POWER ANALYSIS FOR MULTIVARIATE MODELS: NEW RESULTS AND SAS MAKE IT PRACTICAL
Keith E. Muller and Berendis L. Peterson, University of North Carolina, Chapel Hill

ABSTRACT

General approximations have been known for a decade for computing power in multivariate linear models. Tests with Wilks' Lambda, Hotelling-Lawley Trace and the Pillai-Bartlett Trace (Lee, 1971, Sugiura and Fujikoshi, 1969). All use complicated mixtures of noncentral chi-square variates. Recent results (Muller and Peterson, 1983) based on single noncentral F variates provide far more convenient approximations. The basic ideas needed to implement the new approximations are reviewed. A direct approach, using matrix manipulations, and an indirect approach, using linear models programs, are sketched.

INTRODUCTION

Motivation

Power calculations in testing the multivariate general linear hypothesis have presented difficult problems. Although not well known (despite publication in some of the most widely read statistical journals), practical asymptotic approximations have been available for a decade. More convenient methods have recently been proposed. The goals of this article are 1) to outline the recently published methods and 2) to present practical ways to use them. The focus is on general methods which can be implemented easily in computer programs.

Four topics will be discussed. First, methods for calculation of probabilities in the central case are reviewed. Second, earlier work providing asymptotic approximations to power is reviewed. Third, new and more convenient approximations are presented. Finally, a direct method, using matrix manipulations, and an indirect method, using linear models programs, for power calculations in experimental design are highlighted.

Formal Statement of the Problem

A convenient presentation of the methods requires a clear specification of the hypothesis of interest. The general linear multivariate model is

\[ Y = XB + \epsilon \]

Here \( N \) is the number of observations, \( p \) the number of dependent variables and \( r \) the number of columns of the design matrix, \( X \), which is a set of known constants. It is convenient to assume \( N > p + r \). The matrix of unknown parameters is \( B \). Each row of \( Y \) is assumed i.i.d. multivariate normal with mean zero and covariance matrix \( \Sigma \). With appropriate choices of \( X \) and \( Y \) the model encompasses multivariate multiple regression, fixed effect MANOVA and their myriad special cases. General treatments of the material in this section may be found in Timm (1975), Finn (1974), Morrison (1967), Rao (1973), Kshirsagar (1972), Harris (1975) and Anderson (1974), among others.

The general linear hypothesis (GLH) may be stated as

\[ H_0: C B U - \theta = \delta = 0 \]

(2)

For testing purposes, \( r(C) = \alpha r \) and \( r(U) = bsp \) will be required. The matrix of secondary parameters, \( \delta \), contains linear combinations of the matrix of primary parameters, \( B \). The matrix \( C \), the usual contrast matrix in univariate linear models, computes linear combinations of rows of \( B \) which correspond to different predictors in the design matrix. The matrix \( U \) computes linear combinations of columns of \( B \), which correspond to different dependent variables. It is useful and accurate to think of \( U \) as transforming the dependent variables.

The model contains two parameter matrices to be estimated: \( B \) and \( \delta \). Under the assumptions stated, estimates are

\[ \delta = (X'X)^{-1}X'Y \]

(3)

\[ \hat{\theta} = Y'[(I - X'(X'X)^{-1}X')]/N - rk(X) \]

(4)

In turn, testing the GLH requires estimating \( \delta \)

\[ \delta = CBU - \delta \]

(5)

where \( C \) and \( U \) must be such that \( Y \) is estimable (identically equal to some linear function of the expected value of \( Y \), which guarantees invariance with respect to a particular \( B \). If \( X \) is less than full rank, then infinitely many \( B \) estimates exist. If \( X \) is full rank, then a unique optimal estimate, \( B \), exists.

Test statistics are computed from three related matrices of sums of squares and cross products:

\[ H = (C(X'X)^{-1}C)' \]

(6)

\[ E = U'E \]

(7)

\[ T = H + E \]

(8)

These are respectively the hypothesis, error and total SSCP matrices. The above, of course, are sample estimators. Corresponding population parameters may be computed by replacing \( \delta \) with \( \hat{\delta} \) and \( \hat{\delta} \) with \( \delta \).

