THE UNIVERSITY OF MINNESOTA

Statistics 5401 October 24, 2005

Testing MANOVA Hypotheses (corrected)

In multivariate linear models, including the multivariate analysis of variance (MANOVA), many hypothesis tests are based on a comparison of two p by p matrices, a *hypothesis matrix* **H** with degrees of freedom f_h (*hypothesis degrees of freedom*) and an *error matrix* **E** with degrees of freedom f_e (*error degrees of freedom*).

MANOVA generalizes univariate analysis of variance (ANOVA) to many variables, with a direct correspondence of many ideas. For example, **H** corresponds to an ANOVA *hypothesis sum of squares* SS_h, and **E** to an *error sum of squares* SS_e. In the one-way MANOVA discussion in Johnson and Wichern, **H** and **E** are called **B** (between) and **W** (within).

The hypothesis and error matrices H and E

You can compute **H** and **E** using formulas analogous to the corresponding univariate ANOVA SS except that you substitue a p by p matrix product $\mathbf{y}\mathbf{y}'$ wherever a y^2 would appear in the univariate formula, where $\mathbf{y}$ is a p×1 observation vector.

For example in the g-sample one-way *univariate* ANOVA with group sample sizes $n_1, \ldots, n_g$, the hypothesis SS and error SS for testing H_0: $\mu_1 = \mu_2 = \ldots = \mu_g$ are

$$SS_h = \sum_{1\leq j\leq g} n_j(\overline{y}_{.j} - \overline{y}_{..})^2$$

and

$$SS_e = \sum_{1\leq j\leq g} \sum_{1\leq i\leq n_j}(y_{ij} - \overline{y}_{.j})^2 = \sum_{1\leq j\leq g}(n_j-1)s_j^2,$$

where s_j^2 is the sample variance for group j.

In the corresponding g sample one-way MANOVA, the hypothesis matrix for testing the equality of the g mean *vectors* (H_0: $\boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_g$) is

$$\mathbf{H} = \sum_{1\leq j\leq g} n_j(\overline{\mathbf{y}}_{.j} - \overline{\mathbf{y}}_{..})(\overline{\mathbf{y}}_{.j} - \overline{\mathbf{y}}_{..})',$$

with error matrix

$$\mathbf{E} = \sum_{1\leq j\leq g} \sum_{1\leq i\leq n_j}(\mathbf{y}_{ij} - \overline{\mathbf{y}}_{.j})(\mathbf{y}_{ij} - \overline{\mathbf{y}}_{.j})' = \sum_{1\leq j\leq g}(n_j-1)\mathbf{S}_{j},$$

where $\mathbf{S}_j$ is the unbiased sample variance matrix for group j.

Similar explicit formulae are available in other multivariate analogues of more complex ANOVAs including randomized block, split plot, Latin squares, and incomplete block designs.

You can find more general formulas in an explicit linear model framework. Let the "full" model be

$$\underset{N\times p}{\mathbf{Y}} = \underset{N\times r}{\mathbf{Z}}\ \underset{r\times p}{\mathbf{B}} + \underset{N\times p}{\boldsymbol{\varepsilon}} = \underset{N\times p}{\mathbf{M}} + \underset{N\times p}{\boldsymbol{\varepsilon}}, \qquad \mathbf{M} \equiv E[\mathbf{Y}] = \mathbf{ZB},$$

where (a) the rows of $\boldsymbol{\varepsilon}$ are independent with zero means and common variance matrix $\boldsymbol{\Sigma}$, and (b) the r columns of $\mathbf{Z}$ are appropriate predictor variables - dummy variables and/or covariates. Each column of $\mathbf{B}$ contains coefficients for the corresponding column of $\mathbf{Y}$. Each row of $\mathbf{B}$ contains coefficients multiplying the corresponding column of $\mathbf{Z}$.

Linear hypotheses

Any linear hypothesis that can be tested can be put in the form H_0: $\mathbf{LB} = \mathbf{0}$, with some $q\times r$ matrix $\mathbf{L}$ with rank($\mathbf{L}$) = q = f_h. Then the hypothesis matrix used to test H_0 is

$$\mathbf{H} = (\mathbf{L}\hat{\mathbf{B}})'(\mathbf{L}\,(\mathbf{Z}'\mathbf{Z})^{-}\mathbf{L}')^{-1}(\mathbf{L}\hat{\mathbf{B}}), \quad \mathbf{E} = (\mathbf{Y} - \mathbf{Z}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{Z}\hat{\mathbf{B}}),$$

where $(\mathbf{Z}'\mathbf{Z})^{-}$ is a generalized inverse of $\mathbf{Z}'\mathbf{Z}$ (ordinary inverse when $\mathbf{Z}$ is of full rank), and $\hat{\mathbf{B}} = (\mathbf{Z}'\mathbf{Z})^{-}\mathbf{Z}'\mathbf{Y}$ is the usual "least squares" estimator of $\mathbf{B}$ (maximum likelihood estimator under normality). (Note: A generalized inverse $\mathbf{A}^{-}$ of a matrix $\mathbf{A}$ satisfies $\mathbf{AA}^{-}\mathbf{A} = \mathbf{A}$.)

