GENERALIZED CANONICAL ANALYSIS

by

Carlos Agra Coelho

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
1992

Doctoral Committee:

Professor Anant M. Kshirsagar, Chairman
Professor Mark Becker
Professor Charles Kowalski
Professor M. Anthony Schork

To The True God. To my Parents. To the one I love.

To my true friends.

To everybody who made this possible to happen.

'Happy is the man that has found wisdom, and the man that gets discernment, for having it as gain is better than having silver as gain and having it as produce than gold itself. It is more precious than corals, and all other delights of yours cannot be made equal to it.'
– Pro. 3:13-15.

ACKNOWLEDGEMENTS

I would like to thank all the individual persons and the institutions who made this thesis possible to happen.

First of all I would like to thank my parents, who made me possible. I would like also to thank, among other institutions, the Luso-American Cultural Commission, the INVOTAN-JNICT (Comissão Permanente INVOTAN – JNICT), the Luso-American Foundation for Development (FLAD) and the Institute for International Education (IIE), institutions without whose support, both financial and moral, this thesis would have never been possible.

I would also like to thank my advisor, Professor Anant M. Kshirsagar, for all his guidance, effort, patience and time spent, as well as all the other members of my dissertation committee, Professors Mark Becker, Charles Kowalski and Anthony Schork.

I would like to thank the Executive and Scientific Boards of Instituto Superior de Agronomia (Agronomy College) and the Reitoria of the Lisbon Technical University (UTL) for granting me a leave from my teaching assignment. I would like also to thank Professor António St.Aubyn, head of the Mathematics, Statistics and Computer Science Department of Instituto Superior de Agronomia of UTL both for granting my leave and for being the one who introduced me to the areas of Multivariate Statistics and Canonical Analysis.

Finally, but not least, to all the true friends that always encouraged me in the moments of less motivation, or even despair, and that may have been more responsible than what they may think for this thesis to happen.

To these institutions and individual persons I will be forever grateful.

NOTATION

$\underline{x}_i \quad (i = 1, ..., n)$	—	the column vector representing the i^{th} row of the matrix X
$\underline{x}^j \quad (j = 1, ..., p)$	—	the column vector representing the j^{th} column of the matrix X
$\underline{x}_{(k)} \quad (k = 1, ..., m)$	—	a vector of p_k variables (usually with a p_k-variate normal distribution or a multinomial distribution). Not to be confused with $\underline{x}_i$. It will also be clear from the context.
$bdiag(Q_k)$	—	a block-diagonal matrix with typical block Q_k (usually with $k = 1, ..., m$; however, when the range of the index is not explicitly written, it may easily be understood from the context).
E_{nm}	—	denotes an $n \times m$ matrix of all 1's.
$vec(X)$	—	denotes a vector obtained by stacking the columns of X one under the other, from the first to the last.

All other notation is either clear or directly introduced in the text.

TABLE OF CONTENTS

LIST OF TABLES

Table

ABSTRACT

GENERALIZED CANONICAL ANALYSIS

by

Carlos Agra Coelho

Chairman: Anant M. Kshirsagar

Canonical Analysis is the statistical study of the relationships of two vector variables. In this dissertation we have generalized it to the simultaneous study of the relationships among more than two vectors. We call this extension of Canonical Analysis, Generalized Canonical Analysis (GCA).

Besides a detailed exploration of the geometrical properties of GCA, an explicit model and test statistics are developed and it is shown how such model is a generalization of well known univariate and multivariate statistical linear models and methods, such as Canonical Analysis, Multiple Regression Analysis, Univariate and Multivariate Analysis of Variance and Covariance, Discriminant Analysis, Correspondence Analysis and Principal Components Analysis. Other statistical models usually not treated in the literature may still be looked upon as particular cases of the GCA model. Some of these models, which we refer to as Block Regression Models, are treated in more detail. The test statistics developed for this unifying theory of GCA are straightforward and, when used in each of the particular cases,

reduce to the test statistics commonly used in the usual linear models. These statistics are generalizations statistics known in other situations.

The Maximum Likelihood Estimators (MLE's) for the mean vector and variance-covariance matrix (Σ) of a multivariate normal distribution, when different masses are assigned to each observation of a random sample, are obtained. It is shown that the distribution of the MLE of Σ is then approximately Wishart. Based on this result, the distributions of the submatrices, and meaningful functions of such submatrices, of the MLE of Σ are studied. The results obtained enabled us to investigate the distributions of the above test statistics and their relation to Wilks' lambda. This dissertation also gives an asymptotic normal distribution for the generalized Wilks' lambda.

The application of the GCA model and associated test statistics to the study of relationships among categorical variables is also considered and illustrated.

CHAPTER 1

INTRODUCTION

Every scientific activity is concerned with relationships among variables. Classical Canonical Correlation Analysis studies the relationship between two vector variables. However, in many practical situations one needs the analysis of not just two but several vector variables simultaneously. This dissertation is an attempt to extend the classical method to several vector variables. We shall call it Generalized Canonical Analysis (GCA). The statistical hypotheses and associated test criteria and their exact or approximate distributions are discussed in the dissertation. The Generalized Canonical Analysis may also be seen as a generalization of other well known linear statistical models. The GCA model and associated test statistics are derived first for continuous variables having a joint multivariate normal distribution. Later the GCA model is applied to the simultaneous analysis of categorical variables, leading to the Multiple Correspondence Analysis model or its generalization. Other models that are parts of the overall GCA model, and may be looked upon as a generalization of the Multiple Regression model, are also analyzed.

In Chapter 2 – Factorial Analysis – we present a detailed account, starting from the fundamentals, of the two dual factorial analyses (in the row or observations space and the column or variables space) of a data matrix, using general metrics in either space. While all concepts in this account may not be new and the account in some way parallels already known expositions (see for example Cailliez and Pagès (1976)), our exposition is more general than the ones usually found in the literature because it uses, in both spaces, general metrics with appropriate properties. This exposition is made more structured in order to present the main results in a most concise way to the readers, without sacrificing precision. Chapter 2 lays down the basis for a geometrical-algebraic introduction to the Generalized Canonical Analysis.

In Chapter 3 we introduce a geometrical-algebraic approach to the generalization of the usual Canonical Analysis to the simultaneous analysis of more than two sets of variables.

Since the pioneering results of Hotelling (1936) on canonical correlations and canonical variables, several efforts of a generalization to more than two sets of variables have been made. Prominent among these are the methods suggested by Carroll (1968), Kettenring (1971), Dauxois and Pousse (1976) and Geer (1984).

Generalizations of the usual Canonical Analysis are not the only methods to handle the simultaneous analysis of more than two sets of variables (Escoufier, 1973, 1980; Gower, 1975; L'Hermier des Plantes, 1976; Salle, 1979; Glaçon, 1981; Lavit and Roux, 1982; Escofier and Pagès, 1982, 1984; Lafosse, 1985). However, generalizations of the usual Canonical Analysis to more than two sets of variables may be very useful as they may not only work as unifying methods for a large number of well known linear statistical models, but at the same time may stimulate the development of other new models.

The generalization proposed by Carroll (1968) seems to be the best method that fulfills the goal of providing a basis for a general unifying method and model that could have as particular cases most of the well known linear models as well as other interesting and useful models not treated in the literature.

In Chapter 3, we bring together the general background provided by the Factorial Analysis described in Chapter 2, and Carroll's (1968) idea for a generalization of the usual Canonical Analysis, and are able to provide the reader with a condensed yet detailed overview of the main properties of such a generalization that we refer to as Generalized Canonical Analysis (GCA).

By the end of Chapter 3, we expect a reader, who is already familiar with the theory and methods of Linear Models and Canonical Analysis, to realize the usefulness of GCA as a method able to generalize a number of well known statistical linear methods and models. At this stage, the pertinent questions from a reader will possibly be about the hypotheses of interest and the availability of test statistics and an explicit form of the GCA model. Other questions that may be raised are: Does our GCA model reduce to the usual well known linear models when we restrict it in such a way that we would expect so to happen? In these cases, do the test statistics developed for the GCA model give rise to the test statistics commonly used in such particular models? Will the GCA model provide any guidelines in the search for new models, and will the test statistics used in GCA still be applicable? We shall attempt to provide answers to these questions in Chapter 4.

In Chapter 4 we derive the Maximum Likelihood Estimators (MLE's) for the mean vector and the variance-covariance matrix of a p-variate normal distribution (with mean $\underline{\mu}$ and variance-covariance matrix Σ) when masses m_i (with $\sum_{i=1}^{n} m_i = 1$) are assigned to the observations in a random sample of size n from the distribution. Let X $(n{\times}p)$ denote the data matrix of such a random sample, centered for the masses m_i, and $D = diag(m_i)$ $(i = 1, ..., n)$. Then, the MLE of Σ is

$$V = X'DX \ .$$

We show that the distribution of V is approximately Wishart.

Consider a partitioning of Σ and V corresponding to the partitioning of the p-variate vector $\underline{x}$ into m sets (with p_k variables in the k^{th} set of variables and $\sum_{k=1}^{m} p_k$ $= p$). In Chapter 4, we derive the distributions of the submatrices in V and some meaningful functions of them.

Corresponding to this partitioning of $\underline{x}$, Σ, and V, we may consider three types of nested null hypotheses:

H_0 : the m sets of variables are all independent

$H_0^{(k)}$: the k^{th} set of variables is independent of all the sets from $k+1$ to m; $(k = 1, ..., m-1)$

$H_0^{i(k)}$: the k^{th} and i^{th} sets are independent $(i > k)$, conditionally on all the sets $k+1$ through m (but not the i^{th} one); $(k = 1, ..., m-1)$ $(i = k+1, ..., m)$.

These hypotheses are presented in a more formal notation in Chapter 4, but even now it is possible to see that if H_0 holds, $H_0^{(k)}$ also holds for all k $(k = 1, ..., m-1)$ and that if $H_0^{(k)}$ holds then $H_0^{i(k)}$ holds for all i $(i = k+1, ..., m)$.

