Statistical Papers 45, 111-122 (2004)

Statistical Papers

The geometrical interpretation of statistical tests in multivariate linear regression

Øyvind Langsrud

MATFORSK, Osloveien 1, 1430 Ås, Norway
(e-mail: oyvind.langsrud@matforsk.no)

Received: December 20, 2001; revised version: September 5, 2002

A geometrical interpretation of the classical tests of the relation between two sets of variables is presented. One of the variable sets may be considered as fixed and then we have a multivariate regression model. When the Wilks' lambda distribution is viewed geometrically it is obvious that the two special cases, the F distribution and the Hotelling T^2 distribution are equivalent. From the geometrical perspective it is also obvious that the test statistic and the p-value are unchanged if the responses and the predictors are interchanged.

KEY WORDS : Multivariate analysis, Wilks' lambda distribution, MANOVA, Canonical correlation, Random rotation, Invariance.

1 Introduction

The classical regression line formed by two univariate samples depends on the choice of predictor and response, but testing the regression parameters (if zero) in the two regressions yield identical t-statistics and p-values. The t-test generalises to an F-test when there are several predictor variables. When both sets of variables are multivariate there are several available generalisations of the F statistic, e.g. Wilks' lambda. For the two regression models ($\boldsymbol{y}$ on $\boldsymbol{x}$ and $\boldsymbol{x}$ on $\boldsymbol{y}$), these generalised test statistics will still take exactly the same value and the associated null distributions are also the same.

The equivalence of the two tests is unknown to many statisticians, but if the tests are understood geometrically the equivalence is obvious. This is one example of how the geometry of the tests can provide better intuitive understanding. Geometrical formulations are described by some authors (Dempster (1969), Eaton (1983), Wickens (1995) and Saville and Wood (1996)), but unfortunately standard statistical textbooks do not offer much attention to this topic. This paper presents a simple and unifying geometrical description of the most common statistical tests. In particular section 2.2 shows a nice approach to the Wilks' lambda statistic.

2 Testing the hypothesis $\mathrm{H}_0: \boldsymbol{\Sigma}_{xy} = 0$

On the basis of two data matrices with n independent observations, $\boldsymbol{X}(n \times p)$ and $\boldsymbol{Y}(n \times q)$, we are interested in testing if there is any dependence between the p x-variables and the q y-variables. Let the partitioned vector $\boldsymbol{g} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} = [x_1, \ldots, x_p | y_1, \ldots, y_q]^T$ represent the observations on both sets of variables. We assumed that the distribution of $\boldsymbol{g}$ is multinormal and the covariance matrix of $\boldsymbol{g}$ is partitioned as:

$$\mathrm{Cov}(\boldsymbol{g}) = \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix} . \tag{1}$$

We have independence between the two sets of variables if and only if $\boldsymbol{\Sigma}_{xy} = 0$ (Morrison 1976, chap. 7.5). So our hypothesis of interest is

$$\mathrm{H}_0: \quad \boldsymbol{\Sigma}_{xy} = 0 \qquad \text{against} \qquad \mathrm{H}_1: \quad \boldsymbol{\Sigma}_{xy} \neq 0 . \tag{2}$$

In the following we want to simplify the problem of finding a test statistic, which is a function of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

2.1 Reducing the problem by invariance properties

To avoid the intercept terms having influence on the analysis both data sets are first centered, i.e. for each x- and y-variable the mean value is subtracted:

$$\boldsymbol{X}_c = \boldsymbol{X} - \tfrac{1}{n}\mathbf{1}_{(n\times1)}\mathbf{1}^T_{(n\times1)}\boldsymbol{X} \tag{3}$$

$$\boldsymbol{Y}_c = \boldsymbol{Y} - \tfrac{1}{n}\mathbf{1}_{(n\times1)}\mathbf{1}^T_{(n\times1)}\boldsymbol{Y} . \tag{4}$$

Geometrically all column vectors are now projected onto the $(n-1)$-dimensional subspace, in the following called $\mathcal{S}$, that is orthogonal to the constant vector, $\mathbf{1}_{(n\times1)}$.