For example, one way to express the usual one-way MANOVA uses $\mathbf{B} = [\boldsymbol{\mu}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_g]'$. Then you can exprfess the hypothesis of no treatment effects H_0: $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = \ldots = \boldsymbol{\alpha}_g$ as H_0: $\mathbf{LB} = \mathbf{0}$, where $\mathbf{L}$ is the g - 1 by g + 1 matrix

$$\mathbf{L} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 1 & 0 & 0 & 0 & \ldots & 0 & -1 \end{bmatrix}.$$

Whether you proceed by generalizing an ANOVA computation or by using the linear models approach, there are always N×N matrices $\mathbf{Q}_h$ and $\mathbf{Q}_e$ such that

$$\mathbf{H} = \mathbf{Y}'\mathbf{Q}_h\mathbf{Y} \quad \text{and} \quad \mathbf{E} = \mathbf{Y}'\mathbf{Q}_e\mathbf{Y},$$

where $\mathbf{Q}_h$ and $\mathbf{Q}_e$ are mutually orthogonal ($\mathbf{Q}_h\mathbf{Q}_e = \mathbf{Q}_e\mathbf{Q}_h = \mathbf{O}$) symmetric *projection matrices* ($\mathbf{Q}^2 = \mathbf{Q} = \mathbf{Q}'$) with ranks f_h and f_e, respectively. Moreover $\mathbf{Q}_e\mathbf{Z} = \mathbf{O}$ so that $\mathbf{Q}_e\mathbf{M} = \mathbf{Q}_e\mathbf{Z}\mathbf{B} = \mathbf{O}$.

For the hypothesis H_0: $\mathbf{LB} = \mathbf{O}$, these matrices are

$$\mathbf{Q}_h = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^-\mathbf{L}'(\mathbf{L}(\mathbf{Z}'\mathbf{Z})\mathbf{L}')^{-1}\ \mathbf{L}(\mathbf{Z}'\mathbf{Z})^-\mathbf{Z}' \text{ and } \mathbf{Q}_e = \mathbf{I}_N - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^-\mathbf{Z}'$$

The following propositions are true:

(1) H_0 is true if and only if $\mathbf{Q}_h\mathbf{M} = \mathbf{Q}_h\mathbf{Z}\mathbf{B} = \mathbf{O}$. Equivalently, H_0 is true if and only if $\mathbf{M}'\mathbf{Q}_h\mathbf{M} = \mathbf{B}'\mathbf{Z}'\mathbf{Q}_h\mathbf{Z}\mathbf{B} = \mathbf{O}$.

(2) When H_0 is true, $E[\mathbf{H}] = f_h\mathbf{\Sigma}$. When the rows of $\boldsymbol{\varepsilon}$ are independent $N_p(\mathbf{O},\mathbf{\Sigma})$, $\mathbf{H}$ has the Wishart distribution $W_{f_h}(\mathbf{H},\mathbf{\Sigma})$.

(3) $E[\mathbf{E}] = f_e\mathbf{\Sigma}$, <u>whether or not</u> H_0 is true. When the rows of $\boldsymbol{\varepsilon}$ are $N_p(\mathbf{O},\mathbf{\Sigma})$, $\mathbf{E}$ has the Wishart distribution $W_{f_e}(\mathbf{E},\mathbf{\Sigma})$, and is independent of $\mathbf{H}$.

(4) When H_0 is false, $E[\mathbf{H}] = f_h\mathbf{\Sigma} + \mathbf{M}'\mathbf{Q}_h\mathbf{M} \neq f_h\mathbf{\Sigma}$. When $\boldsymbol{\varepsilon}$ is $N_p(\mathbf{O},\mathbf{\Sigma})$, $\mathbf{H}$ is still independent of $\mathbf{E}$ and has what is known as the *non-central Wishart distribution* with *noncentrality matrix* $\mathbf{\Delta} = \mathbf{\Sigma}^{-1}\mathbf{M}'\mathbf{Q}_h\mathbf{M}$.