Also in Chapter 4, test statistics to test $H_0^{(k)}$ and $H_0^{i(k)}$ are derived and their distributions shown to be related to the Wilks' lambda distribution. Moreover, under H_0, the test statistics to test the $m-1$ hypotheses $H_0^{(k)}$ (for $k = 1, ..., m-1$) are shown to be independent, thus providing also a test for H_0.

Finally, an explicit form of the GCA model is suggested and through the associated discussion we try to make it clear how the GCA model is not only a generalization of well known linear statistical models, such as models in Canonical Analysis, Multiple Regression Analysis, Univariate and Multivariate Analysis of Variance and Covariance, Multiple Discriminant Analysis, Principal Components Analysis and Correspondence Analysis, but also the GCA model comprises submodels which are useful in their own right. These are further studied in Chapter 7.

The above hypothesis H_0 may be tested by testing the $m-1$ hypotheses $H_0^{(k)}$ $(k = 1, ..., m-1)$ sequentially. However, this may be a long and time consuming process. If one needs a quick assessment of the independence of the m sets of variables, then there is a test statistic that is a function of the likelihood ratio test statistic to test H_0. Under H_0, it is a product of $m-1$ independent Wilks' lambdas and therefore we shall refer to it as generalized Wilks' lambda. This generalized Wilks' lambda will enable us to test H_0 directly. The distribution of this test statistic is known to the extent that its moments are known and the statistic has a finite range, zero to one, but the exact distribution is too complicated to be useful in practice. Several asymptotic distributions have been studied; some of these are referred to in Chapter 5. All these expansions are in terms of chi-square or F distributed variables. A more recent one uses the Beta distribution. But these are all rather complicated and for this reason we develop an asymptotic normal approximation to the distribution of this generalized Wilks' lambda statistic, in Chapter 5. Incidently, in order to obtain the asymptotic normal expansion for the distribution of the generalized Wilks' lambda it is necessary to obtain a better asymptotic expansion to the distribution of $|V|$. This is also derived in Chapter 5 and provides another useful result in itself.

In Chapter 6, the GCA model is applied to the study of the relationships among m categorical variables, each represented by the set of indicator variables for its categories. It is shown that in this situation the GCA model is the Multiple Correspondence Analysis (MCA) model. All the main results and the test statistics derived for the GCA model are still applicable. If we assume a large enough sample size, the distribution of such test statistics may then be approximated either by the Wilks' lambda distribution or preferably by the chi-square distribution.

Taylor's blood serological data, which was originally analyzed by Fisher (1938) and later by Bartlett (1951), is then used to illustrate some of the steps in building and testing a GCA type model. This leads us to a slightly different type of model. This data set is used not because we think the MCA is the appropriate model to analyze it, but because, as Bartlett has remarked, it is a handy data set for illustrating this type of model building and testing procedures. A recent paper by Gower (1990) creates the impression that MCA is really a better technique to analyze these data. The results of our analysis, however, point towards the need for a different type of model. This different type of model is related to the Canonical Analysis model, but is one in which we would like to be able to test for the significance of one set of variables conditionally on holding the others in the model.

These models, which are a part of the overall GCA model, presented in Chapter 4, may be seen as particular cases of the GCA model, and this aspect is further discussed in Chapter 7. Further, we show in Chapter 7, how the test statistics for the hypotheses $H_0^{(k)}$ and $H_0^{i(k)}$ described in Chapter 4 can be used to test the overall fit of the model and the importance of a set of variables in the model.

Another way to look at the GCA is to see it as a direct generalization of the Principal Components Analysis model, where instead of dealing with individual variables we deal with sets of variables. This perspective is briefly introduced in the last Chapter, Chapter 8, along with several topics of future research arising out of this dissertation.

CHAPTER 2

FACTORIAL ANALYSIS

2.1 Introduction.

Factorial Analysis (not to be confused with 'Factor Analysis') is a method of analysis and representation that follows an essentially algebraic-geometrical approach, avoiding any *a priori* hypotheses. Factorial Analysis allows us to visualize data and obtain a simplified representation of such data in an intelligible reduced space, usually a plane defined by two factorial axes. This reduced space is defined in such a way that it presents some well defined optimal properties. The objective of this Chapter is to present the fundamentals of Factorial Analysis and the definition and most important optimal properties of the subspaces used for analysis and representation of the clouds of points.

2.2 Data and the two dual clouds of points.

Let us suppose that p real variables have been measured over a collection of n observed units. Our data can then be represented by the matrix X $(n \times p)$ with elements x_i^j, i.e.

$$X = \left[\begin{array}{c} x_i^j \end{array} \right] \qquad i = 1, ..., n; \ j = 1, ..., p$$

with x_i^j being the value that the j^{th} variable takes for the observation unit i. The i^{th} row of $\mathbf{X}$, is formed by the values of the p variables on the i^{th} unit and will be represented by $\underline{x}_i'$ $(i = 1, ..., n)$. The j^{th} column of $\mathbf{X}$, is formed by the values of the j^{th} variable over the set of n observation units and will be represented by $\underline{x}^j$ $(j = 1, ..., p)$.

We then have two "dual" clouds of points. If we call I the collection of the n observation units, we have the cloud of observation-points

$$N(I) = \left\{ (\underline{x}_i, m_i), \ \underline{x}_i \in \mathbb{E}^p, \ m_i > 0, \ \sum_{i=1}^{n} m_i = 1 \ \mid \ i \in I \right\} \quad , \qquad I = \{1, ..., n\}$$

where m_i $(i = 1, \ldots, n)$ are the masses given to each of the points $\underline{x}_i$. Since we are working with real variables, we have $\mathbb{E}^p \subseteq \mathbb{R}^p$.

If we call J the collection of the p variables, then the dual cloud of variable-points will be

$$N(J) = \{ (\underline{x}^j, l_j^2), \ \underline{x}^j \in \mathbb{E}^n \ \mid \ j \in J \} \quad , \qquad J = \{1, ..., p\}$$

where $\mathbb{E}^n \subseteq \mathbb{R}^n$ and l_j^2 are weights assigned to the variables from some *a priori* consideration.

Hereon we will represent the points in both clouds by the corresponding vectors, $\underline{x}_i$ and $\underline{x}^j$ respectively.

Since the results of the analysis on the two clouds of points are interconvertible, our analysis may be performed on either of them. First we will center our analysis on $N(I)$ only.

2.3 Definition of a distance and a metric in $\mathbb{E}^p$, the observations space.

First we decide the metric to be used in $\mathbb{E}^p$. This metric is a rule (law) in the observations space to measure distances between points. The metric matrix will be a symmetric positive definite matrix denoted by Q. From now on the metric matrix and the metric itself will be denoted by the same symbol. In general we can write the square of the distance between two observations $\underline{x}_i$ and $\underline{x}_{i'}$ as

$$\begin{aligned} d_Q^2(\underline{x}_i, \underline{x}_{i'}) &= \| \underline{x}_i - \underline{x}_{i'} \|_Q^2 = (\underline{x}_i - \underline{x}_{i'})' Q (\underline{x}_i - \underline{x}_{i'}) \\ &= \sum_{j=1}^{p} \sum_{j'=1}^{p} l_j \; l_{j'} \cos\theta_{jj'} (x_i^j - x_{i'}^j)(x_i^{j'} - x_{i'}^{j'}) \; , \end{aligned} \tag{2.1}$$

where Q is given by

$$Q = \left[\quad l_j \; l_{j'} \cos\theta_{jj'} \quad \right] \quad , \qquad j, j' = 1, ..., p$$

and $\theta_{jj'}$ is the angle between the variables j and j'.

From (2.1) it is clear that if the variables are orthogonal then

$$d_Q^2(\underline{x}_i, \underline{x}_{i'}) = \| \underline{x}_i - \underline{x}_{i'} \|_Q^2 = \sum_{j=1}^{p} l_j^2 (x_i^j - x_{i'}^j)^2 \; , \tag{2.2}$$

in which case $Q = diag(l_j^2)$ and the l_j^2 $(j = 1, ..., p)$ are the weights given to each variable as mentioned earlier. A metric of this type that is frequently used is the metric $Q = diag(1/s_j^2)$, s_j^2 being the sample variance of the j^{th} variable, so that the effect of the scale of measurement of variables is removed. However, if in (2.2) we have $l_j^2 = 1$ then

$$d_Q^2(\underline{x}_i, \underline{x}_{i'}) = \| \underline{x}_i - \underline{x}_{i'} \|_Q^2 = \sum_{j=1}^{p} (x_i^j - x_{i'}^j)^2 \tag{2.3}$$

where $Q = I_p$. In this case (Euclidean distance) each variable is given the same weight but the contribution of each variable to the sum of the square of all the distances will be proportional to their variances (see Appendix 2.A for proof).

2.4 Norm and scalar product in $\mathbb{E}^p$ associated with the metric Q.

Definition 1: The module of the scalar product in $\mathbb{E}^p$ will be defined as

$$\| \underline{z}_i \|_Q = |\underline{x}_i' Q \underline{u}| = |\underline{u}' Q \underline{x}_i| \ ,$$

where $\| \underline{z}_i \|_Q$ is the norm of the projection of the observation point $\underline{x}_i$ on the direction defined in $\mathbb{E}^p$ by the Q-unitary vector $\underline{u}$ (i.e. by the vector $\underline{u}$ such that $\underline{u}'Q\underline{u} = \| \underline{u} \|_Q^2 = 1$). □

Definition 2: $\underline{x}_i$ and $\underline{y} \in \mathbb{E}^p$ will be said to be Q-orthogonal if

$$\underline{x}_i' Q \underline{y} = \underline{y}' Q \underline{x}_i = 0 \ . \ \square$$

2.5 Inertia of the cloud of observation-points.

We consider three different situations.

a) Inertia in relation to a point in $\mathbb{E}^p$.