Test statistics are computed from \( H^{-1} \), \( E^{-1} \) or \( T^{-1} \). The eigenvalues of any one are sufficient for computing all standard multivariate test statistics. Without loss of generality, assume throughout that the eigenvalues have been sorted from largest to smallest. The eigenvalues of \( H^{-1} \) are always in \([0,1]\), with larger values corresponding to larger effects. The eigenvalues of \( T^{-1} \) are generalizations of squared canonical correlations.

The eigenvalues of \( E^{-1} \) and \( T^{-1} \) stand in simple relationships to those of \( H^{-1} \). With \( \lambda_k \) as the kth eigenvalue of \( H^{-1} \), the kth eigenvalue of \( E^{-1} \) is \( c_k = \lambda_k / (1 - \lambda_k) \), and the kth eigenvalue of \( T^{-1} \) is \( r_k = 1 - c_k \). The four common test statistics' names will be abbreviated: Roy's largest root (RLR), Hotelling-Lawley trace (HLT), Wilks' Lambda (W) and Pillai-Bartlett trace (PB). Also, \( \lambda_k \) indicates the kth estimated generalized canonical
The number of non-zero eigenvalues of \( HT \) (and non-zero canonical correlations) is the rank of \( \Omega \) and is at most \( s = \min(a,b) \), the minimum of the number of rows and columns of \( \Omega \), typically the rank of \( \Omega \). Recall that the trace of a square matrix is the sum of its eigenvalues, and the determinant of a (necessarily square) matrix is the product of its eigenvalues.

Under the null, \( H \) and \( T \) follow central Wishart distributions. Under the alternative, \( H \) and \( T \) follow noncentral Wishart distributions, while \( E \) is still central. The noncentrality matrix is defined in terms of the population hypothesis and error matrices, 

\[
\Omega_{pop} = \sigma^2 \left( C(X'X)C' \right)^{-1} \Omega
\]

and is

\[
\Omega = \Omega_{pop} \Omega_{pop}^{-1} \left[ N - rk(X) \right]
\]

For \( \Omega = I \), an important special case, this gives

\[
\Omega = \Omega_{pop} \Omega_{pop}^{-1} \left( N - rk(X) \right)
\]

Clearly the kth eigenvalue of \( \Omega \) is

\[
\lambda_k = \left[ N - rk(X) \right] c_k^2 (1 - \rho_k^2)
\]

Hence, using a relationship mentioned earlier,

\[
A_{W} = 1 - W^{1/s}
\]

This definition is not consistent with Johnson and Kotz (1976, Vol. 4, p. 126) in their discussion of the noncentral Wishart. They multiply \( \Omega \) by \( 1/2 \) to simplify expressions for the Wishart density. However, it is consistent with Johnson and Kotz (1970, Vol. 2, p. 130) in their discussion of the noncentral chi square distribution. Since the latter is a special case of the former which occurs if the Wishart matrix is \( 1 \times 1 \), the noncentrality matrix becomes the noncentrality value. The presence or absence of the \( 1/2 \) must always be checked in using any formula, table or software for power calculations.

NULL CASE CALCULATIONS

**Wilks’ Lambda**

Wilks’ Lambda, \( W \), may be characterized as the determinant of \( \Omega^{-1} \). It is also a simple function of the (generalized) canonical correlations:

\[
W = \prod_{k=1}^{s} (1 - \frac{\rho_k^2}{\sigma^2})
\]

Noting that \( W^{1/s} \) is the geometric mean of the \( (1-\rho_k^2)/\sigma^2 \)'s, a natural measure of multivariate association (Cramer and Anonymous, 1979) is

\[
A_{W} = 1 - W^{1/s}
\]

A similar transformation will be useful later.

For \( s = 1 \), \( W \) is distributed exactly as an \( F \) under the null (Rao, 1973, p. 555). Rao’s \( F \) approximation for the general case (Rao, 1973) is accurate and has the useful property of resolving to the exact forms for \( s = 1 \) or \( s = 2 \). Recent results of Nagarsenker and Suniaga (1983) support this observation. They provided an expression in terms of mixtures of beta distributions, of which Rao’s \( F \) is a special case.

