Power of Tests for Equality of Covariance Matrices*

Richard L. Greenstreet  
V. P. l.  
Blacksburg, Virginia  

and  

Robert J. Connor  
National Cancer Institute  
Bethesda, Maryland

In this paper a Monte Carlo study of four test statistics used to test the null hypothesis that two or more multivariate normal populations have equal covariance matrices is presented. The statistics \(-2 \ln \lambda, -2 \ln W, -2p_1 \ln \lambda\) and \(-2p_2 \ln W\), where \(\lambda\) is the likelihood ratio criterion and \(W\) is Bartlett’s modification of \(\lambda\), are investigated. The results indicate that for small samples the actual significance levels of the first two statistics are somewhat greater than the nominal significance levels set, whereas for the latter two statistics the actual significance levels are close to the nominal levels. For these test statistics, \(-2p_1 \ln \lambda\) and \(-2p_2 \ln W\), the power is seen to be essentially identical; further it increases as the sample size increases, as the inequality of the covariance matrices increases, and as the number of variates increases. Also, their power is seen to be a concave function of the number of populations.

**KEY WORDS**
- Covariance Matrices
- Power of Tests
- Likelihood Ratio Tests
- Multivariate Normal

1. INTRODUCTION AND PROBLEM

Testing for equality of covariance matrices is a topic of much interest and a number of authors have studied this subject. Their attention has been focused primarily on properties of bias, invariance, and monotonicity (e.g. Anderson and Das Gupta (1964), Giri (1963), Nagao (1967), and Sugiuira and Nagao (1968)). Two exceptions are Korin (1967) who, as part of a larger study of multivariate tests, gives limited power results and Ito (1968) who makes an asymptotic study of the effects of non-normality on the test. This paper presents the results of a Monte Carlo study on four test statistics used to test the equality of covariance matrices of multivariate normal populations. The study was an attempt to determine how well the critical values based on the asymptotic distributions of the test statistics fit in small samples, and then for the test statistics for which this fit was good to estimate the power in small samples.

The four statistics considered are based on the likelihood ratio criterion, \(\lambda\). They are \(-2 \ln \lambda\), \(-2 \ln W\), \(-2p_1 \ln \lambda\), and \(-2p_2 \ln W\) where \(W\) is Bartlett’s modification of \(\lambda\) and \(p_1\) and \(p_2\) are constants (see equations (2) and (3)) which improve the approximation to the asymptotic distribution. More specifically, if there are \(q\) normal populations each \(p\)-variate (with unknown mean vectors) the null hypothesis of interest is

\[ H_0 : \Sigma_1 = \cdots = \Sigma_q \]

where \(\Sigma_i\) is the \(p \times p\) covariance matrix of the \(i^{th}\) population. For this null hypothesis the likelihood ratio criterion is

\[ \lambda = \prod_{i=1}^{q} |A_i|^{N_i/2} / |A|^{N_T/2} \cdot \left( N_T / \prod_{i=1}^{q} N_i \right)^{q-1} \]

where \(N_i\) is the sample size from the \(i^{th}\) population, \(N_T = \sum_{i=1}^{q} N_i\),

\[ A_i = \sum_{j=1}^{N_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)' \]

and

\[ A = \sum_{i=1}^{q} A_i \]

with \(X_{ij}\), a \(p\)-dimensional vector, being the \(j^{th}\) observation from population \(i\), and \(\bar{X}_i\), the sample mean from the \(i^{th}\) population. Further, \(W\) is given by (1) if each \(N_i\) is replaced by \(n_i = N_i - 1\) and consequently \(N_T\) by \(n = N_T - q\). The multiplicative constants are

\[ p_1 = 1 - \left[ \left( \sum_{i=1}^{q} \frac{1}{N_i} \right) - \frac{1}{N_T} \right] \cdot \frac{2p^2 + 9p + 11}{6(q - 1)(p + 1)} + \frac{q + p + 2}{N_T(p + 1)} \]

and

\[ p_2 = 1 - \left[ \left( \sum_{i=1}^{q} \frac{1}{n_i} \right) - \frac{1}{n} \right] \cdot \frac{2p^2 + 3p - 1}{6(q - 1)(p + 1)} \]

* This work was done while the authors were at the Division of Computer Research and Technology, National Institutes of Health, Department of Health, Education and Welfare, Bethesda, Maryland.