Definition 3: The inertia of the cloud $N(I)$ of observation points in relation to a given point $\underline{y} \in \mathbb{E}^p$ is defined as

$$\begin{aligned} In_{\underline{y}}(I) &= \sum_{i=1}^{n} m_i \| \underline{x}_i - \underline{y} \|_Q^2 \\ &= \sum_{i=1}^{n} m_i \ (\underline{x}_i - \underline{y})' Q (\underline{x}_i - \underline{y}) \\ &= \sum_{i=1}^{n} m_i \ d_Q^2(\underline{x}_i, \underline{y}) \ . \ \square \end{aligned}$$

Theorem 1: Huygens Theorem (1986)

Let $\underline{g} \in \mathbb{E}^p$ be the center of mass of the cloud $N(I)$, i.e.

$$\underline{g} = \sum_{i=1}^{n} m_i \, \underline{x}_i \quad \text{or} \quad \underline{g} = \begin{bmatrix} g_j \end{bmatrix}, \quad g_j = \sum_{i=1}^{n} m_i \, x_i^j \ ,$$

then

$$In_{\underline{y}}(I) = In_{\underline{g}}(I) + d_Q^2(\underline{y}, \underline{g})$$

or

$$min_{\underline{y} \in \mathbb{E}^p} \ In_{\underline{y}}(I) = In_{\underline{g}}(I) \ . \blacksquare$$

Since $In_{\underline{g}}(I)$ is a characteristic value of $N(I)$, it may be considered as a measure of dispersion of the cloud of points $N(I)$, allowing us to confine the study of the inertia of a cloud of points, in relation to any point, to the study of its inertia in relation to the center of mass.

b) Inertia in relation to the origin ($\underline{O} \in \mathbb{E}^p$) and total inertia.

Definition 4: The inertia of the cloud $N(I)$ in relation to the origin $\underline{O} \in \mathbb{E}^p$ is

$$\begin{aligned} In_{\underline{O}}(I) &= \textstyle\sum_{i=1}^{n} \ m_i \parallel \underline{x}_i \parallel_Q^2 = \sum_{i=1}^{n} \ m_i \ \underline{x}_i' Q \underline{x}_i = \sum_{i=1}^{n} m_i X_i Q X_i' = \\ &= tr(XQX'D) = tr(WD) = tr(X'DXQ) = tr(VQ) \ , \end{aligned} \tag{2.4}$$

where $D = diag(m_i)$, $W = XQX'$, $V = X'DX$ and X_i represents the i^{th} row of $\mathbf{X}$. $\square$

If the data matrix $\mathbf{X}$ had been centered for the masses m_i, then $\underline{O} \equiv \underline{g}$ and in this case $In_{\underline{O}}(I) = In_{\underline{g}}(I)$ is also called total inertia of the cloud of points. Then (see Appendix 2.B for proof)

$$In_{\underline{O}}(I) \ = \ \frac{1}{2} \sum_{i=1}^{n} \sum_{i'=1}^{n} \ m_i \ m_{i'} \ d_Q^2(\underline{x}_i, \underline{x}_{i'}) \ .$$

Without loss of generality we will assume hereonwards that the matrix $\mathbf{X}$ is centered for the masses m_i, which means that that the center of the mass is taken to be the origin itself.

c) Inertia of the cloud $N(I)$ in relation to a given direction in $\mathbb{E}^p$.

Definition 5: The inertia of the cloud $N(I)$ in relation to a given direction in $\mathbb{E}^p$, defined by $\underline{u}$, such that $\underline{u}'Q\underline{u} = \| \underline{u} \|_Q^2 = 1$, that passes through $\underline{y} \in \mathbb{E}^p$ is

$$In_{\underline{u}}(I) = \sum_{i=1}^{n} m_i \| \underline{z}_i - \underline{y} \|_Q^2 \ , \tag{2.5}$$

where $\underline{z}_i \in \mathbb{E}^p$ is the projection of $\underline{x}_i$ on the direction of $\underline{u}$. This is called inertia along the direction $\underline{u}$.

We can also define the inertia around the direction $\underline{u}$ as

$$In_{\underline{u}^\perp}(I) = \sum_{i=1}^{n} m_i \| \underline{x}_i - \underline{z}_i \|_Q^2 = In_{\underline{y}}(I) - In_{\underline{u}}(I) \ . \ \Box \tag{2.6}$$

We easily see that $In_{\underline{y}}(I)$ being fixed for a given cloud $N(I)$ and a given point $\underline{y} \in \mathbb{E}^p$, maximizing one of the inertias $In_{\underline{u}}$ or $In_{\underline{u}^\perp}$ is equivalent to minimizing the other. We usually want to minimize $In_{\underline{u}^\perp}(I)$, which is the same as maximizing $In_{\underline{u}}(I)$.

If the point $\underline{y} \in \mathbb{E}^p$ is the origin, which is also the center of mass, then, from (2.5)

$$In_{\underline{u}}(I) = \sum_{i=1}^{n} m_i \| \underline{z}_i - \underline{O} \|_Q^2 = \sum_{i=1}^{n} m_i \| \underline{z}_i \|_Q^2 \ .$$

But then using Definition 1 we have

$$\begin{aligned} In_{\underline{u}}(I) &= \sum_{i=1}^{n} m_i \left(|\underline{x}_i' Q \underline{u}| \right)^2 \\ &= \sum_{i=1}^{n} (\underline{u}' Q \underline{x}_i) m_i (\underline{x}_i' Q \underline{u}) \\ &= \underline{u}' Q X' D X Q \underline{u} \\ &= \underline{u}' Q V Q \underline{u} \ . \end{aligned}$$

2.6 Factorial axes and subspace of $\mathbb{E}^p$ closest to the cloud of observation-points.

Definition 6: The first factorial axis is defined as the direction $\Delta\underline{u}_1$ in $\mathbb{E}^p$, where $\underline{u}_1$ is the Q-unitary vector in $\mathbb{E}^p$ that maximizes

$$In_{\underline{u}}(I) = \underline{u}'QVQ\underline{u} \ . \ \square$$

As we mentioned earlier, determining $\underline{u}_1$ that maximizes $In_{\underline{u}}(I)$ is equivalent to determining $\underline{u}_1$ that minimizes $In_{\underline{u}^\perp}(I)$. From (2.6) we see that the direction $\Delta\underline{u}_1$ is the direction in $\mathbb{E}^p$ that is closest to the cloud of points $\underline{x}_i$, as it minimizes the weighted average (with weights m_i) of the squared distances between the points $\underline{x}_i$ and their corresponding projections over $\Delta\underline{u}_1$.

Lemma 1: The vector $\underline{u}_1$ is the eigenvector of VQ associated with its largest eigenvalue λ_1.

Proof: The problem of finding $\underline{u}_1$ is the problem of finding the maximum of the quadratic form $In_{\underline{u}}(I)$. Using Lagrange multipliers we equate the derivative of

$$L_1 = \underline{u}_1'QVQ\underline{u}_1 - \lambda_1(\underline{u}_1'Q\underline{u}_1 - 1)$$

to zero and obtain

$$QVQ\underline{u}_1 = \lambda_1 Q\underline{u}_1 \ .$$

Premultiplying by Q^{-1},

$$VQ\underline{u}_1 = \lambda_1\underline{u}_1 \ . \ \blacksquare$$

The following Corollary is then obvious.

Corollary 1: The maximum of the quadratic form $In_{\underline{u}}(I)$ is λ_1. $\blacksquare$

Definition 7: The second factorial axis is the direction $\Delta \underline{u}_2$ in $\mathbb{E}^p$, orthogonal to $\Delta \underline{u}_1$, defined by $\underline{u}_2$, the unitary vector in $\mathbb{E}^p$ that maximizes

$$In_{\underline{u}}(I) - In_{\underline{u}_1}(I) = \underline{u}'QVQ\underline{u} - \underline{u}_1'QVQ\underline{u}_1$$

subject to the restrictions

$$\| \underline{u}_2 \|_Q^2 = \underline{u}_2'Q\underline{u}_2 = 1 \quad \text{and} \quad \underline{u}_2'Q\underline{u}_1 = \underline{u}_1'Q\underline{u}_2 = 0 \ . \ \square$$

Lemma 2: The vector $\underline{u}_2$ is the eigenvector of VQ associated with the second largest eigenvalue of VQ, say λ_2.

Proof: Again using Lagrange multipliers, and equating the derivative of

$$L_2 = \underline{u}_2'QVQ\underline{u}_2 - \lambda_2(\underline{u}_2'Q\underline{u}_2 - 1) - \mu(\underline{u}_2'Q\underline{u}_1)$$

to zero, we obtain

$$QVQ\underline{u}_2 = \lambda_2 Q\underline{u}_2 + (1/2)\mu Q\underline{u}_1 \ .$$

Then premultiplying by $\underline{u}_1'$ gives $\mu = 0$, and then

$$QVQ\underline{u}_2 = \lambda_2 Q\underline{u}_2$$

or, premultiplying by Q^{-1},

$$VQ\underline{u}_2 = \lambda_2 \, \underline{u}_2 \ . \ \blacksquare$$

Corollary 2: As a consequence of Lemmas 1 and 2 and by induction, the subspace of $\mathbb{E}^p$ of dimension $q \leq p$ that lies closest to the cloud of points $N(I)$ is the subspace $\mathcal{F}^q$ generated by $\underline{u}_1, \ldots, \underline{u}_q$, eigenvectors of VQ associated with the eigenvalues $\lambda_1 > \ldots > \lambda_q > 0$, and then we can write

$$In_{\underline{u}_\alpha}(I) = \underline{u}_\alpha'QVQ\underline{u}_\alpha = \lambda_\alpha \qquad (\alpha = 1, \ldots, q)$$

where the vectors $\underline{u}_\alpha$ and the values λ_α are defined by

$$VQ\underline{u}_\alpha = \lambda_\alpha \underline{u}_\alpha \qquad (\alpha = 1, \ldots, q) \ . \ \blacksquare \tag{2.7}$$

Definition 8: The inertia explained by the subspace $\mathcal{F}^q$ is

$$In_{(\underline{u}_1,\ldots,\underline{u}_q)}(I) = \sum_{i=1}^{n} m_i \parallel \underline{z}_i \parallel_Q^2 \tag{2.8}$$

where now $\underline{z}_i$ is the projection of $\underline{x}_i$ on the subspace $\mathcal{F}^q$. □

Lemma 3: The inertia explained by the subspace $\mathcal{F}^q$ is

$$In_{(\underline{u}_1,\ldots,\underline{u}_q)}(I) = \sum_{\alpha=1}^{q} In_{\underline{u}_\alpha}(I) = \sum_{\alpha=1}^{q} \lambda_\alpha \ , \tag{2.9}$$

so that $\mathcal{F}^q$ is the subspace of $\mathbb{E}^p$ for which $In_{(\underline{u}_1,\ldots,\underline{u}_q)}(I)$ is maximum.