Rao’s \( F \) approximation is

\[
W = \frac{ab}{W^2} \frac{1}{g}
\]

\[
g = \frac{n-(b-a+1)/2}{\sigma^2} = \frac{(b-a)/2}{\sigma^2}
\]

Degrees of freedom are \( ab \) and \( g(n-(b-a+1)/2) - (b-a)/2 \). Here \( n = N-rk(X) \), the error degrees of freedom and \( g^2 = [a^2b^2 - 4]/(a^2+b^2 - 5) \).

Note that if \( s = 1 \) or \( s = 2 \) this gives the appropriate exact result. It is easy to show that \( 1 < g < s \). Consider the characterization of \( W \) in (15). Then compare (17) to a form of the univariate \( F \) test stated in terms of the squared multiple correlation, \( \rho^2 \), for the univariate model:

\[
F = \frac{\rho^2}{1 - \rho^2}
\]

**Hotelling-Lawley Trace**

The Hotelling Lawley Trace Statistic is an analog of the ANOVA test statistic. It is

\[
HLT = tr(HC) = \frac{s}{\sigma^2}
\]

In terms of generalized canonical correlations it is

\[
HLT = \sum_k \lambda_k \left( 1 - \frac{\rho_k^2}{\sigma^2} \right)
\]

Note the similarity to the univariate test in (16). As mentioned earlier, for \( s = 1 \) all four multivariate test statistics become equivalent to each other and are described by an \( F \) distribution.

**Pillai’s Trace**

Pillai (1954) provided a convenient \( F \) approximation:

\[
A_{HLT} = \frac{HLT}{s} = \frac{ab}{1}\frac{1}{ab}
\]

\[
s(n-b+1)/2 + 2
\]

Degrees of freedom are \( ab \) and \( s(n-b+1)/2 + 2 \).

A multivariate measure of association based on \( HLT \) is

\[
A_{HLT} = \frac{tr(HE^{-1})}{s} = \frac{ab}{1}\frac{1}{ab}
\]

Clearly \( A_{HLT} \) is in \([0,1]\), and larger values correspond to larger effects. It is easy to show that...
Note the parallel to the univariate $F$ in (18), which, of course, is a special case.

Pillai-Bartlett Trace

The PB statistic can be computed most naturally as $tr(ln I)$. It may be characterized as estimating a generalized Mahalanobis distance. In terms of generalized canonical correlations

$$PB = \frac{s^2}{p^2} \sum_{k=1}^{n-b} \left( I - \frac{1}{s} \right)$$

(24).

Pillai (1954, 1960; also see Olson, 1976) provided an $F$ approximation:

$$F_{PB} = \frac{S}{p} \left( S - PB \right) (n + s - b)$$

(25).

Degrees of freedom are $ab$ and $s(n + s - b)$.

Consider the measure of multivariate association, which is the arithmetic average of the squared generalized canonical correlations, based on the PB:

$$A_{PB} = PB/s$$

(26).

Obviously,

$$F_{PB} = \frac{A_{PB}}{ab} \left( S - PB \right)$$

(27).

Note the parallel to the univariate $F$ in (18), which is, of course, a special case.

Roy's Largest Root

RLR is the largest eigenvalue of $HT^{-1}$ and hence equals the largest squared generalized canonical correlation. Finding expressions for its distribution has provided more trouble than for the other statistics. The most notable success was an approximation for general $s$ under the null (Pillai, 1965) which is sufficiently simple to have been implemented as a FORTRAN program by Harris (1975). The program has fewer than one hundred lines of code and involves one iteration loop running from one to $s$. Pillai indicated that the approximation should be expected to work only for small $p$ values.

One other practical result can be gained in treating the null distribution. As mentioned earlier, for $s = 1$, RLR is exactly distributed as an $F$. Since RLR is the maximum root of $HT^{-1}$, an upper bound can be approximated from an $F$ by using the form of the univariate test. The following gives an optimistically small significance level:

$$F_{max} = \frac{RLR}{n}$$

(28).

See the SAS canonical correlation program (SAS, 1982) for usage. The value of the bound lies in being able to be confident in not rejecting the null if the above $F_{max}$ does not.