We presume that the analysis is invariant under nonsingular linear transformations of either set. In other words, replacing $\boldsymbol{X}_c$ (or $\boldsymbol{Y}_c$) with another $n \times p$-matrix (or $n \times q$-matrix) will not change the results as long as the columns of the new matrix span the same subspace (column space) as the columns of the original matrix. Therefore, instead of working with the centered matrices, we can work with matrices that consist of vectors forming orthonormal bases for the column spaces of $\boldsymbol{X}_c$ and $\boldsymbol{Y}_c$. The simplest way to obtain such matrices is to perform the Gram-Schmidt orthogonalisation process which is also known as the QR decomposition (Strang 1988, chap. 3.4):

$$\boldsymbol{X}_{\text{c}} = \boldsymbol{X}_{\text{o}} \boldsymbol{L}_X \tag{5}$$

$$\boldsymbol{Y}_{\text{c}} = \boldsymbol{Y}_{\text{o}} \boldsymbol{L}_Y \,. \tag{6}$$

Here $\boldsymbol{X}_{\text{o}}$ and $\boldsymbol{Y}_{\text{o}}$ are the "Q"-matrices in the QR-decompositions and $\boldsymbol{L}_X$ and $\boldsymbol{L}_Y$ are the matching "R"-matrices.

The invariance properties have reduced the problem of testing relations between data matrices to testing relations between subspaces. These subspaces are represented by the matrices $\boldsymbol{X}_{\text{o}}$ and $\boldsymbol{Y}_{\text{o}}$ of orthonormal columns. A measure of the relation between the two subspaces can therefore be used as a test statistic. The problem is how to define such a measure.

The QR decomposition is just an example of a decomposition that produces orthogonal bases. A test statistic should not depend on the specific decomposition. In other words, the test statistic should be invariant under the transformation

$$(\boldsymbol{X}_{\text{o}}, \boldsymbol{Y}_{\text{o}}) \rightarrow (\boldsymbol{X}_{\text{o}} \boldsymbol{U}_X, \boldsymbol{Y}_{\text{o}} \boldsymbol{U}_Y) \tag{7}$$

where $\boldsymbol{U}_X(p \times p)$ and $\boldsymbol{U}_Y(q \times q)$ are arbitrary orthogonal matrices.

Another type of invariance follows from the normal distribution. The simultaneous distribution of $(\boldsymbol{X}, \boldsymbol{Y})$ is invariant under the transformation $(\boldsymbol{X}, \boldsymbol{Y}) \rightarrow (\boldsymbol{R}\boldsymbol{X}, \boldsymbol{R}\boldsymbol{Y})$, where the matrix $\boldsymbol{R}$ is an arbitrary orthogonal matrix that satisfies

$$\boldsymbol{R}\boldsymbol{1}_{(n \times 1)} = \boldsymbol{1}_{(n \times 1)} \,. \tag{8}$$

It follows directly that the simultaneous distribution of $\boldsymbol{X}_{\text{o}}$ and $\boldsymbol{Y}_{\text{o}}$ is invariant under the transformation

$$(\boldsymbol{X}_{\text{o}}, \boldsymbol{Y}_{\text{o}}) \rightarrow (\boldsymbol{R}\boldsymbol{X}_{\text{o}}, \boldsymbol{R}\boldsymbol{Y}_{\text{o}}) \,. \tag{9}$$

We refer to $\boldsymbol{R}$ as a rotation matrix that performs rotation in $\mathcal{S}$. In the present paper all orthogonal matrices are called rotation matrices. The result of a rotation matrix multiplied by a data matrix can be interpreted as a representation of the data column vectors in a rotated co-ordinate system or as a rotated version of the original data. For details of rotation matrices, see Appendix.

2.2 The statistic Λ

A test statistic that satisfies the invariance properties, (7) and (9) can be constructed based on the following geometrical volume idea. The column vectors of $\boldsymbol{X}_{\text{o}}$ form a p-dimensional cube. Since all the orthogonal vectors have length one, the volume of the cube equals one. For a general $i \times j$ matrix $\boldsymbol{M}$ where $j \leq i$, the column vectors form a j-dimensional parallelepiped