Proof: Using the Pythagorean Theorem, observing that $\underline{u}_\alpha$ $(\alpha=1,\ldots,q)$ are orthogonal, we obtain

$$\parallel \underline{z}_i \parallel_Q^2 = \sum_{\alpha=1}^{q} (\underline{x}_i' Q \underline{u}_\alpha)^2 \ .$$

Then, from (2.8) above,

$$\begin{aligned} In_{(\underline{u}_1,\ldots,\underline{u}_q)}(I) &= \sum_{i=1}^{n} m_i \sum_{\alpha=1}^{q} (\underline{x}_i' Q \underline{u}_\alpha)^2 = \sum_{i=1}^{n} \sum_{\alpha=1}^{q} (\underline{u}_\alpha' Q \underline{x}_i) m_i (\underline{x}_i' Q \underline{u}_\alpha) \\ &= \sum_{\alpha=1}^{q} \underline{u}_\alpha' Q V Q \underline{u}_\alpha = \sum_{\alpha=1}^{q} \lambda_\alpha = \sum_{\alpha=1}^{q} In_{\underline{u}_\alpha}(I) \ . \end{aligned}$$

The subspace $\mathcal{F}^q$ is the subspace of $\mathbb{E}^p$ for which $In_{(\underline{u}_1,\ldots,\underline{u}_q)}(I)$ is maximum because $\underline{u}_1,\ldots,\underline{u}_q$ were chosen to maximize each of the inertias $In_{\underline{u}_\alpha}(I)$ $(\alpha = 1,\ldots,q)$. ∎

Lemma 4: The inertia of the cloud of points $N(I)$ around the subspace $\mathcal{F}^q$ of $\mathbb{E}^p$ is given by

$$In_{(\underline{u}_1,\ldots,\underline{u}_q)^\perp}(I) = In_{\underline{O}}(I) - In_{(\underline{u}_1,\ldots,\underline{u}_q)}(I) = \sum_{i=1}^{n} \ m_i \parallel \underline{x}_i - \underline{z}_i \parallel_Q^2 \ . \tag{2.10}$$

where as in Definition 8 and Lemma 3, $\underline{z}_i$ is the projection of $\underline{x}_i$ on the subspace $\mathcal{F}^q$.

Proof: We can say that the result in this Lemma is 'intuitive' given the orthogonality of the q axes $\Delta\underline{u}_\alpha$ $(\alpha = 1, ..., q)$ and the definitions of $In_{\underline{O}}(I)$ in (2.4) and $In_{\underline{u}^\perp}(I)$ in (2.6). But if a core proof is necessary, then we may write

$$\parallel \underline{x}_i - \underline{z}_i \parallel_Q^2 \ = \parallel \underline{x}_i \parallel_Q^2 \ + \ \parallel \underline{z}_i \parallel_Q^2 \ - 2 \ \underline{x}_i' Q \underline{z}_i$$

where

$$\underline{z}_i = \sum_{\alpha=1}^{q} \underline{z}_i^\alpha$$

with

$$\underline{z}_i^\alpha = |\underline{x}_i' Q \underline{u}_\alpha| \ \underline{u}_\alpha$$

being the projection of $\underline{x}_i$ on the axis $\Delta \underline{u}_\alpha$. But then we may write

$$\underline{z}_i = \sum_{\alpha=1}^{q} |\underline{x}_i' Q \underline{u}_\alpha| \ \underline{u}_\alpha$$

so that then

$$\underline{x}_i' Q \underline{z}_i = \sum_{\alpha=1}^{q} (\underline{x}_i' Q \underline{u}_\alpha)^2 = \| \ \underline{z}_i \ \|_Q^2$$

or then

$$\| \ \underline{x}_i - \underline{z}_i \ \|_Q^2 = \| \ \underline{x}_i \ \|_Q^2 \ - \| \ \underline{z}_i \ \|_Q^2 \ .$$

Therefore by the definition of $In_{(\underline{u}_1,\ldots,\underline{u}_q)^\perp}(I)$

$$\begin{aligned} In_{(\underline{u}_1,\ldots,\underline{u}_q)^\perp}(I) &= In_{\underline{O}}(I) - In_{(\underline{u}_1,\ldots,\underline{u}_q)}(I) \\ &= \sum_{i=1}^{n} m_i \ \| \ \underline{x}_i \ \|_Q^2 \ - \sum_{i=1}^{n} m_i \ \| \ \underline{z}_i \ \|_Q^2 \\ &= \sum_{i=1}^{n} m_i \ \| \ \underline{x}_i - \underline{z}_i \ \|_Q^2 \ . \ \blacksquare \end{aligned}$$

<u>Lemma 5</u>: $\mathcal{F}^q$ is the subspace of $\mathbb{E}^p$ that minimizes

$$In_{(\underline{u}_1,\ldots,\underline{u}_q)^\perp}(I) \ .$$

Proof: From (2.9) in Lemma 3 and (2.10) above we can write

$$In_{(\underline{u}_1,\ldots,\underline{u}_q)^\perp}(I) = In_{\underline{O}}(I) - In_{(\underline{u}_1,\ldots,\underline{u}_q)}(I) = In_{\underline{O}}(I) - \sum_{\alpha=1}^{q} In_{\underline{u}_\alpha}(I) \ ,$$

where it has been shown in Theorem 1 and Lemma 3 that $In_{\underline{O}}(I)$ is a fixed quantity of the cloud of points and each $In_{\underline{u}_\alpha}(I)$, $(\alpha = 1, \ldots, q)$, has been maximized. $\blacksquare$

Definition 9: The percentage of inertia explained by $\mathcal{F}^q$ is

$$\frac{In_{(\underline{u}_1,\ldots,\underline{u}_q)}(I)}{In_{\underline{Q}}(I)} = \sum_{\alpha=1}^{q} \frac{In_{\underline{u}_\alpha}(I)}{In_{\underline{Q}}(I)} \ ,$$

where

$$\frac{In_{\underline{u}_\alpha}(I)}{In_{\underline{Q}}(I)}$$

is the percentage of inertia explained by the axis $\Delta\underline{u}_\alpha$ $(\alpha = 1, ..., q)$, and where, as shown before, $In_{(\underline{u}_1,\ldots,\underline{u}_q)}(I) = \sum_{\alpha=1}^{q} In_{\underline{u}_\alpha}(I)$. □

These percentages of inertia are a measure of the importance of the subspace, or the axes $\Delta\underline{u}_1, \ldots, \Delta\underline{u}_q$.

Lemma 6:

$$In_{\underline{Q}}(I) = tr(VQ) = \sum_{\alpha=1}^{p} \lambda_\alpha \ ,$$

so that the percentage of inertia explained by the axis $\Delta\underline{u}_\alpha$ can be written

$$\frac{\lambda_\alpha}{tr(VQ)} = \frac{\lambda_\alpha}{\sum_{\alpha=1}^{p} \lambda_\alpha} \ ,$$

and the percentage of inertia explained by the subspace $\mathcal{F}^q$ can be written

$$\frac{In_{(\underline{u}_1,\ldots,\underline{u}_q)}(I)}{In_{\underline{Q}}(I)} = \frac{\sum_{\alpha=1}^{q} \lambda_\alpha}{\sum_{\alpha=1}^{p} \lambda_\alpha} \ .$$

Proof: It is known from (2.4) that

$$In_{\underline{Q}}(I) = tr(VQ)$$

and it is a known result that

$$tr(VQ) = \sum_{\alpha=1}^{p} \lambda_\alpha \ .$$

But, a more detailed proof follows from (2.4), because we can write (see Appendix 2.C, (2.C.4))

$$\| \underline{x}_i \|_Q^2 = \sum_{\alpha=1}^{p} \| \underline{z}_i^\alpha \|_Q^2$$

with

$$\| \underline{z}_i^\alpha \|_Q = |\underline{x}_i' Q \underline{u}_\alpha| \ .$$

Then

$$\begin{aligned} In_{\underline{Q}}(I) = tr(VQ) &= \sum_{i=1}^{n} m_i \sum_{\alpha=1}^{p} \| \underline{z}_i^\alpha \|_Q^2 \\ &= \sum_{i=1}^{n} \sum_{\alpha=1}^{p} m_i \| \underline{z}_i^\alpha \|_Q^2 \\ &= \sum_{\alpha=1}^{p} \underline{u}_\alpha' QVQ \underline{u}_\alpha = \sum_{\alpha=1}^{p} In_{\underline{u}_\alpha}(I) = \sum_{\alpha=1}^{p} \lambda_\alpha \ . \ \blacksquare \end{aligned}$$

2.7 Another look at the Factorial Analysis.

Traditionally in the Factorial Analysis problem we first want to find a 'new' variable – a linear combination of the original variables – with maximum variance (subject to a fixed length). Then we desire a second variable, also a linear combination of the original variables, not correlated with the first, with maximum variance; and so on. This was Pearson's (1901) original approach. Since no distributional assumptions have been made, we really should replace, in the above, 'not correlated' by 'orthogonal' and 'variance' by 'inertia'.