Obviously more information is needed if $F_{max}$ exceeds the $F$ critical value for $a$ and $n$ d.f. Note that for $s = 1$ that the correct exact $F$ results only for univariate tests. It is therefore likely that a less optimistic bound exists which treats $s = 1$ special cases properly.

Non-Null Computations

Since exact general formulae are not available for any of the four test statistics for the null case, none exist for the non-null case, although some exact formulas exist for some special cases. For all but RLR approximate general formulas are available. These are of two types: mixtures of noncentral chi squares or noncentral $F$ approximations.

Chi Square Mixture Approximations

Sugihara and Fujikoshi (1969) provided a general approximation for noncentral probabilities (power) for W. Lee (1971) presented general approximations for W, HT and PB. Typically around ten chi square terms are used with complicated coefficients. The trace of the noncentrality matrix and the traces of its higher powers appear in the formulas for power calculations. Due to their complexity, the formulas are not reported here. Only special case evaluations of largest root power, not based on chi square approximations, are available as of this writing (Pillai and Jayachandran, 1967).

F Approximations

Proper consideration of the central $F$ approximations allow them to be generalized to the non-null case. To provide motivation for the approach, univariate power will be reviewed. The reader should recognize that the approximations are not known analytically to be either optimal or suboptimal except to the extent that such properties are inherited from the null case approximations. Extensive numerical comparison to the chi square approximations has been done.

The univariate central $F$ is given by (18). It will be useful to characterize the univariate noncentrality parameter as

$$\lambda = [N - rk(X)]p^2/(1 - p^2)$$

(29)

which is a special case of (14). Here $p^2$ is the squared multiple correlation for the model, and $N - rk(X) = n$ is the error d.f. Considering the above definition, it may be observed that

$$\lambda = F_d f. model$$

(30),

where

$$F_d f. model = \frac{p^2}{d f. model}$$

(31).

$$F_d f. error = \frac{1 - p^2}{d f. error}$$

(31).

855
FA is the F which would be observed if one obtained \( B = B \) and \( \bar{\xi} = \bar{\xi} \) in a sample, which is equivalent to observing exactly the alternative hypothesis of interest.

The expression given by (30) immediately suggests noncentral \( F \) approximations when considered with expressions given earlier for the central \( F \) approximations. For \( W \), powers may be approximated using a noncentral \( F \) with \( ab \) and \( g(n-(b-a+1)/2)-(ab-2)/2 \) d.f. and noncentrality parameter

\[
\lambda_W = \frac{1 - W^{1/9}}{W^{1/9}} \text{ } g(n-(b-a+1)/2)-(ab-2)/2
\]  

(32)

\[
\Lambda_W = \frac{1 - A_W}{g(n-(b-a+1)/2)-(ab-2)/2}
\]  

(33).

Naturally \( W \) is computed from \( H_{\text{pop}} \) and \( E_{\text{pop}} \).

For \( HLT \), power may be approximated using a noncentral \( F \) with \( ab \) and \( s(n-b+1)/2 \) d.f. and noncentrality parameter

\[
\lambda_{HLT} = \frac{s(n-b+1)/2}{HLT/s}
\]

(34)

\[
\Lambda_{HLT} = \frac{1 - A_{HLT}}{s(n-b+1)/2 - (ab-2)/2}
\]

(35).

\( HLT \) is computed from \( H_{\text{pop}} \) and \( E_{\text{pop}} \). Similarly, for \( PB \), power may be approximated using a noncentral \( F \) with \( ab \) and \( s(n+s-b) \) d.f. and noncentrality parameter

\[
\lambda_{PB} = \frac{PB}{s - PB}
\]

(36)

\[
\Lambda_{PB} = \frac{1 - A_{PB}}{s(n+s-b)}
\]

(37).

As with the other statistics, here \( \lambda_{PB} \) is computed from \( H_{\text{pop}} \) and \( E_{\text{pop}} \). Finally, the approach provides an upper bound on power for RLR. A noncentral \( F \) is used with \( a \) and \( n \) d.f. and noncentrality parameter

\[
\lambda_{RLR} = \frac{1 - \rho^2}{\rho^2}
\]

(38)

Here \( \rho^2 \) is the largest eigenvalue of \( H_{\text{pop}}(H_{\text{pop}} + E_{\text{pop}})^{-1} \). The power estimate may be very optimistic.