(Strang 1988, chap. 4.1). The volume of this parallelepiped can be calculated from the relation:

$$\text{Volume}^2 = \left|M^T M\right| . \tag{10}$$

If all column vectors are restricted to have the length one, then the maximal volume is one. The maximum is obtained when all column vectors are mutually orthogonal and the volume decreases as the "dependence" between the column vectors increases. To measure the relation between X_o and Y_o, an idea is to calculate the volume of the $(p+q)$-dimensional parallelepiped formed by the p column vectors of X_o together with the q column vectors of Y_o (assume $p+q < n$). Since both X_o and Y_o have orthonormal columns, the volume is a function of how the X_o-columns are related to the Y_o-columns. Furthermore, one can view the $(p+q)$-dimensional parallelepiped as being "spanned by" the subspace represented by X_o and the subspace represented by Y_o. Hence, the volume is one way to measure the relationship between the two subspaces. We denote the squared volume by Λ:

$$\Lambda = \left|[X_o \;\; Y_o]^T [X_o \;\; Y_o]\right| \tag{11}$$

$$= \left|I_p - {X_o}^T Y_o {Y_o}^T X_o\right| \tag{12}$$

$$= \left|I_q - {Y_o}^T X_o {X_o}^T Y_o\right| . \tag{13}$$

Here (12) and (13) are obtained by standard matrix algebra together with the fact that ${X_o}^T X_o = I_p$ and ${Y_o}^T Y_o = I_q$.

In the special cases when one of the subspaces is one dimensional, the statistic Λ can be interpreted differently. When $q = 1$ we can calculate the angle θ between the vector Y_o and the X-subspace. It turns out that the relation between θ and Λ can be written as

$$\Lambda = \sin^2 \theta . \tag{14}$$

The statistic Λ described above is precisely Wilks' Λ-statistic which is a simple transformation of the likelihood ratio statistic. Following Mardia (1979, chap 5.3), the test statistic is written as

$$\Lambda = \frac{\left|S_{yy} - S_{yx} S_{xx}^{-1} S_{xy}\right|}{\left|S_{yy}\right|} = \left|I - S_{yy}^{-1} S_{yx} S_{xx}^{-1} S_{xy}\right| \tag{15}$$

where

$$\begin{bmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{bmatrix} = \frac{1}{n} \begin{bmatrix} {X_c}^T X_c & {X_c}^T Y_c \\ {Y_c}^T X_c & {Y_c}^T Y_c \end{bmatrix} . \tag{16}$$

From the relations (5)-(6) it is straight forward to verify that (15) and the expressions (11)-(13) are equivalent.

Table 1 Alternative defining matrices for the eigenvalues $\lambda_1, \ldots, \lambda_s$.

Defining matrix	ith eigenvalue
$\boldsymbol{H}(\boldsymbol{H}+\boldsymbol{E})^{-1}$ or $\boldsymbol{S}_{yy}^{-1}\boldsymbol{S}_{yx}\boldsymbol{S}_{xx}^{-1}\boldsymbol{S}_{xy}$ or $\boldsymbol{Y}_o{}^T\boldsymbol{X}_o\boldsymbol{X}_o{}^T\boldsymbol{Y}_o$	λ_i
$\boldsymbol{E}(\boldsymbol{H}+\boldsymbol{E})^{-1}$ or $\boldsymbol{I}_q - \boldsymbol{Y}_o{}^T\boldsymbol{X}_o\boldsymbol{X}_o{}^T\boldsymbol{Y}_o$	$1-\lambda_{q+1-i}$
$\boldsymbol{H}\boldsymbol{E}^{-1}$ or $\boldsymbol{Y}_o{}^T\boldsymbol{X}_o\boldsymbol{X}_o{}^T\boldsymbol{Y}_o\left(\boldsymbol{I}_q - \boldsymbol{Y}_o{}^T\boldsymbol{X}_o\boldsymbol{X}_o{}^T\boldsymbol{Y}_o\right)^{-1}$	$\lambda_i/(1-\lambda_i)$

Table 2 Four invariant test statistics and the corresponding block design criteria.

