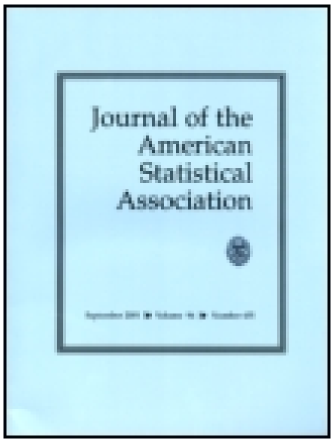

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

A Test of Equality of Two Normal Population Means and Variances

S. K. Perng [a] & Ramon C. Littell [b]

[a] Department of Statistics , Kansas State University , Manhattan , KA , 66506 , USA

[b] Department of Statistics , University of Florida , Gainesville , FL , 32611 , USA

Published online: 05 Apr 2012.

A Test of Equality of Two Normal Population Means and Variances

S. K. PERNG and RAMON C. LITTELL*

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two independent random samples from normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. The statistical problem is to test H_0: $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$ against H_a: $\mu_1 \neq \mu_2$ or $\sigma_1^2 \neq \sigma_2^2$. Using Fisher's method of combining two independent test statistics, we suggest a test and prove that it is asymptotically optimal in the sense of Bahadur efficiency.

1. INTRODUCTION

Let two independent samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be taken from normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Define $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ and $\Omega = \{\theta: -\infty < \mu_i < +\infty, 0 < \sigma_i^2 < +\infty, i = 1, 2\}$. The classical Behrens-Fisher problem is to test the null hypothesis that the means, but not necessarily the variances, are equal. There are cases of practical interest, however, when it is appropriate to test for equality of variances at the same time one is testing for equality of means. For instance, it has been recognized [11, p. 324] that the application of different treatments to otherwise homogeneous experimental units often results in the treatment groups differing not only in means but also in variances. Thus the inequality of means may go along with inequality of variances, so that a null hypothesis of no treatment differences entails $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$. We therefore wish to consider tests of H_0: $\theta \in \Omega_0 = \{\theta: \mu_1 = \mu_2, \sigma_1^2 = \sigma_2^2\}$ against H_a: $\theta \in \Omega - \Omega_0$. In this connection, Fisher's clarifying comments [5, p. 125] on the "assumptions" of the t test are of note, in which he points out that the t test is a valid test of H_0 against H_a. This is because the conditions for the null distribution of t are specified in H_0, and thus t provides valid significance probabilities.

The problem of testing H_0 against H_a received attention from other early writers. Neyman and Pearson [10] derived the likelihood ratio statistic for the problem and obtained its moments and limiting distribution. Sukhatme [12] derived the distribution of the likelihood ratio statistic in terms of incomplete beta functions. He also proposed an alternative test of H_0 against H_a which employed Fisher's method of combining independent tests. It was demonstrated graphically that this alternative test has critical regions which closely resemble those of the likelihood ratio test.

In Section 2 of this paper, we suggest a test of H_0: $\theta \in \Omega_0$ against H_a: $\theta \in \Omega - \Omega_0$ which is essentially the same as the alternative test proposed by Sukhatme. In Section 3, we prove that the suggested test is asymptotically optimal in the sense of Bahadur efficiency [1].

2. THE TEST

First we define the following notation:

$$T_{mn} = (\bar{Y}_n - \bar{X}_m)\left[\frac{(m+n)(S_1^2 + S_2^2)}{(m+n-2)m\cdot n}\right]^{-\frac{1}{2}}, \tag{2.1}$$

$$F_{mn} = [S_2^2(m-1)][S_1^2(n-1)]^{-1}, \tag{2.2}$$

$$H_{mn} = \begin{cases} 2[1 - G_{mn}(F_{mn})], & \text{if } F_{mn} \geq \operatorname{med} G_{mn} \\ 2G_{mn}(F_{mn}), & \text{if } F_{mn} < \operatorname{med} G_{mn}, \end{cases} \tag{2.3}$$

$$W_{mn} = -2 \log H_{mn}, \tag{2.4}$$

where

$$S_1^2 = \sum_{i=1}^{m} (X_i - \bar{X}_m)^2, \quad \bar{X}_m = m^{-1} \sum_{i=1}^{m} X_i,$$

$$S_2^2 = \sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2, \quad \bar{Y}_n = n^{-1} \sum_{i=1}^{n} Y_i,$$

G_{mn} is the distribution function of central F with $n-1$ and $m-1$ degrees of freedom, and $\operatorname{med} G_{mn}$ is the median of central F with $n-1$ and $m-1$ degrees of freedom. We note that under H_0 (i.e., $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$), T_{mn} has the central t distribution with $(m+n-2)$ degrees of freedom and F_{mn} has the central F distribution with $n-1$ and $m-1$ degrees of freedom. Before we define the test, we prove two lemmas.