We recall that our initial objective was to maximize

$$In_{\underline{u}_\alpha}(I) = \underline{u}_\alpha' QVQ \underline{u}_\alpha \ \text{ subject to } \ \underline{u}_\alpha' Q \underline{u}_\alpha = 1 \ \text{ and } \ \underline{u}_\alpha' Q \underline{u}_{\alpha-i} = 0 \ (i = 1, ..., \alpha - 1) \ .$$

But then if we let $\underline{a}^\alpha = Q\underline{u}_\alpha$, we will have $\underline{a}^\alpha \in I\!E^{p*}$, where $I\!E^{p*}$ is the dual space of $I\!E^p$, namely the space of linear combinations, or linear forms. Then our objective will be the maximization of

$$In_{\underline{u}_\alpha}(I) = (\underline{a}^\alpha)' V \underline{a}^\alpha \ \text{ subject to } \ (\underline{a}^\alpha)' Q^{-1} \underline{a}^\alpha = 1 \ \text{ and } \ (\underline{a}^\alpha)' Q^{-1} \underline{a}^{\alpha-i} \quad (i = 1, ..., \alpha - 1) \ .$$

The maximized quantity, $(\underline{a}^\alpha)' V \underline{a}^\alpha$, is itself the inertia of $\underline{a}^\alpha$ (V being the inertia matrix). Moreover, the vectors $\underline{a}^\alpha \in I\!E^{p*}$ are Q^{-1}-unitary and Q^{-1}-orthogonal to all other linear combinations of the original variables of similar type, norm and orthogonality that are defined in $\mathbb{E}^{p*}$ with respect to the metric Q^{-1}. Thus, $\underline{a}^\alpha$ $(\alpha = 1, ..., q)$ are, for each order α, the linear combinations we are looking for in order to fulfill our objectives. We will call $\underline{a}^\alpha$ as the 'factor' of order α.

It is easy to see that while the vectors $\underline{u}_\alpha$ were the eigenvectors of VQ associated with the eigenvalues λ_α, the $\underline{a}^\alpha$ are the eigenvectors of QV associated with the same eigenvalues.

Now we may note that maximizing

$$(\underline{a}^\alpha)'V\underline{a}^\alpha \ \text{ subject to } \ (\underline{a}^\alpha)'Q^{-1}\underline{a}^\alpha = 1 \qquad (\alpha = 1, \ldots, q)$$

is equivalent to maximizing

$$\frac{(\underline{a}^\alpha)'V\underline{a}^\alpha}{(\underline{a}^\alpha)'Q^{-1}\underline{a}^\alpha} \ . \tag{2.11}$$

The vector $\underline{a}^1$ that maximizes (2.11), is the eigenvector of QV associated with the first (largest) eigenvalue of QV, similarly the second eigenvector, $\underline{a}^2$, is the vector for which (2.11) attains its second maximum, under the restriction $(\underline{a}^2)'Q^{-1}\underline{a}^1 = 0$, and so on, for $\underline{a}^1, ..., \underline{a}^q$, eigenvectors of QV associated with the eigenvalues λ_α. Then it is easy to see that $(\underline{a}^\alpha)'V\underline{a}^\alpha = \lambda_\alpha$ and

$$\frac{(\underline{a}^\alpha)'V\underline{a}^\alpha}{(\underline{a}^\alpha)'Q^{-1}\underline{a}^\alpha} = \lambda_\alpha \qquad (\alpha = 1, \ldots, q) \ .$$

Note that if $Q^{-1} = V$ then all λ_α are equal to 1 as $QV = I_p$.

Factorial Analysis and all other such methods in the Factorial Analysis approach may be seen as methods of comparison of two metrics, V and Q^{-1}, over the space $\mathbb{E}^{p*}$, or as the comparison of the metrics Q and V^{-1} over $\mathbb{E}^p$. The deviations of the eigenvalues λ_α of VQ or QV from 1 are measures of the deviation of Q^{-1} from V. These help us to understand to what extent Q and V^{-1} shape the cloud $N(I)$ in $\mathbb{E}^p$, or generate the distances in $N(I)$, in a similar or a different way. In particular, when Q and V^{-1} are identical then, as observed above, all the eigenvalues of VQ are unity and $N(I)$ is then a sphere as the variation is the same in any direction or is isotropic.

Thus, Factorial Analysis and the several Factorial Analysis Methods allow us to detect in 'which way' we deviate from the hypothesis of equality of Q and V^{-1}, and which are the directions of the space that best provide evidence of such deviation.

In order to make inference and be able to test hypotheses, we need to assume some statistical distribution for our variables.

Some aspects of the parametric approach, mainly related with tests of hypotheses of independence among sets of variables and the distribution of the corresponding test statistics will be dealt with later.

2.8 Factorial variables.

Since the q vectors $\underline{u}_\alpha(\alpha = 1, \ldots, q \leq p)$ form a basis for $\mathcal{F}^q \subseteq \mathbb{E}^p$ and Q is the metric used in $\mathbb{E}^p$, the representation of the n observation points in the new basis $\underline{u}_1, \ldots, \underline{u}_q$ will be given by the rows of

$$C = XQU \ , \tag{2.12}$$

where

$$U = \left[\ \underline{u}_1 \mid \ldots \mid \underline{u}_\alpha \mid \ldots \mid \underline{u}_q \ \right] \qquad (q \leq p) \ .$$

The i^{th} row of C is the vector

$$\underline{c}_i' = (U'Q\underline{x}_i)' \qquad (i = 1, ..., n) \ . \tag{2.13}$$

It is the representation or Q-orthogonal projection of $\underline{x}_i$ on $\mathcal{F}^q$, a subspace of $\mathbb{E}^p$ of (reduced) dimension $q(\leq p)$, subspace generated by the columns of U. The vectors $\underline{c}_i$ will give 'the best' representation of $\underline{x}_i$ in a subspace of $\mathbb{E}^p$ of dimension $q(\leq p)$ in the sense that $In_{(\underline{u}_1,\ldots,\underline{u}_q)^\perp}(I) = \sum_{i=1}^n m_i \parallel \underline{x}_i - \underline{c}_i \parallel_Q^2$ is minimum. The vectors $\underline{c}_i$ $(i = 1, ..., n)$ give the best q-dimensional representation of the cloud of points

$N(I)$ because the original distances among the points $\underline{x}_i$ are best preserved and least deformed.

The columns of the matrix C are the vectors

$$\underline{c}^\alpha = XQ\underline{u}_\alpha \quad \in \quad \mathbb{E}^n \ , \tag{2.14}$$

each one of which has the coordinates of all the n observations on the new factorial axis of order α. Each vector $\underline{c}^\alpha$ $(\alpha = 1, \ldots, q)$ represents a 'new' variable, that we will call 'factorial variable' or 'principal variable'.

2.9 The dual analysis. The variables space.

To the j^{th} variable we may associate the vector $\underline{x}^j \in \mathbb{E}^n$, where

$$\underline{x}^j = \sum_{i=1}^{n} x_i^j \underline{f}^i$$

with $(\underline{f}^1, ..., \underline{f}^n)$ as the canonical basis of $\mathbb{E}^n$, the variables space, and x_i^j the coordinate of the vector-variable $\underline{x}^j$ relative to $\underline{f}^i$.

As seen in 2.2 above we then have the cloud $N(J)$ of variable-points. To each one of these points (assuming that the matrix Q, defined in 2.3, is diagonal with elements l_j^2), we shall assign a positive weight l_j^2. The cloud will then be represented as

$$N(J) = \left\{(\underline{x}^j, l_j^2),\ \underline{x}^j \in \mathbb{E}^n \mid j \in J\right\} \quad , \quad J = \{1, ..., p\} \ .$$

We should notice that this cloud $N(J)$ is in general not centered even if the cloud $N(I)$ has been centered.

In the more general case where the matrix Q is not diagonal the masses l_j^2 associated in this space with each variable-point $\underline{x}^j$ will not be the diagonal entries in Q but rather 'compound' weights based on all the entries in a row of Q.

2.10 Definition of a metric and a distance in $\mathbb{E}^n$.

When we measure the distance between two variable-points in the cloud $N(J)$ it makes sense to give each observation-axis a weight m_i in the computation of such a distance, as each observation-point in $N(I)$ had received a mass m_i. Then we may write the squared distance between any two vector variables $\underline{x}^j$ and $\underline{x}^{j'}$, $j, j' \in \{1, ..., p\}$ as

$$d^2(\underline{x}^j, \underline{x}^{j'}) = \sum_{i=1}^{n} m_i (x_i^j - x_i^{j'})^2 \ ,$$

or,

$$d_D^2(\underline{x}^j, \underline{x}^{j'}) = (\underline{x}^j - \underline{x}^{j'})' D (\underline{x}^j - \underline{x}^{j'}) \tag{2.15}$$

where $D = diag(m_i)$ is the metric.

This is the main reason why the two analyses, namely the analysis of the cloud $N(I)$ in $\mathbb{E}^p$ and the analysis of $N(J)$ in $\mathbb{E}^n$ are usually called dual analyses of each other. The metric in one analysis is the matrix of weights in the other analysis and vice-versa.

We will now show how the goals of the Factorial Analysis may also be stated in the space $\mathbb{E}^n$ leading to results that are convertible to the results obtained in $\mathbb{E}^p$. In practice we will only carry one of these two analysis since the results of the other analysis may always be then obtained through the transformation formulas as we will see later.