Received July 1969; revised May 1973
The asymptotic distribution is \( \chi^2 \) with degrees of freedom 
\[ f = \frac{(q - l)p(p + 1)}{2} \]
to order \( n^{-1} \) for \( -2 \ln \lambda \) and \( -2 \ln W \) and to order \( n^{-2} \) for \( -2p_1 \ln \lambda \) and \( -2p_2 \ln W \). For a more complete discussion of these statistics see Anderson (1958) or Kendall and Stuart (1966).

In estimating the power for these test statistics with good agreement between the actual and nominal significance levels the alternative hypotheses were of the form

\[ H_A : \Sigma_1 = \cdots = \Sigma_{q-1} \neq \Sigma_q \]

where \( \Sigma_q \) is such that eigenvalues of \( \Sigma_q \Sigma_q^{-1} \) are equal. Note that it is possible to simultaneously transform any two random normal vectors with positive definite covariance matrices, say \( \Sigma_1 \) and \( \Sigma_2 \), so that the transformed vectors are random normal vectors with the covariance matrices being the identity matrix, \( I \), and a diagonal matrix, \( D \), whose elements are the eigenvalues of \( \Sigma_1 \Sigma_2^{-1} \).

Further, under such transformations \( \lambda \) and \( W \) are invariant, and hence the four statistics are invariant. Thus, this paper's problem can be considered in canonical form. That is, it is sufficient to study the statistics for \( \Sigma_i = I \) for \( i = 1, \cdots, q - 1 \) and \( \Sigma_q = D \).

For simplicity the present paper treats only the cases with sample sizes equal and eigenvalues being at least one and equal, that is \( D = dI \) and \( d \geq 1 \). Note that the roots of \( \Sigma_q \Sigma_q^{-1} \) equal to a constant \( d \) implies \( \Sigma_i = d \Sigma_2 \).

The power was studied with several values of the following parameters: \( \alpha \), the nominal significance level; \( q \), the number of populations; \( p \), the number of variables; \( d \), the value of the diagonal elements of \( D \) and \( N \), the size of the sample from each population.

In the next section the four test statistics are compared with respect to their significance levels, actual versus nominal, and then numerical estimates of power are given with the parameters ranging from small to moderately large values for the statistics for which the actual and nominal significance levels agreed. These results should be of interest to researchers in a number of fields who are concerned with multivariate analysis.

2. Methods and Results

The power for each combination of the parameters \( q, p, d, \) and \( N \) was estimated as follows. First, 5000 samples were generated, then the sample value for each of the four test statistics was calculated and the proportion greater than the \( \chi^2(f) \) point was noted.

2.1 Generation of Samples

The \( p \)-variate normal samples with identity covariance matrix were obtained by first generating uniformly distributed pseudo-random numbers following the method suggested by MacLaren and Marsaglia (1965). These uniformly distributed numbers were transformed to standardized random normal variates using the method of Box and Muller (1958). Finally, each consecutive set of \( p \) normal variates was taken as one \( p \)-dimensional vector observation. For the samples from the \( q \)th population the same process was followed but the \( p \)-dimensional vectors were multiplied by \( \sqrt{d} \). The calculations, performed on an IBM 360/50, were done for \( \alpha = .001, .010, .025, .050; q = 2, 3, 6; p = 2, 4, 10; N = 10, 20, 100; \) and \( d = 1.0, 1.5, 5.0, 10.0 \).

Results for \( N = 5, p = 2 \) for all values of \( \alpha \) and \( q \) are reported also. Note, that for \( p = 10 \) and \( N = 10 \) no results are possible because in this case the sample covariance matrices are singular.

2.2 Evaluation of Test Statistics

The four statistics behaved in a consistent fashion for the cases studied. The results for \( -2 \ln \lambda \) were similar to those for \( -2 \ln W \) and the results for \( -2p_1 \ln \lambda \) matched those for \( -2p_2 \ln W \).