Test Statistic	Expression	Design Criteria
Wilks' Λ	$\prod_{i=1}^{s}(1-\lambda_i)$	D-optimality
Roy's Largest Root	λ_1	E-optimality
Hotelling-Lawley Trace Statistic	$\sum_{i=1}^{s}(\lambda_i/(1-\lambda_i))$	A-optimality
Pillay-Bartlett Trace Statistic	$\sum_{i=1}^{s}\lambda_i$	M-optimality

2.3 Other invariant statistics

It can be shown that the two invariance properties, (7) and (9), imply that any test statistic can be written as a function of the $s = \min(p, q)$ non-zero eigenvalues of $\boldsymbol{Y}_o{}^T\boldsymbol{X}_o\boldsymbol{X}_o{}^T\boldsymbol{Y}_o$. We denote these eigenvalues by $\lambda_1, \ldots, \lambda_s$ and Table 1 lists some alternative defining matrices. The λ_i's are also known as the squared canonical correlation coefficients (Gittins, 1985). The largest such coefficient, $\sqrt{\lambda_1}$, is the maximal correlation coefficient between a pair of linear combinations from the two centered data sets. In terms of canonical angles, $\arccos(\sqrt{\lambda_1})$ is the smallest angle between a pair of vectors from the two actual subspaces. The second correlation coefficient is obtained by restricting the vectors to be orthogonal to those from the first pair.

The statistic Λ (11) can be written as:

$$\Lambda = \prod_{i=1}^{s}(1-\lambda_i). \tag{17}$$

Olson (1976) reviews the four most popular test statistics. These are listed in Table 2. As a digression this table also lists four block design criteria (John and Williams 1982, 1995) that correspond to the four test statistics. We assume two categorical design variables. A treatment variable with t levels and a blocking variable with b levels. A model with only treatment effects can be written in a regression form with $(t-1)$ column vectors that are orthogonal to the constant vector. Similarly, a block effect model can be written with $(b-1)$ vectors. In an optimal design, the subspace spanned by the $(t-1)$ treatment vectors and the subspace spanned by the $(b-1)$

block vectors are "as orthogonal as possible". As a criterion for good design some measure of relation between the subspaces is needed. It is interesting to note that the block design criteria and the test statistics are equivalent.

To view the block design criteria more closely we consider a design where the t treatments are replicated r times and where each of the b blocks is of size $k = tr/b$. The design can be represented by $\boldsymbol{C}(tr \times t)$ and $\boldsymbol{D}(tr \times b)$ whose columns are the dummy variables (values 0 or 1) for the treatments and the blocks respectively. The matrix

$$\boldsymbol{N} = \{n_{ij}\} = \boldsymbol{C}^T \boldsymbol{D} \tag{18}$$

is the incidence matrix where n_{ij} represents the number of times that the ith treatment appears in the jth block. The different block design criteria are functions of the canonical efficiency factors, which are the non-zero eigenvalues of

$$\boldsymbol{A} = \boldsymbol{I}_t - (1/rk)\boldsymbol{N}\boldsymbol{N}^T . \tag{19}$$

We can rewrite $\boldsymbol{A}$ as

$$\boldsymbol{A} = \boldsymbol{I}_t - \left(\tfrac{1}{\sqrt{r}}\boldsymbol{C}\right)^T \left(\tfrac{1}{\sqrt{k}}\boldsymbol{D}\right) \left(\tfrac{1}{\sqrt{k}}\boldsymbol{D}\right)^T \left(\tfrac{1}{\sqrt{r}}\boldsymbol{C}\right) . \tag{20}$$

This expression for $\boldsymbol{A}$ is very similar to the matrices in (12)-(13) which are alternative defining matrices for the eigenvalues $\lambda_1, \ldots, \lambda_s$ (see Table 1). It follows directly that the different block design criteria, as expressed in John and Williams (1982), and the test statistics (Table 2) are equivalent. Note that, similarly to $\boldsymbol{X}_o$ and $\boldsymbol{Y}_o$, $\frac{1}{\sqrt{r}}\boldsymbol{C}$ and $\frac{1}{\sqrt{k}}\boldsymbol{D}$ have orthonormal columns. The only difference between (20) and the matrices in (12)-(13) is that the columns of $\frac{1}{\sqrt{r}}\boldsymbol{C}$ and $\frac{1}{\sqrt{k}}\boldsymbol{D}$ are not orthogonal to the constant vector. In fact, the constant vector is contained in the column space of both these matrices. The consequence is that $\boldsymbol{A}$ has as "an extra" eigenvalue which is zero.