Lemma 2.1: The statistics T_{mn} and W_{mn} are independent if $\sigma_1^2 = \sigma_2^2$.

Proof: We know that $(\bar{Y}_n - \bar{X}_m)$, $(S_1^2 + S_2^2)$, and S_2^2/S_1^2 are independent. (See [7, p. 163]). Since T_{mn} is a function of $(\bar{Y}_n - \bar{X}_m)$ and $(S_1^2 + S_2^2)$, and W_{mn} is a function of S_2^2/S_1^2, T_{mn}, and W_{mn} are stochastically independent.

Lemma 2.2: Under the null hypothesis (i.e., $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$), W_{mn} has the central chi-squared distribution with two degrees of freedom.

Proof: First we note that the distribution of $G_{mn}(F_{mn})$, conditional on $F_{mn} \geq \operatorname{med} G_{mn}$, is uniform on $[\frac{1}{2}, 1]$,

* S.K. Perng is Associate Professor, Department of Statistics, Kansas State University, Manhattan, KA 66506. Ramon C. Littell is Associate Professor, Department of Statistics, University of Florida, Gainesville, FL 32611.

© Journal of the American Statistical Association
December 1976, Volume 71, Number 356
Theory and Methods Section

and the distribution of $G_{mn}(F_{mn})$, conditional on F_{mn} $<$med G_{mn}, is uniform on $[0, \frac{1}{2}]$. Thus H_{mn} is distributed uniformly on $[0, 1]$, and hence W_{mn} has the central chi-squared distribution with two degrees of freedom.

Note that the observed values of H_{mn} are simply the significance levels or tail probabilities of the usual two-sided F test with equal tail probabilities for testing the equality of two normal population variances. Since W_{mn} is a monotone function of H_{mn}, the test based on W_{mn}, with large values of W_{mn} being significant, is equivalent to the usual two-sided F test. Now T_{mn} is usually used to test the equality of two normal population means, the two-sided F test is usually used to test the equality of two normal variances, and T_{mn} and W_{mn} are stochastically independent under H_0. Thus it seems reasonable to combine the two tests based on T_{mn} and W_{mn} by Fisher's method [5, pp. 99–101] to form a new test for the problem defined in Section 1 as follows:

$$\begin{aligned} &\text{Reject } H_0 \text{ if } Q_{mn} \geq c_0 \ ; \\ &\text{Accept } H_0 \text{ if } Q_{mn} < c_0 \ , \end{aligned} \tag{2.5}$$

where

$$Q_{mn} = -2 \log P_{H_0}[|T_{mn}| > t] - 2 \log P_{H_0}[W_{mn} > w] \ ,$$

t equals the observed value of $|T_{mn}|$, and w is the observed value of W_{mn}. It is well-known that under H_0, Q_{mn} has the central chi-squared distribution with four degrees of freedom. Hence, the critical value c_0 can be easily determined by using chi-squared tables.

We note that the Q test (2.5) can also be put in the following form:

$$\begin{aligned} &\text{Reject } H_0 \text{ if } Q_{mn} \geq c_0 \ , \\ &\text{Accept } H_0 \text{ if } Q_{mn} < c_0 \ , \end{aligned} \tag{2.6}$$

where

$$Q_{mn} = -2 \log P_{H_0}[|T_{mn}| > t] - 2 \log P_{H_0}[H_{mn} < h] \ ,$$

t is the observed value of $|T_{mn}|$, and h is the observed value of H_{mn}.