The \( F \) approximations for \( W, HLT \) and \( PB \) presented here appear to provide nearly two digits of accuracy in computing any power over a wide range of patterns of generalized canonical correlations. Such accuracy is sufficient for almost any practical situation. Furthermore, the \( F \) approximations are far simpler to implement and understand, and are less expensive to compute.

Random Effects and Repeated Measures

It should be emphasized that only fixed effect models are treated here. Univariate random effect tests can also yield convenient power calculations (based on multiples of central \( F \)'s). See Koele (1982) for a useful introduction to the ideas in this section. General power functions are unknown for tests of interaction involving both fixed and random effects.

Repeated measures can be analyzed with either a univariate or multivariate approach. Many authors (Keselman, Rogan, Mendoza and Breen, 1980; Huynh and Feldt, 1960) have documented the non-robustness of the univariate approach to typical repeated measures variance heterogeneity. The use of a multivariate approach avoids the problem and generates tests for which power can conveniently be computed. In practice the recommended Geisser-Greenhouse or Huynh-Feldt corrections cloud the power issue even more.

Computational Methods

Power calculations can suggest the need for dramatic changes in proposed experiments. For univariate analysis, the most useful sources are Cohen (1977), Koele (1982) and O'Brien (1982). O'Brien suggested a method which allows estimating power by having a linear models program provide most of the calculations. His method is extended below to the multivariate case. A more direct method is also reviewed.

Given the power estimates, simple programs can be written to do sample size calculations or effect size calculations, as with univariate problems. This will typically involve numerically inverting the power functions, which is quite simple to do. Details are omitted for the sake of brevity.

Two methods will be described for power computation: the "direct" method and an "indirect" method. The direct method assumes access to matrix manipulation software. The task may be split into three sub-tasks: 1) model specification, 2) test statistic calculation and 3) power calculations.

For direct and indirect methods, first specify

1) \( N \) and \( X \).
2) a conjecture for \( \bar{\xi} \).
3) an estimate for \( \bar{\xi} \).
4) \( C, U \) and \( \phi \).

The covariance matrix should usually be estimated from earlier data. Care should be taken to use the residuals from the earlier experiment. Since only residuals are needed, the earlier data need only include the covariates and dependent variables (Y variables), and the
previous study need not have involved the treatments of the proposed study.

**Direct Method**

If the conjecture for \( \beta \) is not taken from actual data, care must be taken to insure that impossible values are not chosen. \( \beta \) must be at least positive semi-definite, and usually positive definite. It is easy to specify matrices which look like covariance matrices but cannot be. Solutions include 1) specify matrices with simple patterns known to be possible, 2) express the matrix as a spectral decomposition, based on eigenvalue and eigenvector speculations or 3) use the model behind principal components (a special case of the factor analysis model).

For the direct method, the following computations are needed (equation numbers are those given earlier):

\[
\begin{align*}
\Omega &= C BU - \varnothing^2 \\
H_{\text{pop}} &= \Omega'[C(X'X)C']^{-1}\Omega \\
E_{\text{pop}} &= U'\Sigma U [N - \text{rk}(X)]
\end{align*}
\]

Any of the defining matrices may be used for computing the test statistics. Consider

\[
M = H_{\text{pop}}E_{\text{pop}}^{-1}
\]

The singular values of the singular value decomposition of \( M \) are also its eigenvalues. A routine for symmetric matrix eigenanalysis may be applied to the (symmetric) matrix

\[
M_a = F_{\text{pop}}F_{\text{pop}}^{-1}
\]

where

\[
FF' = E
\]

A convenient choice for \( F \) is the Cholesky factor of \( E \). \( M_a \) and \( M \) can easily be shown to have the same eigenvalues and usually distinct eigenvectors. \( W, HLT \) and \( PB \) are computed from the eigenvalues.