2.4 The null distribution of the test statistics

Recall that all the invariant test statistics can be written as a function of the two orthogonal matrices, $\boldsymbol{X}_o$ and $\boldsymbol{Y}_o$. The invariance (9) follows from the simultaneous multinormal distribution of $\boldsymbol{x}$ and $\boldsymbol{y}$. Under the null hypothesis $\boldsymbol{x}$ and $\boldsymbol{y}$ is independently distributed. Therefore the invariance (9) can be written separately for $\boldsymbol{x}$ and for $\boldsymbol{y}$. In other words, $\boldsymbol{X}_o$ and $\boldsymbol{Y}_o$ are invariant under the transformation

$$(\boldsymbol{X}_o, \boldsymbol{Y}_o) \to (\boldsymbol{R}_1\boldsymbol{X}_o, \boldsymbol{R}_2\boldsymbol{Y}_o) \tag{21}$$

where $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are two arbitrary rotation matrices that perform rotations in S. From general theory of left invariant distributions (Dempster, 1969) it follows that the null distribution of the test statistics is uniquely characterised by this invariance property. The null distribution of a general test statistic, $G(\boldsymbol{X}_o, \boldsymbol{Y}_o)$, can be viewed as the distribution of

$$G(\boldsymbol{R}_1\boldsymbol{X}_o, \boldsymbol{R}_2\boldsymbol{Y}_o) \tag{22}$$

where $\boldsymbol{X}_o$ and $\boldsymbol{Y}_o$ are fixed and where $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ represent two independent random rotations in $\mathcal{S}$ (see Appendix). This distribution has an intuitive geometrical interpretation. The test statistic is a measure of relation between two subspaces (represented by $\boldsymbol{X}_o$ and $\boldsymbol{Y}_o$) and the null distribution follows from the relation between two randomly orientated subspaces (represented by $\boldsymbol{R}_1\boldsymbol{X}_o$ and $\boldsymbol{R}_2\boldsymbol{Y}_o$).

However, all invariant test statistics depend on $\boldsymbol{X}_o$ and $\boldsymbol{Y}_o$ through ${\boldsymbol{X}_o}^T\boldsymbol{Y}_o$. In other words all test statistics can be written as a function, $\tilde{G}({\boldsymbol{X}_o}^T\boldsymbol{Y}_o)$. Hence we can represent the null distribution as the distribution of

$$\tilde{G}\left({\boldsymbol{X}_o}^T{\boldsymbol{R}_1}^T\boldsymbol{R}_2\boldsymbol{Y}_o\right). \tag{23}$$

From the properties of random rotation matrices (see Appendix) it follows that the product matrix $\left({\boldsymbol{R}_1}^T\boldsymbol{R}_2\right)$ is also a matrix that performs a random rotation in $\mathcal{S}$. Moreover, for the product matrix to be random it is sufficient that one of the two matrices, $\boldsymbol{R}_1$ or $\boldsymbol{R}_2$, is random. It follows, as pointed out by Fisher (1939), that the assumption that $\boldsymbol{x}$ is normally distributed can be dropped. The null distribution will be the same as long as one of the sets is normally distributed. Geometrically this means that the relation between two subspaces is random as long as one of the two subspaces is randomly orientated.

The null distribution of Wilks' Λ-statistic is known as the Wilks' lambda distribution and has three parameters. These parameters correspond to the dimension of the "working space", $\mathcal{S}$ and the dimensions of the two subspaces. The notation for these parameters is, however, not standardised [see Mardia et al. (1979, chap. 3.7) and Kshirsagar (1972, chap. 8.2)].

2.5 Testing subhypotheses

The test described so far concerns overall relation between the two sets of variates, but in several situations one is often interested in testing subhypotheses regarding particular x-variables. Such tests can be described in a way that is very similar to the overall test. First, the matrix $\boldsymbol{X}$ is spilt into two parts $\boldsymbol{X}_1(n \times p_1)$ and $\boldsymbol{X}_2(n \times p_2)$ so that the last matrix consists of the variables for the hypothesis. The centering step, which is a projection onto $\mathcal{S}$, is now replaced by projection onto $\mathcal{S}^+$, which is the orthogonal complement of the space spanned by the columns of $\boldsymbol{X}_1$ together with the constant vector. We can express the projected data as

$$\boldsymbol{X}_c^+ = \boldsymbol{X}_2 - \boldsymbol{Z}\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{X}_2 \tag{24}$$

$$\boldsymbol{Y}_c^+ = \boldsymbol{Y} - \boldsymbol{Z}\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{Y} \tag{25}$$

where

$$Z = [\, \mathbf{1}_{n\times 1} \quad \boldsymbol{X}_1] \,. \tag{26}$$

To formulate the null distribution of the test statistic, the rotations in $\mathcal{S}$ are replaced by rotations in $\mathcal{S}^+$. Now, the "X-subspace" has dimension p_2 instead of p and the space "we are working in", $\mathcal{S}^+$, has dimension $(n-1-p_1)$ instead of $(n-1)$. Tests of hypotheses concerning a linear combination of x-variables can in principle be constructed in a similar way by reparameterising the x-variables. Also note that the vector $\mathbf{1}_{n\times 1}$ does not need to be a part of the matrix $\boldsymbol{Z}$.