Note that $\mathbf{M}'\mathbf{Q}_h\mathbf{M}$ is formed from the expectation matrix $\mathbf{M} = E[\mathbf{Y}] = \mathbf{ZB}$ exactly the same way as $\mathbf{H} = \mathbf{Y}'\mathbf{Q}_h\mathbf{Y}$ is formed from $\mathbf{Y}$. This means that, even without constructing $\mathbf{Q}_h$, if you know how to compute $\mathbf{H}$ from $\mathbf{Y}$, you can compute $\mathbf{M}'\mathbf{Q}_h\mathbf{M}$ for any $\mathbf{M}$ by using $\mathbf{M}$ in place of $\mathbf{Y}$ in your calculation. For example, in one-way MANOVA, where

$$\mathbf{M} = E[\mathbf{Y}] = [\boldsymbol{\mu}_1,\boldsymbol{\mu}_1,\ \dots\ ,\boldsymbol{\mu}_1,\ \dots\ ,\ \boldsymbol{\mu}_g,\ \dots\ ,\boldsymbol{\mu}_g]', \text{ (note the transpose)}$$

with $\boldsymbol{\mu}_i = \boldsymbol{\mu} + \boldsymbol{\alpha}_i$ and with n_1 $\boldsymbol{\mu}_1$'s, n_2 $\boldsymbol{\mu}_2$'s, etc. Because

$$\mathbf{H} = \textstyle\sum_{1\leq j\leq g} n_j(\overline{\mathbf{y}}_{.j} - \overline{\mathbf{y}}_{..})(\overline{\mathbf{y}}_{.j} - \overline{\mathbf{y}}_{..})',$$

a formula for $\mathbf{M}'\mathbf{Q}_h\mathbf{M}$ is

$$\mathbf{M}'\mathbf{Q}_h\mathbf{M} = \textstyle\sum_{1\leq j\leq g} n_j[(\boldsymbol{\mu}_j - \overline{\boldsymbol{\mu}}_.)(\boldsymbol{\mu}_j - \overline{\boldsymbol{\mu}}_.)'],\quad \overline{\boldsymbol{\mu}}_. \equiv \textstyle\sum_{1\leq j\leq g} n_j\boldsymbol{\mu}_j/N.$$

Similarly, if you express $\mathbf{H}$ as $\mathbf{H} = (\mathbf{L}\hat{\mathbf{B}})'(\mathbf{L}(\mathbf{Z}'\mathbf{Z})^-\mathbf{L}')^{-1}(\mathbf{L}\hat{\mathbf{B}})$, then $\mathbf{M}'\mathbf{Q}_h\mathbf{M} = (\mathbf{LB})'(\mathbf{L}(\mathbf{Z}'\mathbf{Z})^-\mathbf{L}')^{-1}(\mathbf{LB})$, the same as $\mathbf{H}$ without the "hats" on $\mathbf{B}$.

This means you can consider $\mathbf{M}'\mathbf{Q}_h\mathbf{M}$ to be a *population* analogue of the *sample* hypothesis matrix $\mathbf{H}$, and the null hypothesis is equivalent to asserting $\mathbf{M}'\mathbf{Q}_h\mathbf{M} = \mathbf{0}$, that is, asserting that $E[\mathbf{H}] = f_h\boldsymbol{\Sigma}$.

Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ be the eigenvalues of $\mathbf{M}'\mathbf{Q}_h\mathbf{M}$ relative to $f_e\boldsymbol{\Sigma}$. These are also the ordinary eigenvalues of the (non-symmetric) *noncentrality matrix* $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{-1}\mathbf{M}'\mathbf{Q}_h\mathbf{M}$. Then, another way to state the null hypothesis $\mathbf{M}'\mathbf{Q}_h\mathbf{M} = \mathbf{0}$ is H_0: $\lambda_1 = \lambda_2 = \ldots = \lambda_p = 0$. Because the λ's are in decreasing order and are non-negative, you can restate the null hypothesis simply as H_0: $\lambda_1 = 0$. The λ_i's are population analogues of $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p$, the eigenvalues of $\mathbf{H}$ relative to $\mathbf{E}$.

It is important to remember that both $\mathbf{H}$ and its population analogue $\mathbf{M}'\mathbf{Q}_h\mathbf{M}$ together with relative eigenvalues $\{\lambda_i\}$ and $\{\hat{\lambda}_i\}$ are associated with a specific null hypthesis. For each null hypothesis being tested there is a different $\mathbf{H}$ and a different $\mathbf{M}'\mathbf{Q}_h\mathbf{M}$ with different relative eigenvalues.