As \( N \) varied the behavior of the statistics varied considerably. For large \( N, N = 100 \), the four statistics yielded results which were essentially identical in all cases. Of special interest is that for \( N = 100 \) the significance levels observed for the four statistics were close to the nominal levels.

For small \( N, N = 5, 10 \) or 20, there were large differences in the results. In particular, for small \( N \) the pair \( -2 \ln \lambda \) and \( -2 \ln W \) had significance levels consistently higher than expected while the pair \( -2p_1 \ln \lambda \) and \( -2p_2 \ln W \) achieved levels very close to the significance levels that were desired. For example with \( N = 5, p = 2 \) and \( q = 6 \) the observed significance levels for \( \alpha \) set at .05, .025, .01, and .001 were .189, .177, .061 and .012 for \( -2p_1 \ln W \);

with \( N = 20, p = 4 \) and \( q = 6 \) the observed levels were .118, .069, .031, and .006 for \( -2 \ln W \) and \( .047, .021, .009 \) and .0014 for \( -2p_2 \ln W \).

Because the results for small \( N \) indicate that while the critical values for the asymptotic \( \chi^2 \) distribution do not fit well for \( -2 \ln \lambda \) and \( -2 \ln W \) they do fit well for \( -2p_1 \ln \lambda \) and \( -2p_2 \ln W \), the first pair of statistics are not considered further and the power estimates are given only for the latter pair. In particular the results for \( -2p_2 \ln W \) are presented; however, everything said applies equally to \( -2p_1 \ln \lambda \) because all the results are essentially the same for this pair of statistics.

2.3 Power of \( -2p_2 \ln W \)

The power results are presented in Figures 1, 5, 9 for \( \alpha = .050 \); in Figures 2, 6, 10 for \( \alpha = .025 \); in
Figures 3, 7, 11 for $\alpha = .010$; and in Figures 4, 8, 12 for $\alpha = .001$. Also the power for $\alpha = .050$ is given in Table 1. Inspection of the results reveals the following consistent properties of the statistics under question. The power increases with $N$, the common sample size, with $d$, the measure of difference of the covariance matrices (e.g. Figures 1, 2, 3 and 4) and with $p$ (e.g. Figures 1, 5, 9). An unexpected empirical result is that power is seen to be a concave function of $q$ (e.g. Figures 9, 10, 11 and 12). It is not clear why this occurs.

These results give an indication of the test's sensitivity to the departures from equality. For $d = 1.5$ the test is not likely to detect the inequality unless $N = 100$ while for $d = 10.0$ the inequality is quite likely to be found even for $N = 10$. Further the results show that the values of $p$ and $q$ are of much less consequence than the values of $N$ and $d$. For example, at $\alpha = .050$ over all the values of $p$ and $q$ considered when $N = 20$ and $d = 1.5$ the power is between .10 and .20. For $d = 5.0$ and $N = 20$ the power for all $p$ and $q$ is between .98 and 1.00. When $N = 100$ for $d = 1.5$ the power is between .64 and .91 for all values of $p$ and $q$. For $d = 5.0$ the power is 1.00.

For an investigator who considers the two population case, i.e. $q = 2$, with $p = 4$ and is concerned when $\xi_1 = 5\xi_2$, Table 1 shows that a sample of size 10 from each population insures a power of around .728 of rejecting $H_0$.

2.4 Conclusions

The test statistics $-2 \ln \lambda$ and $-2 \ln W$ for small samples are inferior to $-2p_1 \ln \lambda$ and $-2p_2 \ln W$ in that their actual significance levels are con-

TECHNOMETRICS®, VOL. 16, NO. 1, FEBRUARY 1974
siderably greater than the desired nominal values. Also it is seen that \( W \), Bartlett's modification of \( \lambda \), the likelihood ratio criterion, has little effect with respect to power.

For the cases considered it is seen that \(-2p_2 \ln W\) has reasonably good power when there are moderate to large departures from the null hypothesis, that is for \( d \geq 5.0\), even for sample sizes as small as 10. However, for small departures, that is \( d = 1.5\), large samples are required in order to have good power.

3. Acknowledgments

We are grateful to Mr. Ray Danner for designing the program and making the computer runs necessary for this research.

References