3. AN ASYMPTOTIC OPTIMALITY PROPERTY OF THE TEST

In this section we shall show that the test suggested in Section 2 is asymptotically optimal in the sense of Bahadur efficiency. We first introduce Bahadur efficiency [1, 2] in the two-sample context. For each m and n, let $U_{mn}^{(1)}$ and $U_{mn}^{(2)}$ be two competing test statistics for testing $H_0\colon \theta \in \Omega_0$ against $H_1\colon \theta \in \Omega - \Omega_0$. We assume that large values of $U_{mn}^{(1)}$ (or $U_{mn}^{(2)}$) are significant. For any t, let

$$F_{mn}^{(i)}(t) = \inf\,(P_\theta[U_{mn}^{(i)} < t] | \theta \in \Omega_0)$$

for $i = 1, 2$. Define the significance level attained by $U_{mn}^{(i)}$ as $L_{mn}^{(i)} = 1 - F_{mn}^{(i)}(U_{mn}^{(i)})$ for $i = 1, 2$. Furthermore, assume that m and n increase in such a way that $m(m+n)^{-1}$ tends to γ and $0 < \gamma < 1$. In typical cases, for $i = 1, 2$,

$$\lim_{N \to \infty} (-2/N)[\log L_{mn}^{(i)}] = c_i(\theta)$$

with probability one when $\theta \in \Omega - \Omega_0$, where $N = n + m$, and $c_i(\theta)$ is a positive function defined over $(\Omega - \Omega_0)$. The function $c_i(\theta)$ is called the exact slope of the sequence $\{U_{mn}^{(i)}\}$. The exact Bahadur efficiency of $\{U_{mn}^{(1)}\}$ relative to $\{U_{mn}^{(2)}\}$ at θ is given by $c_1(\theta)/c_2(\theta)$. The sequence of test statistics $\{U_{mn}^{(1)}\}$ is said to be more efficient than the sequence of test statistics $\{U_{mn}^{(2)}\}$ at θ if

$$c_1(\theta)/c_2(\theta) > 1 \ .$$

To find the exact slope of the sequence of test statistics $\{Q_{mn}\}$, we need to find the exact slopes of the sequences of test statistics $\{|T_{mn}|\}$ and $\{W_{mn}\}$. Throughout this paper we shall assume that $m(n+m)^{-1}$ tends to γ and $0 < \gamma < 1$ as $N \to \infty$. The following lemma provides the exact slopes of $\{|T_{mn}|\}$ and $\{W_{mn}\}$.

Lemma 3.1: The exact slopes of $\{|T_{mn}|\}$ and $\{W_{mn}\}$ are

$$c_1(\theta) = \log\left\{1 + \frac{\gamma(1-\gamma)(\mu_2 - \mu_1)^2}{\gamma\sigma_1^2 + (1-\gamma)\sigma_2^2}\right\} \tag{3.1}$$

and

$$c_2(\theta) = \log\left\{\frac{\gamma\sigma_1^2 + (1-\gamma)\sigma_2^2}{\sigma_1^{2\gamma}\sigma_2^{2(1-\gamma)}}\right\}, \tag{3.2}$$

respectively, for $\theta \in (\Omega - \Omega_0)$.

Proof: The exact slope $c_1(\theta)$ of $(|T_{mn}|)$ will be obtained via Theorem 7.2 of [2]. Observe that

$$\begin{aligned} f(t) &= \lim_{N \to \infty} (-1/N) \log P_{H_0}[|T_{mn}| > N^{\frac{1}{2}}t] \\ &= \lim_{N \to \infty} (-1/N) \log 2P_{H_0}[T_{mn} > N^{\frac{1}{2}}t] \\ &= \tfrac{1}{2} \log (1 + t^2) \end{aligned}$$

where $N = n + m$. The last equality follows from [2, Eq. (8.2)]. For $\theta \in \Omega - \Omega_0$

$$b(\theta) = \lim_{N \to \infty} \frac{|T_{mn}|}{N^{\frac{1}{2}}} = \frac{(\gamma(1-\gamma))^{\frac{1}{2}} |\mu_2 - \mu_1|}{(\gamma\sigma_1^2 + (1-\gamma)\sigma_2^2)^{\frac{1}{2}}}$$

with probability one [θ], which yields

$$c_1(\theta) = \log\left[1 + \frac{\gamma(1-\gamma)(\mu_2 - \mu_1)^2}{\gamma\sigma_1^2 + (1-\gamma)\sigma_2^2}\right]$$

for every $\theta \in \Omega - \Omega_0$.