Distribution of relative eigenvalues

The null distribution (distribution when H_0:$\mathbf{M}'\mathbf{Q}_h\mathbf{M} = \mathbf{0}$ is true) of the s non-zero eigenvalues $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_s$ is quite complicated. However, with multivariate normal errors it depends only on three integer or half integer quantities.

These quantities are, in the standard notation,

$s \equiv \min(p, f_h) \geq 1$

$m \equiv (|p - f_h| - 1)/2 \geq -1/2$

$n \equiv (f_e - p - 1)/2 \geq -1/2$ (when $f_e < p$, $\mathbf{E}^{-1}$ does not exist).

The quantities s, m and n are somewhat analogous to degrees of freedom in univariate F-tests. When $p = 1$ (univariate), $s = 1$, $m = f_h/2 - 1$, $n = f_e/2 - 1$.

The non-null distribution (distribution when H_0 is false, $\mathbf{M}'\mathbf{Q}_h\mathbf{M} \neq \mathbf{0}$) depends on these same quantities as well as the population eigenvalues $\lambda_1, \ldots, \lambda_s$.

There an interesting duality between f_h and p in the definitions of these quantities and corresponding identities in sampling distributionsw. If you substitute $\tilde{f}_h \equiv p$ for f_h, $\tilde{p} \equiv f_h$ for p, and $\tilde{f}_e \equiv f_e + f_h - p$ for f_e, then s, m, and n are unchanged and it can be verified that the distribution of non-zero eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_s$ is unchanged. For example,

$$(f_e - p - 1)/2 \Rightarrow (\tilde{f}_e - \tilde{p} - 1)/2 = (f_e + f_h - p - f_h - 1)/2 = n.$$

Test statistics for multivariate linear hypotheses

Many test statistics have been proposed for testing H_0: $E[\mathbf{H}] = f_h\mathbf{\Sigma}$, that is, $\mathbf{M}'\mathbf{Q}_h\mathbf{M} = \mathbf{0}$. Several are based on the sample relative eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p$ of $\mathbf{H}$ relative to $\mathbf{E}$ (ordinary eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$). By the preceding, the null distribution of such test statistics depends only on s, m and n.

Remark: when $f_h < p$, then $\hat{\lambda}_{f_h+1} = \ldots = \hat{\lambda}_p = 0$, that is, there are at most $s \equiv \min(p, f_h)$ non-zero $\hat{\lambda}$'s. Also, whether or not H_0 is true, the population values $\lambda_{f_h+1} = \ldots = \lambda_p = 0$, so the null hyothesis is equivalent to H_0: $\lambda_1 = \lambda_2 = \ldots = \lambda_s = 0$.

When H_0 is not true (that is, $E[\mathbf{H}] \neq f_h\mathbf{\Sigma}$ or $\mathbf{M}'\mathbf{Q}_h\mathbf{M} \neq \mathbf{0}$ or $\lambda_1 > 0$), the power (probability of rejecting H_0) of any test based on the sample eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_s$ depends only on the population eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_s$ defined above.

Special cases when s = 1 (p = 1 or f_h = 1)

When p = 1 (the univariate case), under H_0, there is only one relative eigenvalue $\hat{\lambda}_1$ and

$$\hat{\lambda}_1 = SS_h/SS_e = (f_h/f_e)F(f_h, f_e)$$
(*central F* on f_h and f_e d.f.),

In the null null case (H_0 false)

$$\hat{\lambda}_1 = (f_h/f_e)F(f_h, f_e; \delta^2)$$
(*non-central F* on f_h and f_e d.f. and *non-centrality parameter* δ^2),

where $\delta^2 = \lambda_1 = \mathbf{M}'\mathbf{Q}_h\mathbf{M}/\sigma^2$ ($\mathbf{M}$ is a vector when p = 1).

When $f_h = 1$ (essentially the case of Hotelling's T^2), the only non-zero eigenvalue is $\hat{\lambda}_1 = \mathrm{tr}\, \mathbf{E}^{-1}\mathbf{H} = T^2/f_e \ \tilde{=}\ \{p/(f_e - p + 1)\}\, F(p, f_e-p+1)$ when H_0 is true. When H_1 is true, $\hat{\lambda}_1 \ \tilde{=}\ \{p/(f_e - p + 1)\}F(p, f_e-p+1; \lambda_1)$ (non-central F).