2.11 Norm and scalar product in $\mathbb{E}^n$.

Definition 10: The norm of $\underline{x}^j$ is

$$\| \underline{x}^j \|_D^2 = d_D^2(\underline{x}^j, \underline{O}) = (\underline{x}^j - \underline{O})' D (\underline{x}^j - \underline{O}) = (\underline{x}^j)' D \underline{x}^j \ . \ \square \tag{2.16}$$

As said in 2.5, we are using the variables centered for the masses m_i and since, from (2.16),

$$\| \underline{x}^j \|_D^2 = (\underline{x}^j)' D \underline{x}^j = \sum_{i=1}^{n} m_i (x_i^j)^2 ,$$

we really see that the squared norm of each variable-point $\underline{x}^j$ is numerically equivalent to the sample variance of the j^{th} variable (where variance is defined in a more general sense then the usual one, that is, referring to the use of the masses m_i).

The distance between any two variable vectors $\underline{x}^j$ and $\underline{x}^{j'}$, $(j, j' \in \{1, ..., p\})$ (from (2.15)) may also be written as

$$d_D^2(\underline{x}^j, \underline{x}^{j'}) = \| \underline{x}^j - \underline{x}^{j'} \|_D^2 = \| \underline{x}^j \|_D^2 + \| \underline{x}^{j'} \|_D^2 - 2(\underline{x}^j)' D \underline{x}^{j'}$$

where $(\underline{x}^j)' D \underline{x}^{j'} = (\underline{x}^{j'})' D \underline{x}^j$ is the scalar product of $\underline{x}^j$ and $\underline{x}^{j'}$ defined in $\mathbb{E}^n$ using the metric D. But then such distance may be seen as numerically equivalent to the sum of the sample variances for the two variables minus twice the sample covariance between the two variables, thus giving a wider interpretative meaning to the scalar product in $\mathbb{E}^n$.

If we denote by $\theta_{jj'}$ the angle between the directions of $\underline{x}^j$ and $\underline{x}^{j'}$, we may write

$$\cos \theta_{jj'} = \frac{(\underline{x}^j)' D \underline{x}^{j'}}{\| \underline{x}^j \|_D \| \underline{x}^{j'} \|_D} ,$$

showing that if both $\underline{x}^j$ and $\underline{x}^{j'}$ are D-unitary then the scalar product in $\mathbb{E}^n$ of the two vectors $\underline{x}^j$ and $\underline{x}^{j'}$ will be numerically equivalent to the cosine of the angle between the two vectors or say to the sample correlation between the two variables. In general such a scalar product will be equal to the cosine multiplied by the norms of $\underline{x}^j$ and $\underline{x}^{j'}$ and thus numerically equivalent to the sample covariance between the two variables.

In $\mathbb{E}^n$ we now define the norm of the projection of a vector $\underline{x}^j$ over the direction of the unit vector $\underline{v}$ as

$$\| \underline{z}^j \|_D = |(\underline{x}^j)'D\underline{v}| = |\underline{v}'D\underline{x}^j| = \| \underline{x}^j \|_D \; |\cos\theta_{jv}| \;, \tag{2.17}$$

θ_{jv} being the angle between $\underline{x}^j$ and the direction of $\underline{v}$.

Definition 11: Two vectors $\underline{x}^j$ and $\underline{w}$ in $\mathbb{E}^n$ are said to be D-orthogonal if

$$(\underline{x}^j)'D\underline{w} = \underline{w}'D\underline{x}^j = 0 \;. \;\square$$

If we are working with centered variables, we are really working in a subspace $\mathbb{E}^{n-1}$ of $\mathbb{E}^n$, since the vector $\underline{1}_n$, (vector with all n components equal to 1) will be D-orthogonal to any $\underline{x}^j$ $(j = 1, ..., p)$, i.e.

$$\underline{1}_n'D\underline{x}^j = (\underline{x}^j)'D\underline{1}_n = 0 \qquad j = 1, ..., p \;.$$

2.12 Inertia of the cloud of variable-points.

We now consider again three situations.

a) Inertia in relation to the origin in $\mathbb{E}^n$.

Definition 12: The inertia of the cloud $N(I)$ relative to a point $\underline{w} \in \mathbb{E}^n$, when the weight given to $\underline{x}^j$ is l_j^2 is defined as

$$In_{\underline{w}}(J) = \sum l_j^2 \| \underline{x}^j - \underline{w} \|_D^2 = \sum_{j=1}^{p} l_j^2 (\underline{x}^j - \underline{w})'D(\underline{x}^j - \underline{w}) \;. \;\square \tag{2.18}$$

If in (2.18) we replace $\underline{w}$ by $\underline{O}$, the origin in $\mathbb{E}^n$, we will have the inertia of the cloud $N(J)$ relative to the origin,

$$In_{\underline{O}}(J) = \sum_{j=1}^{p} l_j^2 \parallel \underline{x}^j - \underline{O} \parallel_D^2 = \sum_{j=1}^{p} l_j^2 (\underline{x}^j)' D \underline{x}^j = tr(X'DXQ) = tr(VQ) = In_{\underline{O}}(I) \ , \tag{2.19}$$

the last equality being obtainable from (2.4), Definition 4.

Equation (2.19) shows how the inertia relative to the origin of the two dual clouds of points is the same. This is a curious result but also, by now, an expected one and is intimately related to the fact that the two analyses are duals.

We may notice that for (2.19) to hold, the centering of $N(I)$ is not essential.

We shall be using the matrix $W = XQX'$ later. In that case

$$In_{\underline{O}}(J) = tr(VQ) = tr(X'DXQ) = tr(XQX'D) = tr(WD) \ . \tag{2.20}$$

b) Inertia in relation to a given direction in $\mathbb{E}^n$.

Let $\underline{v} \in \mathbb{E}^n$ be a D-unitary vector, i.e. $\underline{v}'D\underline{v} = \parallel \underline{v} \parallel_D^2 = 1$.

<u>Definition 13</u>: The inertia of the cloud of variable points $N(J)$ explained by (or relative to) the direction $\Delta \underline{v}$ in $\mathbb{E}^n$ that passes through the point $\underline{w} \in \mathbb{E}^n$ is defined by

$$In_{\underline{v}}(J) = \sum l_j^2 \parallel \underline{z}^j - \underline{w} \parallel_D^2 \ , \tag{2.21}$$

where $\underline{z}^j$ is the projection of $\underline{x}^j$ onto the direction $\Delta \underline{v}$. □

<u>Lemma 7</u>: If in Definition 13 above the point $\underline{w}$ is the origin in $\mathbb{E}^n$, then

$$In_{\underline{v}}(J) = \sum_{j=1}^{p} l_j^2 \parallel \underline{z}^j \parallel_D^2 = \underline{v}'DWD\underline{v} \tag{2.22}$$

where $W = XQX'$ is the 'inertia matrix' in $\mathbb{E}^n$.

Proof: The first equality is simply obtained by replacing $\underline{w}$ by $\underline{O}$ in (2.21). But then using (2.17)

$$\begin{aligned} In_{\underline{v}}(J) &= \textstyle\sum_{j=1}^{p} l_j^2 \parallel \underline{z}^j \parallel_D^2 = \sum_{j=1}^{p} l_j^2 \left((\underline{x}^j)'D\underline{v}\right)^2 = \sum_{j=1}^{p} (\underline{v}'D\underline{x}^j)\, l_j^2 \,((\underline{x}^j)'D\underline{v}) \\ &= \underline{v}'DXQX'D\underline{v} = \underline{v}'DWD\underline{v} \ . \ \blacksquare \end{aligned}$$

2.13 Factorial axes in $\mathbb{E}^n$ and subspace of $\mathbb{E}^n$ closest to the cloud of variable-points.

Let us suppose that we are looking for a subspace of $\mathbb{E}^n$ of dimension $r \leq n$ that under some optimality definition lies closest to the cloud $N(J)$.

As stated in 2.9, the cloud $N(J)$, of variable points is usually not centered. Therefore, $In_{\underline{O}}(J) \neq In_{\underline{g}}(J)$, where $\underline{g}$ represents now the center of mass for the cloud $N(J)$. But then by Huygens' Theorem, $In_{\underline{O}}(J) > In_{\underline{g}}(J)$. So the subspace of $\mathbb{E}^n$ we are going to obtain is not the subspace of $\mathbb{R}^n$ of dimension r that lies closest to the cloud $N(J)$ of variable points, which would have the center of mass of $N(J)$ as the origin.

However, as we shall see later this subspace defined by the first r factorial axes of the cloud $N(J)$ (orthogonal directions in $\mathbb{E}^n$, passing through the origin and relative to which the inertia of the cloud $N(J)$ attains successive maximums) will still be the subspace of $\mathbb{E}^n$ that is closest to the cloud $N(J)$. The elements of its orthonormal basis are the vectors in $\mathbb{E}^n$ that maximize (as we shall see later) a weighted sum of the squares of their cosines (statistically the same as correlations) with the p variables $\underline{x}^j$ $(j = 1, ..., p)$.

Similar to Definition 6 in 2.6 we have now the following definition for the first factorial axis in $\mathbb{E}^n$.

<u>Definition 14</u>: The first factorial axis is defined as the direction $\Delta\underline{v}^1$ in $\mathbb{E}^n$, where $\underline{v}^1$ is the D-unitary vector in $\mathbb{E}^n$ that maximizes

$$In_{\underline{v}}(J) = \underline{v}'DWD\underline{v} \ . \ \square$$

<u>Lemma 8</u>: The vector $\underline{v}^1$ is the D-unitary eigenvector of WD associated with its largest eigenvalue λ_1, where λ_1 is the same as in Lemma 1.

