1 = 1.0$, $\sigma_1 = 2.0$, $cv_1 = 2.0$		$\mu_2 = 3.0$, $\sigma_2 = 2.5$, $cv_2 = .833$
Test statistic	α-Level			α-Level			α-Level		
	.01	.05	.10	.01	.05	.10	.01	.05	.10
Bartlett's	2	4	13	24	53	63	3	9	20
λ_1*	24	49	70	0	5	7	8	29	41
t	29	57	73	16	32	37	9	39	55
t'	35	61	75	21	34	40	6	45	58
Con. slip.	9	28	46	4	7	12	3	17	38

	10			11		
	$\mu_1 = 4.0$, $\sigma_1 = 2.0$, $cv_1 = .5$		$\mu_2 = 6.0$, $\sigma_2 = 2.5$, $cv_2 = .416$	$\mu_1 = 1.0$, $\sigma_1 = 2.0$, $cv_1 = 2.0$		$\mu_2 = 3.0$, $\sigma_2 = 4.0$, $cv_2 = 1.33$
Test statistic	α-Level			α-Level		
	.01	.05	.10	.01	.05	.10
Bartlett's	2	15	24	24	46	60
λ_1*	27	60	67	39	58	71
t	2	48	60	9	24	36
t'	27	53	67	11	26	36
Con. slip.	8	34	54	6	36	51

Section 9 indicates that, for small differences in standard deviation and larger differences in coefficients of variation with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$, the t-test and the t'-test are the most powerful.

In section 10 we see little difference in the procedures that test mean differences. If we let $\mu_2 = \mu_1 + 1$ and $\sigma_2 = \sigma_1 + 2$, where $\sigma_2^2/\sigma_1^2 = 4.0$ as in section 11, λ_1^*

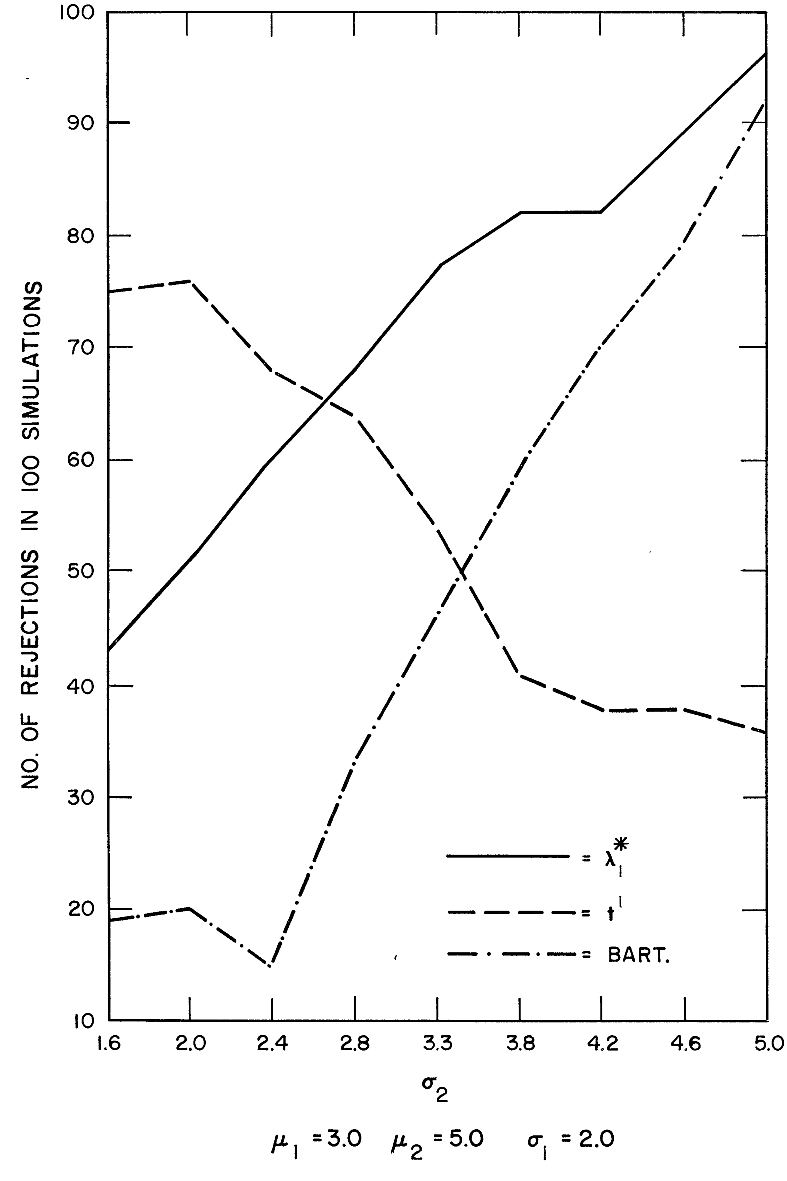

FIG. 2

is the most powerful test for this simulation with this particular choice of parameters.