3 Testing under other models

As mentioned in section 2.4 the assumption that $\boldsymbol{x}$ is normally distributed can be dropped. The sections 3.1-3.2 below treat the special case of non-normal $\boldsymbol{x}$ where $\boldsymbol{x}$ is fixed.

3.1 Multivariate multiple regression

Assume an ordinary multivariate multiple regression model defined by:

$$\boldsymbol{Y} = \mathbf{1}_{(n\times 1)}\boldsymbol{B}_0 + \boldsymbol{X}\boldsymbol{B} + \boldsymbol{F} \tag{27}$$

where $\boldsymbol{B}_0^T$ is the vector of intercept terms and $\boldsymbol{B}$ is the matrix of the other regression parameters. The matrix $\boldsymbol{F}$ consist of random disturbances whose rows are uncorrelated multinormal with zero mean and a common covariance matrix $\boldsymbol{\Sigma}$. Following Mardia (1979, chap. 6.3), Wilks' Λ for the hypothesis

$$\mathrm{H}_0: \quad \boldsymbol{B} = 0 \qquad \text{against} \qquad \mathrm{H}_1: \quad \boldsymbol{B} \neq 0 \tag{28}$$

is in this case written as

$$\Lambda = \frac{|\boldsymbol{E}|}{|\boldsymbol{H}+\boldsymbol{E}|} \tag{29}$$

where the matrices $\boldsymbol{E}$ and $\boldsymbol{H}$ can be defined from maximum likelihood estimates of $\boldsymbol{\Sigma}$ under H_0 and H_1:

$$\boldsymbol{E}+\boldsymbol{H} = n\widehat{\boldsymbol{\Sigma}}_{H0} = \boldsymbol{L}_Y^T\boldsymbol{L}_Y \tag{30}$$

$$\boldsymbol{E} = n\widehat{\boldsymbol{\Sigma}}_{H1} = \boldsymbol{L}_Y^T\left(\boldsymbol{I}_q - \boldsymbol{Y}_o{}^T\boldsymbol{X}_o\boldsymbol{X}_o{}^T\boldsymbol{Y}_o\right)\boldsymbol{L}_Y \,. \tag{31}$$

The expression (29) is just another way to write the test statistic (11) treated earlier. Since $\boldsymbol{Y}$ is multinormal the statistic will still follow the Wilks' lambda distribution. Also note that the expression (29) corresponds to the common definition of Wilks' lambda distribution based on Wishart distributions (Mardia 1979, chap. 3.7).

When we have one y-variable it is common to use the equivalent F statistic:

$$F = \tfrac{n-1-p}{p}\ \frac{1-\Lambda}{\Lambda}. \tag{32}$$

The relation to θ follows from (14):

$$F = \tfrac{n-1-p}{p}\ \frac{1}{\tan^2\theta}. \tag{33}$$

In this case Box and Draper (1987, fig. 3.4-3.5) present nice illustrations of the geometry. On their figures the significance level is a proportion of a spherical surface area.

The tests of subhypotheses under this regression model can be formulated as described in section 2.5.

3.2 Multivariate analysis of variance

All MANOVA models are in principle multivariate linear regression models. The model can always be written in a regression form where p equals the degrees of freedom for the model. The column vectors of $\boldsymbol{X}_0$ are then coefficient vectors for orthogonal contrasts. When comparing g groups in one-way MANOVA one possibility is to use the dummy variables (values 0 or 1) corresponding to the $(g-1)$ first groups as X-variables. In the one-way MANOVA case $\boldsymbol{E}$ is the matrix of "within sum of squares and cross products" and $\boldsymbol{H}$ is the matrix of "between sum of squares and cross-products" (Johnson and Wichern 1992, chap. 6.4). The two matrices are often denoted by "$\boldsymbol{W}$" and "$\boldsymbol{B}$".