These two cases cover all the possibilities when $s = \min(f_h, p) = 1$. Thus the only distributional difficulties for test statistics based on the $\hat{\lambda}_i$'s are when when s > 2, that is when $p \geq 2$ and $f_h \geq 2$.

Test statistics based on relatative eigenvalues

Here are some of the test statistics based on the relative eigenvalues that have been proposed, together with information about their exact or approximate distributions.:

1. *Roy's maximum root test*

Reject H_0 for "large" $\hat{\lambda}_1$. This is equivalent to rejecting H_0 for "large" $\hat{\theta}_1 \equiv \hat{\lambda}_1/(1 + \hat{\lambda}_1)$, since $\hat{\theta}$ is an increasing function of $\hat{\lambda}$.

Charts for the 1% and 5% critical values of $\hat{\theta}_1$ for $2 \leq s \leq 5$ as computed by Heck (1960) are in a handout posted on the class web site (`http://www.stat.umn.edu/~kb/classes/5401/files/RoysTest.pdf`). I don't know of a useful approximation to the distribution of Roy's test.

For s = 2, $P(\hat{\theta}_1 \leq x) = I_x(2m+2,2n+2) - Cx^{m+1}(1-x)^{n+1}I_x(m+1,n+1)$, where $I_x(a,b)$ is an imcomplete beta function computed in MacAnova by **`cumbeta(x,a,b)`** and $C = \sqrt{\pi}\times\Gamma(m+n+5/2)/\{\Gamma(m+3/2)\Gamma(n+3/2)\}$, where $\Gamma(z)$ is the gamma function. When N is an integer, $\Gamma(N) = (N-1)!$ and $\Gamma(N+1/2) = \sqrt{\pi}\times 1\times 3\times\ldots\times(2N-1)/2^N$.

2. *Likelihood ratio based test (Wilks' test)*

Reject for "small" values of

$$\Lambda^* \equiv \det(\mathbf{E})/\det(\mathbf{E} + \mathbf{H}) = 1/\det(\mathbf{I}_p + \mathbf{E}^{-1}\mathbf{H}) = \Pi_{1\leq i\leq s}\{1/(1 + \hat{\lambda}_i)\}.$$

Rao (1948) (see also Anderson (1965)) gave the following expression for an approximation to the cumulative distribution function (cdf) of a multiple of $\log \Lambda^*$. Define

$$m_1 \equiv f_e - (p - f_h + 1)/2 = 2n + m + s + 1 \text{ and } P_g \equiv P(\chi_g^2 > x).$$

Then the upper tail probability for $-m_1\log \Lambda^* = m_1\sum_{1\leq i\leq s}\log(1 + \hat{\lambda}_i)$ under H_0 is

$$P(-m_1\log \Lambda^* > x) =$$
$$P_f + \beta_1(P_{f+4} - P_f)/m_1^2 + \{\beta_2(P_{f+8} - P_f) - \beta_1^2(P_{f+4} - P_f)\}/m_1^4 + O(1/m_1^6),$$

where

$$f \equiv pf_h$$
$$\beta_1 \equiv (pf_h/48)(p^2 + f_h^2 - 5),$$

$$\beta_2 \equiv \beta_1{}^2/2 + (pf_h/1920)(3p^4+3f_h{}^4+10p^2f_h{}^2 - 50(p^2 + f_h{}^2) + 150).$$

Using just the first term, that is, using the approximation $P(-m_1\log \Lambda^* > x) \cong P_f = P(\chi_f{}^2 > x)$ is equivalent to treating

$$-m_1\log \Lambda^* = \{f_e - (p-f_h+1)/2\}\sum_{1\leq j\leq s}\log(1 + \hat{\lambda}_j)$$

as a $\chi_f{}^2$ random variable, where $f = pf_h$. This is a widely used approximation that is generally sufficiently accurate. The additional terms, which go rapidly to zero as $m_1 \rightarrow \infty$, serve to correct this first approximation.

MacAnova macro **cumwilks()** with keyword phrase **useF:F** uses this series when $\min(p, f_h) > 2$.