The exact slope of $\{W_{mn}\}$ is computed as follows. Clearly for every $\theta \in \Omega - \Omega_0$,

$$c_2(\theta) = \lim_{N \to \infty} (-2/N) \log H_{mn}$$

with probability one [θ]. Now if $\sigma_2^2 > \sigma_1^2$, then with probability one, $H_{mn} = 2[1 - G_{mn}(F_{mn})]$ for sufficiently large N. From Killeen, Hettmansperger, and Sievers [6], for $t > 1$,

$$\lim_{N\to\infty} (1/N) \log [1 - G_{mn}(t)] = \lim_{N\to\infty} (m/N)(1/m) \log [1 - G_{mn}(t)] = \tfrac{1}{2} \log (t^{(1-\gamma)}/(\gamma + (1-\gamma)t)) \quad (3.3)$$

and F_{mn} tends to σ_2^2/σ_1^2 with probability one. Thus if $\sigma_2^2 > \sigma_1^2$, we have for every $\theta \in \Omega - \Omega_0$

$$c_2(\theta) = \lim_{N\to\infty} (-2/N) \log [1 - G_{mn}(F_{mn})] = \log \left(\frac{\gamma\sigma_1^2 + (1-\gamma)\sigma_2^2}{\sigma_2^{2(1-\gamma)}\sigma_1^{2\gamma}} \right). \quad (3.4)$$

Similarly, for $\sigma_1^2 > \sigma_2^2$

$$c_2(\theta) = \log \left(\frac{\gamma\sigma_1^2 + (1-\gamma)\sigma_2^2}{\sigma_2^{2(1-\gamma)}\sigma_1^{2\gamma}} \right).$$

The following lemma gives the exact slope of the sequence of test statistics $\{Q_{mn}\}$.

Lemma 3.2: The exact slope of $\{Q_{mn}\}$ is

$$c_Q(\theta) = \log \left[\frac{\gamma\sigma_1^2 + (1-\gamma)\sigma_2^2 + \gamma(1-\gamma)(\mu_2 - \mu_1)^2}{\sigma_1^{2\gamma}\sigma_2^{2(1-\gamma)}} \right] \quad (3.5)$$

for every $\theta \in \Omega - \Omega_0$.

Proof: Equation (3.5) follows from results of Littell and Folks [8] which show that $c_Q(\theta) = c_1(\theta) + c_2(\theta)$ for every $\theta \in \Omega - \Omega_0$.

We shall now show that the Q test is asymptotically optimal, using Corollary 3 of Bahadur and Raghavachari [3]. Let

$$X_m^* = (X_1, \ldots, X_m) \ ,$$
$$Y_n^* = (Y_1, \ldots, Y_n) \ ,$$
$$\theta_1 = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \in \Omega - \Omega_0 \ ,$$
$$\theta_0 = (\mu, \mu, \sigma^2, \sigma^2) \in \Omega_0 \ ;$$

and let $f(X_m^*, Y_n^*, \theta_i)$ for $i = 0, 1$, be the joint density function of X_m^* and Y_n^* when $\theta = \theta_i$. Define

$$r_N(X_m^*, Y_n^*, \theta_1, \theta_0) = \frac{f(X_m^*, Y_n^*, \theta_1)}{f(X_m^*, Y_n^*, \theta_0)}.$$

For given $\theta_1 \in \Omega - \Omega_0$ and $\theta_0 \in \Omega_0$, let the constant $K^{(1)}(\theta_1, \theta_0)$ be given by

$$K^{(1)}(\theta_1, \theta_0) = \lim_{N\to\infty} N^{-1} \log r_N(X_m^*, Y_n^*, \theta_1, \theta_0)$$
$$= \log [\sigma/(\sigma_1^{\gamma}\sigma_2^{(1-\gamma)})] - \tfrac{1}{2}\{1 - \sigma^{-2}[\gamma\sigma_1^2 + \gamma(\mu_1 - \mu)^2 + (1-\gamma)\sigma_2^2 + (1-\gamma)(\mu_2 - \mu)^2]\} \ .$$

(The limit holds with probability one $[\theta_1]$.) Define

$$J^{(1)}(\theta_1) = \inf \{K^{(1)}(\theta_1, \theta_0) : \theta_0 \in \Omega_0\} \ .$$

Then,

$$J^{(1)}(\theta_1) = \tfrac{1}{2} \log \left[\frac{\gamma(1-\gamma)(\mu_1 - \mu_2)^2 + \gamma\sigma_1^2 + (1-\gamma)\sigma_2^2}{\sigma_1^{2\gamma}\sigma_2^{2(1-\gamma)}} \right].$$

Hence,

$$c_Q(\theta_1) = 2J^{(1)}(\theta_1) \ .$$

Now Corollary 3 of Bahadur and Raghavachari [3] states that a test is asymptotically optimal if it has slope equal to $2J^{(1)}(\theta_1)$. We thus conclude that the Q test suggested in Section 2 is asymptotically optimal in the sense of Bahadur efficiency. Now we state the result just obtained as the main theorem of this paper.