When $g = 2$ it is common to use the equivalent Hotelling T^2 statistic:

$$T^2 = (n-2)\frac{1-\Lambda}{\Lambda}. \tag{34}$$

Now, $\boldsymbol{X}_o$ is a single vector. We can therefore refer to the angle θ between this vector and the y-space. The relation between T^2 and θ follows from (14):

$$T^2 = \frac{n-2}{\tan^2\theta}. \tag{35}$$

From this viewpoint it is obvious that the Hotelling T^2 and the F distribution are the same (except for a proportionality constant). In both cases we test whether there is any relation between a subspace and a vector. The difference is that in an F-test $\boldsymbol{Y}_o$ is the vector, but for the T^2-test here $\boldsymbol{X}_o$ is the vector.

3.3 The hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$

Now, only one set of variates is considered and multinormality is assumed:

$$\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) . \tag{36}$$

The likelihood ratio test for the hypothesis

$$\mathrm{H}_0 : \quad \boldsymbol{\mu} = \boldsymbol{\mu}_0 \qquad \text{against} \qquad \mathrm{H}_1 : \quad \boldsymbol{\mu} \neq \boldsymbol{\mu}_0 \tag{37}$$

is based on the Hotelling T^2 test statistic (Mardia 1979, chap. 5.2):

$$T^2 = (n-1)\left(\overline{\boldsymbol{Y}} - \boldsymbol{\mu}_0\right)^T \boldsymbol{S}^{-1}\left(\overline{\boldsymbol{Y}} - \boldsymbol{\mu}_0\right) \tag{38}$$

where

$$\overline{\boldsymbol{Y}} = \frac{1}{n}\mathbf{1}^T_{(n\times 1)}\boldsymbol{Y} \tag{39}$$

and where $\boldsymbol{S}$ is the maximum likelihood estimate of $\boldsymbol{\Sigma}$. This T^2 statistic can also be interpreted geometrically as Λ or θ:

$$T^2 = (n-1)\frac{1-\Lambda}{\Lambda} = \frac{n-1}{\tan^2\theta} . \tag{40}$$

The statistic Λ can be calculated by first subtracting the hypothesised mean $\boldsymbol{\mu}_0$ from all the original observations $\boldsymbol{Y}$. Since we are testing the mean, no centering is performed before the QR decomposition. The unit vector $\frac{1}{\sqrt{n}}\mathbf{1}_{(n\times 1)}$ will in this case function as the matrix $\boldsymbol{X}_o$. Then, θ is the angle between this vector and the space spanned by the "$\boldsymbol{\mu}_0$-subtracted" Y-data. A small θ indicates that $\boldsymbol{\mu}_0$ is not the true mean. The formulation of this test is based on random rotations that are not subspace restricted (8). We can say that our "working space" now is the whole n-dimensional space instead of the $(n-1)$-dimensional subspace $\mathcal{S}$. This explains why the proportionality constant in (40) differs from those in (34) and (35).

4 Concluding remarks

In this paper we have assumed that $\boldsymbol{Y}$ (or $\boldsymbol{X}$) follows a multinormal distribution. The results are, however, valid for any distributions of $\boldsymbol{Y}$ that are invariant under the transformation $\boldsymbol{Y} \to \boldsymbol{RY}$ ($\boldsymbol{R}$ represents a rotation in $\mathcal{S}$). This invariance together with the independence of the rows implies normality (James, 1954). Therefore, the practical relevance of the theory will be mostly for multinormal distributed data. However, Dawid (1977) mention that the generalised distributions "do arise naturally in certain applications". Theory for such generalised distributions and other related distributions are extensively described by Fang and Anderson (1990) and Fang and Zhang (1990).

APPENDIX: Random rotation matrices

A rotation matrix is an orthogonal matrix with determinant 1 (Curtis 1979). The determinant of an orthogonal matrix is always 1 or -1. However, in this paper we refer to all orthogonal matrices when we talk about rotation matrices.