Fujikoshi (1973) derived a similar, more complicated expression for the non-null distribution when $\lambda_1 > 0$. The leading term is

$$P(-m_1 \log \Lambda^* > x) = P_f(\delta^2) + O(1/m_1),\ P_f(\delta^2) = P(\chi_f{}^2(\delta^2) > x),$$

$$\delta^2 = \mathrm{tr}\boldsymbol{\Delta} = \mathrm{tr}\ \boldsymbol{\Sigma}^{-1}\mathbf{M}'\mathbf{Q}_h\mathbf{M} = \sum_{1\leq j\leq s}\lambda_j.$$

Here $\chi_f{}^2(\delta^2)$ represents the non-central chi-squared distribution with non-centrality parameter δ^2 and, as before, the λ_j's are the eigenvalues of $\mathbf{M}'\mathbf{Q}_h\mathbf{M}$ relative to $\boldsymbol{\Sigma}$. You can use this to compute the approximate power of the likelihood ratio test for specified δ^2.

Rao's approximation is essentially an adjustment to the standard large sample result for the ratio λ of maximized likelihoods. This result says that, for large samples, $-2 \log \lambda$ is approximately $\chi_f{}^2$, where f is the number of restrictions H_0 imposes on the parameters. In this situation, $\lambda = (\Lambda^*)^{2/N}$, where N is the number of rows (cases) in $\mathbf{Y}$ and $f = pf_h$, the number of elements in the f_h by p matrix $\mathbf{LB}$, all of which are hypothesized to be 0. Thus

$$-2 \log \lambda = N \log \Lambda^* = N\sum_{1\leq j\leq s}\log(1 + \hat{\lambda}_j).$$

Rao's adjustment replaces N by m_1. Generally, for large N, $m_1/N \rightarrow 1$. In the one-way MANOVA case where $f_h = g-1$ and $f_e = N - g$, $m_1 = N - g - (p+g-1+1)/2$ and hence $m_1/N = 1 - (1/2)(g + p + 1)/N \rightarrow 1$ for large N.

Rao (1951, 1973) derived a different approximation to the null distribution of Λ^* when $s = \min(p, f_h) > 1$ based on the F distribution:

$$(1 - (\Lambda^*)^{1/t})/(\Lambda^*)^{1/t} \cong \{pf_h/(m_1t - v)\}\ F(pf_h,\ m_1t-v),$$

where

$$t = \{(p^2f_h^2 - 4)/(p^2 + f_h^2 - 5)\}^{1/2} \text{ and } v = (pf_h - 2)/2.$$

Thus $(m_1t - v)(1 - (\Lambda^*)^{1/t})/(pf_h\,(\Lambda^*)^{1/t})$ is approximately $F(pf_h,\ m_1t-v)$.

MacAnova macro **cumwilks()** uses this approximation to the distribution of Λ^* when **useF:F** is not an argument.

Exact distribution of Λ^* when $s \leq 2$

When $s \leq 2$, Rao's formula involving the F distribution is *exactly* correct (not an approximation). It reduces to the following special cases:

$s = 1$: $t = 1$, $m_1t - v = 2n + 2$

$$\{(n + 1)/(m + 1)\}(1 - \Lambda^*)/\Lambda^* \cong F(2m + 2,\ 2n + 2)$$

$s = 2$: $t = 2$, $m_1t - v = 4n + 4$,

$$\{(2n + 2)/(2m + 3)\}\{1 - (\Lambda^*)^{1/2}\}/(\Lambda^*)^{1/2} \cong F(4m + 6,\ 4n + 4).$$

In terms of p, f_h, and f_e, these cases are

$f_h = 1$, any p: $\{(f_e - p + 1)/p\}(1 - \Lambda^*)/\Lambda^* \cong F(p,\ f_e-p+1)$

$f_h = 2$, any $p \geq 2$: $\{(f_e - p + 1)/p\}\{1 - (\Lambda^*)^{1/2}\}/(\Lambda^*)^{1/2} \cong F(2p,\ 2(f_e-p+1))$

$p = 1$, any f_h: $\{f_e/f_h\}(1 - \Lambda^*)/\Lambda^* \cong F(fh,\ f_e)$

$p = 2$, any $f_h \geq 2$: $\{(f_e-1)/f_h\}\{1 - (\Lambda^*)^{1/2}\}/(\Lambda^*)^{1/2} \cong F(2f_h,\ 2(f_e-1))$

3. Hotelling's generalized T_0^2 or trace test:

$$\text{Reject } H_0 \text{ for "large" } T_0^2 \equiv f_e\,\mathrm{tr}(\mathbf{E}^{-1}\mathbf{H}) = f_e \textstyle\sum_{1\leq j\leq s} \hat{\lambda}_j\ .$$

When $s = 1$, $T_0^2 = f_e\hat{\lambda}_1 = f_e\Lambda^*/(1 - \Lambda^*) \cong f_e\{(m+1)/(n+1)\}F(2m+2,\ 2n+2)$. When $f_h = 1$, $T_0^2 = T^2$ (Hotelling's ordinary T^2) and has null distribution $\{(pf_e)/(f_e - p + 1)\}F(p,\ f_e-p+1)$. When $p = 1$, $T_0^2 \cong F(f_h,f_e)$.