Theorem: For the hypothesis testing problem given in Section 1, the Q test defined in Section 2 is asymptotically optimal in the sense of Bahadur efficiency.

It should be pointed out that the theorem can also be proved using the Bahadur-Raghavachari result in conjunction with the fact, shown by Neyman and Pearson [10], that the likelihood ratio for testing H_0 against H_a can be expressed as the product of the likelihood ratio for testing H_0 against H': $\theta \in \Omega'$ and the likelihood ratio for testing H': $\theta \in \Omega'$ against H_a, where $\Omega' = \{\theta: \sigma_1^2 = \sigma_2^2\}$. This approach would not, however, have provided the exact slopes in their explicit form as functions of θ. Having the functional form allows one to compute relative efficiencies. The relative efficiency of the combined test to the t test is $(c_1(\theta) + c_2(\theta))/c_1(\theta)$, which, for example, has the value 1.35 for $\sigma_2^2 = 3\sigma_1^2$ and $|\mu_2 - \mu_1| = 2\sigma_1$. The functional form of

$$c_1(\theta) = \log [1 + \gamma(1-\gamma)(\mu_2 - \mu_1)^2/(\gamma\sigma_1^2 + (1-\gamma)\sigma_2^2)]$$

also reveals the effect of unequal variances on the efficiency of the t test. In particular, the fact that $c_1(\theta) = 0$ for $\mu_2 = \mu_1$ supports Fisher's statement [5, p. 124] that "The theoretical possibility, that a significant value of t should be produced by a difference between the variances only, seems to be unimportant in the application of the method . . . ."

The theorem bears out, in terms of asymptotic efficiency, the graphical observation of Sukhatme [12] that the Q test and the likelihood ratio test behave similarly. On the grounds that the Q test requires no special tables, is readily computed, and has optimal asymptotic efficiency, it appears to be a good test of the hypothesis that two samples came from identical normal populations.

[*Received March 1974. Revised February 1976.*]

REFERENCES

[1] Bahadur, R.R., "Rates of Convergence of Estimates and Test Statistics," *Annals of Mathematical Statistics*, 38 (April 1967), 303–24.

[2] ———, *Some Limit Theorems in Statistics*, Philadelphia: SIAM, 1971.

[3] ——— and Raghavachari, M., "Some Asymptotic Properties of Likelihood Ratios on General Sample Spaces," *Proceedings*

of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, Berkeley and Los Angeles: University of California Press, 1972, 129–52.

[4] Basu, D., "On Statistics Independent of a Complete Sufficient Statistics," *Sankhya,* 15 (April 1955), 377–80.

[5] Fisher, R.A., *Statistical Methods for Research Workers,* 11th ed., London: Oliver & Boyd, Ltd., 1950.

[6] Killeen, T.J., Hettmansperger, T.P. and Sievers, G.L., "An Elementary Theorem on the Probability of Large Deviations," *Annals of Mathematical Statistics,* 43 (February 1972), 181–92.

[7] Lehmann, E.L., *Testing Statistical Hypotheses,* New York: John Wiley & Sons, Inc., 1959.

[8] Littell, R.C. and Folks, J.L., "Asymptotic Optimality of Fisher's Method of Combining Independent Tests," *Journal of the American Statistical Association,* 66 (December 1971), 802–6.

[9] ——— and Folks, J.L., "Asymptotic Optimality of Fisher's Method of Combining Independent Tests II," *Journal of the American Statistical Association,* 68 (March 1973), 193–4.

[10] Neyman, J. and Pearson, E.S., "On the Problem of Two Samples," *Bulletin de l'Academie Polonaise des Sciences et des Lettres,* A (1930), 73–96.

[11] Snedecor, G.W. and Cochran, W.G., *Statistical Methods,* 6th ed., Ames: Iowa State University Press, 1967.

[12] Sukhatme, P.V., "A Contribution to the Problem of Two Samples," *Proceedings India Academy Science,* Series A, Vol. 2, (December 1935), 384–604.