Proof: We maximize $(\underline{v}^1)'DWD\underline{v}^1$ subject to $(\underline{v}^1)'D\underline{v}^1$ by equating the derivative of the Lagrangian

$$L' = (\underline{v}^1)'DWD\underline{v}^1 - \lambda_1 \left((\underline{v}^1)'D\underline{v}^1 - 1\right)$$

to zero. We then obtain

$$DWD\underline{v}^1 = \lambda_1 D\underline{v}^1 \ ,$$

or, premultiplying by D^{-1},

$$WD\underline{v}^1 = \lambda_1 \underline{v}^1 \ . \tag{2.23}$$

The maximum is thus

$$(\underline{v}^1)'DWD\underline{v}^1 = \lambda_1 (\underline{v}^1)'D\underline{v}^1 = \lambda_1 \ . \tag{2.24}$$

But from (2.23) and (2.24) we see that $\underline{v}^1$ is the D-unitary eigenvector of WD associated with λ_1, the quantity maximized, and as such the largest eigenvalue of WD.

The eigenvalues of VQ and WD are the same because premultiplying (2.7) by XQ, we get

$$XQX'DXQ\underline{u}_\alpha = \lambda_\alpha XQ\underline{u}_\alpha$$
$$\Leftrightarrow \quad WD\underline{c}^\alpha = \lambda_\alpha \underline{c}^\alpha \ .$$

This shows that the eigenvalues of VQ and WD are the same. Therefore λ_1 above is the same as λ_1 in Lemma 1. ∎

Proceeding in a similar manner, with the further restriction that any factorial axis must be D-orthogonal to the previous ones, and using a procedure similar to the one used in 2.6, we have the following Corollary.

Corollary 3: The subspace of $\mathbb{E}^n$ of dimension $r \leq n$ that lies closest to the cloud $N(J)$ is the subspace $\mathcal{G}^r$ spanned by $\underline{v}^1, \ldots, \underline{v}^r$, the D-unitary eigenvectors of WD associated with the eigenvalues $\lambda_1 > \ldots > \lambda_r > 0$, and then

$$In_{\underline{v}^\alpha}(J) = (\underline{v}^\alpha)'DWD\underline{v}^\alpha = \lambda_\alpha \qquad (\alpha = 1, ..., r)$$

where the vectors $\underline{v}^\alpha$ and the values λ_α satisfy

$$WD\underline{v}^\alpha = \lambda_\alpha \underline{v}^\alpha \qquad (\alpha = 1, ..., r) \ . \ \blacksquare \tag{2.25}$$

Then, from Corollaries 2 and 3 the following Corollary is immediate.

Corollary 4:

$$In_{\underline{v}^\alpha}(J) = In_{\underline{u}_\alpha}(I) = \lambda_\alpha \ . \ \blacksquare \tag{2.26}$$

We may notice that the vectors $\underline{v}^\alpha$ are proportional to the vectors $\underline{c}^\alpha$ used in the proof of Lemma 8 or the same as the ones defined in (2.14). Both $\underline{v}^\alpha$ and $\underline{c}^\alpha$ $(\alpha = 1, ..., r)$ are eigenvectors of WD but while, from Corollary 3, $\| \underline{v}^\alpha \|_D^2 = 1$, from (2.14) and (2.7), $\| \underline{c}^\alpha \|_D^2 = \underline{u}'_\alpha QX'DXQ\underline{u}_\alpha = \underline{u}'_\alpha QVQ\underline{u}_\alpha = \lambda_\alpha$.

Theorem 2: The vectors $\underline{v}^\alpha$ are the vectors in $\mathbb{E}^n$ that maximize the weighted sum of their squared cosines with the p variables $\underline{x}^j$ $(j = 1, ..., p)$, with weights being $l_j^2 \ \| \underline{x}^j \|_D^2$.

Proof: Using the fact that for $\alpha = 1, ..., r$, λ_α was the quantity maximized, and using (2.26), (2.22) and the definition of the scalar product in $\mathbb{E}^n$ in (2.17) we may write

$$\begin{aligned} \lambda_\alpha &= In_{\underline{v}^\alpha}(J) = (\underline{v}^\alpha)'DWD\underline{v}^\alpha = (\underline{v}^\alpha)'DXQX'D\underline{v}^\alpha \\ &= \sum_{j=1}^{p} \left((\underline{v}^\alpha)'D\underline{x}^j\right) l_j^2 \left((\underline{x}^j)'D\underline{v}^\alpha\right) \ = \ \sum_{j=1}^{p} l_j^2 \left((\underline{x}^j)'D\underline{v}^\alpha\right)^2 \\ &= \sum_{j=1}^{p} l_j^2 \ \| \underline{z}_\alpha^j \|_D^2 \ = \ \sum_{j=1}^{p} l_j^2 \ \| \underline{x}^j \|_D^2 \cos^2\theta(\underline{x}^j, \underline{v}^\alpha) \ , \end{aligned}$$

where $\theta(\underline{x}^j, \underline{v}^\alpha)$ represents the angle between the directions of $\underline{x}^j$ and $\underline{v}^\alpha$. $\blacksquare$

Here in $\mathbb{E}^n$ also we have the equivalent of Lemma 3. Given the orthogonality of the vectors $\underline{v}^\alpha$ $(\alpha = 1, ..., r)$, we could easily show that

$$In_{(\underline{v}^1, ..., \underline{v}^r)}(J) = \sum_{\alpha=1}^{r} In_{\underline{v}^\alpha}(J) = \sum_{\alpha=1}^{r} \lambda_\alpha \ ,$$

using an argument similar to the one in the proof of Lemma 3. Together with Theorem 2 this shows that $\mathcal{G}^r$ is the subspace of $\mathbb{E}^n$ that lies closest to the cloud of variables $N(J)$, not only in the sense that it maximizes $In_{(\underline{v}^1,...,\underline{v}^r)}(J)$, and correspondingly minimizes $In_{(\underline{v}^1,...,\underline{v}^r)^\perp}(J)$, but also in the sense that the vectors $\underline{v}^\alpha$ $(\alpha = 1, ..., r)$, that form an orthonormal basis of $\mathcal{G}^r$, verify the property stated in Theorem 2 above.

Definition 15: The percentage of inertia explained by $\mathcal{G}^r$ is

$$\frac{In_{(\underline{v}^1,...,\underline{v}^r)}(J)}{In_{\underline{Q}}(J)} = \sum_{\alpha=1}^{r} \frac{In_{\underline{v}^\alpha}(J)}{In_{\underline{Q}}(J)} = \frac{\sum_{\alpha=1}^{r} \lambda_\alpha}{\sum_{\alpha=1}^{n} \lambda_\alpha}$$

where

$$\frac{In_{\underline{v}^\alpha}(J)}{In_{\underline{Q}}(J)} = \frac{\lambda_\alpha}{\sum_{\alpha=1}^{n} \lambda_\alpha}$$

is the percentage of inertia explained by the axis $\Delta\underline{v}^\alpha$ $(\alpha = 1, ..., r)$. □

These percentages are measures of the importance of the subspace $\mathcal{G}^r$, or the axes $\Delta\underline{v}^1, ..., \Delta\underline{v}^r$.

It was proved in 2.6, Lemma 6, that

$$In_{\underline{Q}}(I) = tr(VQ) = \sum_{\alpha=1}^{p} \lambda_\alpha$$

and since, from (2.19) and (2.20),

$$In_{\underline{Q}}(J) = In_{\underline{Q}}(I) = tr(VQ) = tr(WD)$$

we have the equality

$$In_{\underline{Q}}(J) = \sum_{\alpha=1}^{p} \lambda_\alpha = \sum_{\alpha=1}^{n} \lambda_\alpha \ .$$

This equality could also be formally proved in a way similar to the proof of Lemma 6.

Note that the equality

$$\sum_{\alpha=1}^{p} \lambda_\alpha = \sum_{\alpha=1}^{n} \lambda_\alpha$$

may appear contradictory, but is not really so, as $n-p$ or $p-n$ eigenvalues are zero depending on whether $n>p$ or $p>n$. It may be proper, therefore to write the above as

$$In_{\underline{Q}}(I) = In_{\underline{Q}}(J) = \sum_{\alpha=1}^{\min(n,p)} \lambda_\alpha \ .$$

In all this discussion it is assumed that the non-zero eigenvalues are all distinct and that $rank(VQ) = \min(n,p)$.

2.14 Representation of the variable-points on the new basis.

The representation of the p variables on the D-orthonormal basis of $\mathcal{F}^r$ formed by the vectors $\underline{v}^1, ..., \underline{v}^r$ will be given by the rows of

$$B = X'D\ \Phi \tag{2.27}$$

where

$$\Phi = \begin{bmatrix} \underline{v}^1 \mid ... \mid \underline{v}^\alpha \mid ... \mid \underline{v}^r \end{bmatrix} \quad (r \leq n) \ .$$

The j^{th} row of B is the vector

$$(\underline{b}^j)' = (X'D\underline{v}^\alpha)' \qquad (j = 1, ..., p) \ .$$

It is the representation or D-orthogonal projection of $\underline{x}^j$ on $\mathcal{G}^r$, subspace of $\mathbb{E}^n$ of (reduced) dimension $r(\leq n)$, a subspace spanned by the columns of Φ. The vectors $\underline{b}^j$ will give 'the best' representation of $\underline{x}^j$ in a subspace of $\mathbb{E}^n$ of dimension $r(\leq n)$ in the sense that, from Theorem 2, $\sum_{\alpha=1}^{r}\sum_{j=1}^{p} l_j^2 \parallel \underline{x}^j \parallel_D^2 \cos^2\theta(\underline{x}^j, \underline{v}^\alpha)$ is maximum $\Big($and also $In_{(\underline{v}^1,...,\underline{v}^r)^\perp}(J) = In_{\underline{Q}}(J) - \sum_{\alpha=1}^{r} In_{\underline{v}^\alpha}(J) = \sum_{j=1}^{p} \parallel \underline{x}^j - \underline{b}^j \parallel_D^2$ is minimum$\Big)$. In this sense the p vectors $\underline{b}^j$ $(j = 1, ..., p)$ thus give 'the best' r-dimensional representation of the cloud of points $N(J)$.