Remark When $f_h > 1$, T_0^2 is in fact a generalization of Hotelling's T^2 in that it can be put in the form $T_0^2 = (\hat{\boldsymbol{\theta}} - \mathbf{0})'[\hat{V}[\hat{\boldsymbol{\theta}}]]^{-1}(\hat{\boldsymbol{\theta}} - \mathbf{0})'$. You only need to take all $f = f_hp$ elements of $\mathbf{LB}$ and string them into a long vector $\boldsymbol{\theta}$ of length f and similarly string out the elements of $\hat{\mathbf{B}}$ into $\hat{\boldsymbol{\theta}}$, and take as $\hat{V}[\hat{\boldsymbol{\theta}}]$ the "natural" estimator obtained by substituting $\mathbf{S} = (f_e)^{-1}\mathbf{E}$ for $\mathbf{S}$ in an equation for $V[\hat{\boldsymbol{\theta}}]$. From what is known of statistics of this type, in large enough samples, the distribution of T_0^2 is approximately χ_f^2.

For moderately large f_e, Fujikoshi (1973) found an adjustment to T_0^2 whose distribution is better approximated by χ_f^2. Moreover he found terms to adjust the P-value computed from χ_f^2 to get a better approximation. Let $m_2 \equiv f_e - p - 1 = 2n$ and define the adjusted statistic to be

$$T \equiv m_2 \,\mathrm{tr}(\mathbf{E}^{-1}\mathbf{H}) = (m_2/f_e)T_0^2 = (1 - (p+1)/f_e)\, T_0^2.$$

Then, approximately, upper tail probabilities are

$$P(T > x) = P_f + \{f\times(p+f_h+1)/(4m_2)\}(P_f - 2P_{f+2} + P_{f+4})$$
$$+ \{f/(96m_2^2)\} \textstyle\sum_{0\leq j\leq 4}(-1)^j\, h_jP_{g+2j} + O(1/m_2^2),$$

where $f = pf_h$ and $P_g \equiv P(\chi_g^2 > x)$ as before, and

$$h_0 = (3f-8)(p+f_h+1)^2 + 4g \qquad h_1 = 12f(p+f_h+1)^2$$
$$h_2 = 6(3f+8)(p+f_h+1)^2 \qquad h_3 = 4((3f+16)(p+f_h+1)^2 + 4g)$$
$$h_4 = (3f+24)(p+f_h+1)^2 + 12g \qquad g = (p+1)(f_h+1) + 2.$$

Fujikoshi also gives the $O(1/f_e^3)$ term.

MacAnova macro **cumtrace()** uses this approximation by default.

Using just the leading term ($P(T > x) \cong P_f$) is often sufficiently accurate. This means you treat the modified statistic

$$T \equiv (m_2/f_e)\times T_0^2 = (f_e - p - 1)\mathrm{tr}(\mathbf{E}^{-1}\mathbf{H}) = (f_e - p - 1) \textstyle\sum_{1\leq j\leq s} \hat\lambda_j$$

as χ_f^2, $f = pf_h$. To order $1/m_2$, the null upper α probability points (critical values) of T are

$$T(\alpha) = \chi_f^2(\alpha) - m_2^{-1}\{(p+f_h+1)/2\}\{\chi_f^2(\alpha) - (\chi_f^2(\alpha))^2/(f+2)\}.$$

When H_0 is false, a Fujikoshi (1973) found a similar more complicated series involving non-central χ^2. The leading term is the same as for $-m_1\log\Lambda^*$, that is, $P(T > x) = P_f(\delta^2) + O(1/f_e)$.

4. *Pillai's trace criterion*

Pillai's trace test is

$$\text{Reject } H_0 \text{ for "large" } V \equiv m_3\times\mathrm{tr}(\mathbf{H}+\mathbf{E})^{-1}\mathbf{H} = m_3\times\textstyle\sum_{1\leq i\leq s}\hat\lambda_i/(1+\hat\lambda_i),$$
$$m_3 \equiv (f_e+f_h) = 2(n + m + s + 1).$$

You can obtain a large f_e approximation for the null hypothesis tail probability $P(V > x)$ by changing the sign of the second term in the expression for $P(T > x)$ above and replacing m_2 and m_2^2 by m_3 and m_3^2.