The last step for the direct method most conveniently uses the \( F \) approximations. First a significance level, \( \alpha \), must be specified. The parameters of the test then imply a critical value. This requires

\[
F_{\text{crit}} = F^{-1}(1 - \alpha, d.f., d.f.)
\]

where \( F^{-1}( \cdot \) \) indicates an inverse \( F \) cumulative distribution function. For HLT \( d.f., = ab \) and \( d.f., = s(n-b-1) + 2 \). For \( PB \) \( d.f., = ab \) and \( d.f., = q(f-(b-2)+1)/2 - (ab - 2)/2 \), with \( g^2 = [(a^2b^2-4)(a^2+b^2-5)] \). Next the noncentrality parameter, \( \lambda \), is computed from (32), (34) or (36), depending upon the test statistic of interest. Finally the power is approximated by

\[
P(X, B, \varnothing, \beta) = 1 - F(F_{\text{crit}}, \lambda, d.f., d.f.)
\]

Here \( F(\cdot) \) indicates a noncentral \( F \) cumulative distribution function. All of these calculations are straightforward in PROC MATRIX. See Peterson and Muller (1984) for convenient implementations.

**Indirect Method**

Without access to matrix manipulation software, the above direct method is impractical. The indirect method described below allows power computations using data transformation, multivariate linear models and principal components software. The indirect method is a modification and generalization of one given by O'Brien (1982) for univariate power calculation.

The indirect method differs in the steps used to compute the test statistics. As with the direct method, first specify \( X, B \), \( \varnothing \) and \( \beta \). Then create

\[
\begin{align*}
X_a &= \begin{bmatrix} X \\ X \end{bmatrix} \\
Y &= \begin{bmatrix} I_p \\ \sqrt{(2N - \text{rk}(X))} \end{bmatrix} \\
Q &= \begin{bmatrix} (N - p) \times p & 2 \end{bmatrix}
\end{align*}
\]

Let \( F_{\text{crit}} \) be such that

\[
F_{\text{crit}}F' = \Sigma
\]

\( \Sigma \) may be the Cholesky factor of \( \Sigma \). Alternatively consider the spectral decomposition of \( \Sigma \):

\[
\Sigma = V_{\Sigma}D_g(e_{\Sigma})V_{\Sigma}'
\]

\( F_{\text{crit}} \) may be chosen to be \( V_{\Sigma}D_g(e_{\Sigma}) \). If software only for principal components on a correlation matrix is available, \( D_g(e_{\Sigma}) \) may be used. Here \( D_g(\varnothing) \) is the diagonal matrix of standard deviations and \( P \) is the correlation matrix. It is easy to see that \( F_{\text{crit}} \) could alternately be chosen to be \( D_g(\varnothing)P \).

The spectral decomposition of \( P \) is

\[
P = V_{\Sigma}D_g(e_{\Sigma})V_{\Sigma}'
\]

Obviously the decomposition of \( \Sigma \) defined by \( F_{\text{crit}} \) is not unique. Next compute

\[
Y_* = \begin{bmatrix} X - vF' \Sigma \\ \Sigma \end{bmatrix}
\]

Use \( Y_* \) as input to a multivariate linear models program to fit the model

\[
Y_* = X_\beta + \varepsilon
\]
It is easy to show that, with obvious notation,
\[ E_\ast = B \]  \hspace{1cm} (52)
\[ E_\ast = X \]  \hspace{1cm} (53)
\[ E_\ast = \Omega \]  \hspace{1cm} (54)
\[ E_\ast = E_{pop} \]  \hspace{1cm} (56).

Error d.f. computed by the program are incorrect since 2N observations are used, but the error d.f. do not adversely affect any calculations. The above results indicate that providing a test of \( H_0 : \Omega = \Omega \) will automatically generate the proper generalized squared canonical correlations, the roots of \( H^T \Omega H \). Some program will only provide the roots of \( H^T \Omega H \). The simplest relationship between the roots of the two matrices make that no obstacle. The remaining calculations are the same as for the direct method.

CONCLUSIONS

Completely general methods for computing power in testing the general linear multivariate hypothesis (including MANOVA tests) are available, except for RLR. New F approximations greatly simplify the task while reducing accuracy only inconsequentially in practice. Standard multivariate linear models programs can be used to simplify power calculations.

REFERENCES


Peterson, B. L. and Muller, K. E. Estimating power for tests of the multivariate general linear hypothesis. Presented at the meeting of the SAS User's Group, 1984.