In Figure 2 the values are fixed for μ_1, μ_2, and σ_1, with σ_2 varying with an α-level of .10. At the point $\sigma_2 = 2.0$, t' is the most powerful test. As σ_2 increases to 3.3, the assumptions of λ_1^* are satisfied; we see that λ_1^* is the most powerful at this point and remains best as σ_2 increases for the simulations presented in Figure 2.

6. Conclusions and acknowledgments. If an experimenter feels that the assumption of homogeneous coefficients of variation is a more valid assumption than the assumption of homogeneous variances when testing mean differences of two populations, then the λ_1^* test will generally be more powerful than the t-test. This is especially true if the coefficient of variation is between .25 and 1.0.

I am very much indebted to my major professor, Dr. W. J. Conover, for his excellent advice and suggestions in the preparation of this paper. I also wish to acknowledge Dr. R. N. Carr, Mr. Walter Knowles, Mr. Richard Beckman, and Dr. R. K. Zeigler for their helpful suggestions.

REFERENCES

[1] BARTLETT, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc. Ser. A* **160** 268–282.

[2] BEHRENS, W. V. (1929). Ein Beitrag zur Fehlerberechnung bei weinige Beobachtungen. *Landiv. Jb.* **68** 807–837.

[3] BOX, G. E. P. (1954a). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effects of inequality of variance in one-way classifications. *Ann. Math. Statist.* **25** 290–302.

[4] BOX, G. E. P. (1954b). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variances and of correlation between errors in the two-way classification. *Ann. Math. Statist.* **25** 484–498.

[5] BOX, G. E. P. and MULLER, M. E. (1958). A note on the generation of random normal deviates. *Ann. Math. Statist.* **29** 610–611.

[6] BOZIVICH, H., BANCROFT, T. A., and HARTLEY, H. O. (1956). Power of analysis of variance test procedures for certain incompletely specified models, I. *Ann. Math. Statist.* **27** 1017–1043.

[7] COCHRAN, W. G. and COX, G. M. (1957). *Experimental Designs.* Wiley, New York.

[8] CONOVER, W. J. (1965a). Several k-sample Kolmogorov-Smirnov tests. *Ann. Math. Statist.* **36** 1019–1026.

[9] CONOVER, W. J. (1965b). Notes from a Course in Nonparametric Statistics. Kansas State University.

[10] CONOVER, W. J. (1968). Two k-sample slippage tests. *J. Amer. Statist. Assoc.* **63** 614–626.

[11] DAVID, F. N. and JOHNSON, N. L. (1951a). A method of investigating the effect of non-normality and heterogeneity of variance on tests of the general linear hypothesis. *Ann. Math. Statist.* **22** 382–392.

[12] DAVID, F. N. and JOHNSON, N. L. (1951b). The effect of non-normality on the power function of the F-test in the analysis of variance. *Biometrika* **38** 43–57.

[13] FISHER, R. A. (1939). The comparison of samples with possibly unequal variances. *Ann. Eugenics* **9** 174–180.

[14] FISHER, R. A. and YATES, F. (1948). *Statistical Tables.* Oliver and Boyd, London.

[15] FRYER, H. C. (1966). *Concepts and Methods of Experimental Statistics.* Allyn and Bacon, Boston.

[16] Graybill, F. A. (1961). *An Introduction to Linear Statistical Models.* McGraw-Hill, New York.

[17] Gronow, D. G. C. (1951). Test for the significance of the difference between means in two normal populations having unequal variances. *Biometrika* **38** 252–256.

[18] Hohn, F. E. (1965). *Elementary Matrix Algebra.* MacMillan, New York.

[19] Kendall, M. G. and Stuart, A. (1961). *The Advanced Theory of Statistics Vol. 2.* Griffin, London.

[20] Lohrding, R. K. (1969). Likelihood ratio tests of equal mean when the variances are heterogeneous. Ph.D. dissertation. Kansas State University, Manhattan, Kansas.

[21] Lohrding, R. K. (1966). Nonparametric analogues of analysis of variance. Masters Report. Kansas State University, Manhattan, Kansas.

[22] Peterson, T. S. 1(947). *College Algebra.* Harper, New York.

[23] Welch, B. L. (1937). The significance of the difference between two means when the two population variances are unequal. *Biometrika* **29** 350–362.

[24] Wilks, S. S. (1963). *Mathematical Statistics.* Wiley, New York.