MacAnova macro **cumtrace()** with keyword phrase **pillai:T** uses this approximation.

Upper tail probability points (critical values) accurate to $O(m_3^{-1})$ are

$$V(\alpha) = \chi_f^2(\alpha) - m_3^{-1}\{(p+f_h+1)/2\}\{\chi_f^2(\alpha) - (\chi_f^2(\alpha))^2/(f+2)\},\ f = pf_h.$$

When H_0 is false, the leading term of the power function is again

$$P(V > x) = P_f(\delta^2) + O(1/(f_e+f_h)).$$

Differences among the tests

How should you chose among these tests? The large sample (actually large f_e) form of the power functions (non-null rejection probabilities) for $-m_1 \log \Lambda^*$, T_0^2, and V (but not $\hat{\lambda}_1$) are all the same, that is, $P_f(\delta^2)$, based on non-central χ_f^2 with non-centrality parameter $\delta^2 = \mathrm{tr}\boldsymbol{\Delta} = \mathrm{tr}\ \boldsymbol{\Sigma}^{-1}\mathbf{M}'\mathbf{Q}_h\mathbf{M} = \sum_{1\leq i\leq s}\lambda_j$.

You might have expected this. In a "neighborhood" of H_0, that is, when $\lambda_1,\ldots,\lambda_s$ are all small, when N is large, all three test statistics are essentially equivalent. This is the only case that matters in large samples, since otherwise virtually any test will have power close to 1. For example, when f_e is large and therefore m_2 and m_3 are also large,

$$-m_1 \log \Lambda^* = -m_2(1 - (p+f_h+1)/2m_2) \sum_{1\leq i\leq s}\log(1 + \hat{\lambda}_j) \cong m_2 \sum_{1\leq i\leq s}\hat{\lambda}_j = T$$

and

$$V = m_3 \sum_{1\leq j\leq s}(\hat{\lambda}_j/(1+\hat{\lambda}_j)) \cong m_2(1 + (p+f_h+1)/m_2)\sum_{1\leq j\leq s}\hat{\lambda}_j \cong T.$$

Thus for large f_e with H_1 "near" H_0, the three statistics are essentially the same. In fact, in power computations that have appeared in the literature, there does not seem to be much to choose between the various tests, at least for large f_e. The differences in power tend to be on the order of 0.02, an amount insignificant compared to the uncertainty arising from the inexactness of guesses for a value for δ^2.

Roy's maximum root test is different. When you think the appropriate alternative hypothesis has rank 1 in the sense that $\mathbf{M}'\mathbf{Q}_h\mathbf{M}$ has rank 1 or $\lambda_1 > 0$, $\lambda_j = 0$, $j > 1$, or almost of rank 1 ($\lambda_1 >> \lambda_2 \geq \lambda_3 \geq \ldots$), Roy's maximum root test using only $\hat{\lambda}_1 = \hat{\lambda}_{max}$ is probably to be preferred, focussing as it does on the single largest sample eigenvalue. Some of the power that the other tests implicitly expend on testing for non-zero λ_2, λ_3, etc. is "wasted" in this case. Conversely, if several λ's are non-zero, without λ_1 being dominant, Roy's test may fail to reject H_0 because it ignores the values of $\hat{\lambda}_j$, $j > 1$. If it is expected that $\mathbf{M}'\mathbf{Q}_h\mathbf{M}$ has rank 2 with both λ_1 and λ_2 large, a test based on $\hat{\lambda}_2$, $\hat{\lambda}_1 + \hat{\lambda}_2$ or $\log(1 + \hat{\lambda}_1) + \log(1 + \hat{\lambda}_2)$ might be even better.

REFERENCES

Anderson, T.W. (1958) *An Introduction to Multivariate Statistical Analysis*, Wiley

Fujikoshi, Yasunori (1973) Asymptotic formulas for the distributions of three statistics for multivariate linear hypothesis, *Ann. Inst. Statist. Math.* **25** 423-437

Heck, D. L. (1960) Charts of some upper percentage points of the distribution of the largest characteristic root, *Ann.Math.Statist.* **31** 625-642

Rao, C. Radhakrishna (1948) Tests of significance in multivariate analysis, *Biometrika* **35** 58-79

Rao, C. Radhakrishna (1951) An asymptotic expansion of the distribution of Wilks' criterion *Bull. Inst. Inter. Statist.* **33** (2) , 1-28

Rao, C. Radhakrishna (1973) *Linear Statistical Inference and Its Applications*, 2nd Edition, Wiley