The columns (and also the rows) of an orthogonal $n \times n$ matrix form a basis for the n-dimensional space. In a random rotation matrix, the first column is a random unit vector. Given the first column, the second column is a random unit vector orthogonal to the first. A simple way to construct a random rotation matrix is to fill a $n \times n$ matrix with n^2 independent standard normal deviates and then perform a QR decomposition. The resulting Q-matrix is then a random rotation matrix. Also note that the transpose (= inverse) matrix is also a random rotation matrix. Moreover, it is easy to see that an $n \times n$ random rotation matrix multiplied by another rotation matrix (fixed or independent random) is still a random rotation matrix. Note that a random rotation matrix is also known as a random "uniformly distributed" orthogonal matrix.

A matrix that perform random rotation in a $i < n$ dimensional subspace can be constructed as

$$\boldsymbol{R} = \boldsymbol{U}_1 \tilde{\boldsymbol{R}} \boldsymbol{U}_1{}^T + \boldsymbol{U}_2 \boldsymbol{U}_2{}^T \tag{41}$$

where $\boldsymbol{U}_1$ is a $n \times i$ matrix whose columns form an orthonormal basis for the subspace for rotation. Similarly $\boldsymbol{U}_2$ contains a basis for the complement. The $i \times i$ matrix $\tilde{\boldsymbol{R}}$ is an ordinary random rotation matrix. A class of matrices that perform rotation in a specific subspace is closed under the transpose operation and multiplication with matrices from the same class. This can be seen from the fact that these operations on "$\boldsymbol{R}$-matrices" can be reduced to similar operations on the "$\tilde{\boldsymbol{R}}$-matrices". The class of rotation matrices that performs rotation in the subspace $\mathcal{S}$, which is orthogonal to the constant vector, plays a key role in this paper. Also note that this class contains all permutation matrices.

REFERENCES

Box, G. E. P. and Draper, N. R. (1987), *Empirical Model-Building and Response Surfaces*, New York: John Wiley & Sons.

Curtis, M. L. (1979), *Matrix Groups*, New York: Springer-Verlag.

Dawid, A. P. (1977), Spherical Matrix Distributions and a Multivariate Model, *Journal of Royal Statistical Society, Series B* , **39**, 254–261.

Dempster, A. P. (1969), *Elements of Continuous Multivariate Analysis*, Reading, Mass. : Addison-Wesley.

Eaton, M. L. (1983), *Multivariate Statistics: A Vector Space Approach*, New York: John Wiley & Sons.

Fang, K. T. and Anderson T. W. (1990), *Statistical Inference in Elliptically Contoured and Related Distributions*, New York: Allerton Press.

Fang, K. T. and Zhang Y. T. (1990), *Generalized Multivariate Analysis*, New York: Springer-Verlag.

Fisher, R. A. (1939), The Sampling Distribution of some Statistics Obtained from Non-Linear Equations, *Annals of Eugenics*, **9**, 238–249.

Gittins, R. (1985), *Canonical Analysis: A Review with Applications in Ecology*, New York: Springer-Verlag.

James, A. T. (1954), Normal Multivariate Analysis and The Orthogonal Group, *The Annals of Mathematical Statistics*, **25**, 40–75.

John, J. A. and Williams, E. R. (1982), Conjectures for Optimal Block Designs, *Journal of Royal Statistical Society, Series B* , **44**, 221–225.

John, J. A. and Williams, E. R. (1995), *Cyclic and Computer Generated Designs, Sec. Ed.*, New York: Chapman and Hall.

Johnson, R. A. and Wichern, D. W. (1992), *Applied Multivariate Statistical Analysis, 3rd. ed.*, London: Prentice-Hall.

Kshirsagar, A. M. (1972), *Multivariate Analysis*, New York: Marcel Dekker.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979), *Multivariate Analysis*, London: Academic Press Limited.

Morrison, D. F. (1976), *Multivariate Statistical Methods, Sec. ed.*, New York: McGraw-Hill.

Olson, L. (1976), On Choosing a Test Statistic in Multivariate Analysis of Variance, *Psychological Bulletin*, **83**, 579–586.

Saville, D. J. and Wood, G. R. (1996), *Statistical Methods : A Geometric Primer*, New York: Springer-Verlag.

Strang, G. (1988), *Linear Algebra and its Applications, 3rd ed.*, San Diego: Harcourt Brace Javanovich.

Wickens, T. D. (1995), *The Geometry of Multivariate Statistics*, Hillsdale, New Jersey: Lawrence Erlbaum Associates.