The columns of the matrix B are the vectors

$$\underline{b}_\alpha = X'D\underline{v}^\alpha \quad \in \quad \mathbb{E}^p \tag{2.28}$$

each one of which has the coordinates of all the p variables on the new factorial axis of order α in $\mathbb{E}^n$. They will be shown to be proportional to the vectors $\underline{u}_\alpha$ as the vectors $\underline{v}^\alpha$ are proportional to the vectors $\underline{c}^\alpha$ defined in (2.14).

2.15 Transformation formulas.

As noticed in section 13 (immediately after Corollary 4), both the vectors $\underline{c}^\alpha$ defined in (2.14) and the vectors $\underline{v}^\alpha$ defined in (2.25) are eigenvectors of WD, with the only difference that $\underline{v}^\alpha$ are D-unitary, but $\underline{c}^\alpha$ have a squared D-norm equal to λ_α. They are thus proportional, satisfying

$$\underline{v}^\alpha = \frac{1}{\sqrt{\lambda_\alpha}}\underline{c}^\alpha \ . \tag{2.29}$$

From the relation between vectors $\underline{v}^\alpha$ and $\underline{c}^\alpha$ and from Corollary 4, it is appropriate to take $r = q$, this is we take the same dimension for the subspaces $\mathcal{F}^q$ in $\mathbb{E}^p$ and $\mathcal{G}^r$ in $\mathbb{E}^n$.

From (2.29) we see that the matrix C defined in (2.12) satisfies

$$C = \Phi\Lambda^{1/2}$$

where $\Lambda = diag(\lambda_1, ..., \lambda_r)$, or correspondingly that the matrix B defined in (2.27) may be written as

$$B = X'DC\Lambda^{-1/2} = X'DXQU\Lambda^{-1/2} = U\Lambda^{1/2} \ . \tag{2.30}$$

All this means that once the eigenvectors $\underline{u}_\alpha$ are obtained the vectors $\underline{v}^\alpha$ may be obtained directly from them through one of the transformations and it is not necessary to diagonalize WD, or conversely, if we first obtained the eigenvectors $\underline{v}^\alpha$ from WD we may obtain the vectors $\underline{u}_\alpha$ directly from the vectors $\underline{v}^\alpha$ through the other transformation formula and it is not necessary to diagonalize VQ.

The two transformations are

$$\underline{v}^{\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} XQ\underline{u}_\alpha \tag{2.31}$$

and

$$\underline{u}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} X'D\underline{v}^{\alpha} \ . \tag{2.32}$$

The transformation (2.31) is directly obtained from (2.14) and (2.29), while (2.32) is obtained from (2.28) and (2.30).

Due to the equivalence of the two dual analyses, it is obvious that a practitioner should carry out only that which is simpler.

We may notice from (2.31) and (2.32) that

$$\Phi = XQU\Lambda^{-1/2} \tag{2.33}$$

and

$$U = X'D\Phi\Lambda^{-1/2} \ , \tag{2.34}$$

where $\Lambda = diag(\lambda_\alpha)$. But then, from (2.33) and (2.34)

$$\begin{aligned} \Phi\Lambda^{-1/2}U' &= XQU\Lambda^{-1/2}\Lambda^{1/2}\Lambda^{-1/2}\Phi'DX \\ &= XQUU' = \Phi\Phi'DX = \tilde{X} \end{aligned}$$

where $\tilde{X} \equiv X$ if the r columns of Φ (or correspondingly if the r columns of U), form a complete basis for $\mathbb{E}^n$ ($\mathbb{E}^p$), showing that the Factorial Analysis is ultimately a special singular value decomposition of the data matrix X with metrics Q in the row space and D in the columns space.

Appendix 2.A

The Euclidean distance and the contribution of each variable to the sum of the squared distances among observation points.

When using the Euclidean distance as in (2.3) the contribution of each variable to the sum of the square of all the distances in $N(I)$ will be proportional (equal) to their variances.

Proof:

Twice the sum of the squared distances among all the observation points in $N(I)$ is, when using (2.3)

$$\sum_{i=1}^{n}\sum_{i'=1}^{n} m_i \, m_{i'} \sum_{j=1}^{p}(x_i^j - x_{i'}^j)^2 \ ,$$

which can also be written, after a little algebra, as

$$2 \sum_{j=1}^{p} s_j^2 \ ,$$

where

$$s_j^2 = \sum_{i=1}^{n} m_i \, (x_i^j - \overline{x}^j)^2 \ .$$

We may note that s_j^2 represents the inertia of the cloud of observation-points $N(I)$ relative to the direction $\Delta\underline{e}^j$ $(j = 1, ..., p)$ ($\underline{e}^j$ $(j = 1, ..., p)$ being the j^{th} canonical basis vector of $\mathbb{E}^p$, i.e. a vector with all zero and a 1 in the j^{th} position) or the sample variance of the j^{th} variable, when the definition of variance is taken in a broader sense, referring to the use of the masses m_i. Here is finally explained the equivalence between inertia of the cloud $N(I)$ relative to the direction $\Delta\underline{e}^j$ and estimated or sample variance of the j^{th} variable.

Appendix 2.B

Proof of the equality,

$$In_{\underline{O}}(I) = \frac{1}{2} \sum_{i=1}^{n} \sum_{i'=1}^{n} d_Q^2(\underline{x}_i, \underline{x}_{i'}) \ ,$$

for centered data (centered for the masses m_i).

Proof:

$$\sum_{i=1}^{n} \sum_{i'=1}^{n} m_i \ m_{i'} \ d_Q^2(\underline{x}_i, \underline{x}_{i'}) \ = \ \sum_{i=1}^{n} \sum_{i'=1}^{n} m_i \ m_{i'} \ \| \ \underline{x}_i - \underline{x}_{i'} \ \|_Q^2$$

$$= \sum_{i=1}^{n} \sum_{i'=1}^{n} m_i \ m_{i'} \ \| \ \underline{x}_i \ \|_Q^2 \ + \sum_{i=1}^{n} \sum_{i'=1}^{n} m_i \ m_{i'} \ \| \ \underline{x}_{i'} \ \|_Q^2 \ - \\ - \ 2 \sum_{i=1}^{n} \sum_{i'=1}^{n} m_i \ m_{i'} \ \underline{x}_i' Q \underline{x}_{i'}$$

$$= \sum_{i'=1}^{n} m_{i'} \sum_{i=1}^{n} m_i \ \ \| \ \underline{x}_i \ \|_Q^2 + \sum_{i=1}^{n} m_i \sum_{i'=1}^{n} m_{i'} \ \ \| \ \underline{x}_{i'} \ \|_Q^2 \ - \\ -2 \sum_{i'=1}^{n} m_{i'} \ \Big(\sum_{i=1}^{n} m_i \ \underline{x}_i' \Big) Q \underline{x}_{i'} \qquad (2.B.1)$$

$$= tr(VQ) + tr(VQ) - 2 \sum_{i'=1}^{n} m_{i'} \ \underline{g}' Q \underline{x}_{i'} \qquad (2.B.2)$$

$$= tr(VQ) + tr(VQ) - 2 \ \underline{g}' Q \Big(\sum_{i'=1}^{n} m_{i'} \ \underline{x}_{i'} \Big)$$

$$= tr(VQ) + tr(VQ) - 2 \ \underline{g}' Q \underline{g}$$

$$= 2 \ tr(VQ) = 2 \ In_{\underline{O}}(I) \ .$$

where we used, as defined above, $V = X'DX$, with $D = diag(m_i)$, and

$$\underline{g} = \sum_{i=1}^{n} m_i \ \underline{x}_i$$

that for centered data for the masses m_i as we assume we are dealing with is confounded with the origin $\underline{O}$, so that $\underline{g}' Q \underline{g} \ = \ 0$ (zero), and where from (2.B.1) to (2.B.2) we used (2.4) and the fact that $\sum_{i=1}^{n} m_i = 1$.

Appendix 2.C

Matrix form of the vector $\underline{z}_i$ and related topics.

Let $\underline{z}_i$ denote the projection of $\underline{x}_i$ onto $\mathcal{F}^q \subseteq \mathbb{E}^p$, the subspace of $\mathbb{E}^p$ spanned by the q vectors $\underline{u}_1, ..., \underline{u}_q$, and let $\underline{z}_i^\alpha$ denote the projection of $\underline{x}_i$ onto the direction $\Delta\underline{u}_\alpha$ $(\alpha = 1, ..., q \leq p)$.

Before obtaining the matrix representation for the vector $\underline{z}_i$ let us first consider the matrix representation of each $\underline{z}_i^\alpha$.

It is clear that being $\underline{z}_i^\alpha$ the projection of $\underline{x}_i$ on the direction $\Delta\underline{u}_\alpha$, then, from Definition 1 in 2.4, we have

$$\underline{z}_i^\alpha = \pm\ \underline{u}_\alpha \parallel \underline{z}_i^\alpha \parallel = \underline{u}_\alpha(\underline{u}_\alpha' Q\underline{x}_i)$$

where the sign $\pm$ represents the result of $sign(\underline{u}_\alpha' Q\underline{x}_i)$.

Then, $\underline{z}_i$ being the projection of $\underline{x}_i$ onto a subspace of $\mathbb{E}^p$ of dimension q spanned by the q Q-orthogonal vectors $\underline{u}_\alpha$ $(\alpha = 1, ..., q)$

$$\underline{z}_i = \sum_{\alpha=1}^{q} \underline{z}_i^\alpha = \sum_{\alpha=1}^{q} \underline{u}_\alpha \underline{u}_\alpha' Q\underline{x}_i = UU'Q\underline{x}_i\ , \tag{2.C.1}$$

where $U = \left[\underline{u}_1 \middle| ... \middle| \underline{u}_\alpha \middle| ... \middle| \underline{u}_q\right]$ is the matrix defined in 2.8.

Note that

$$\underset{(p\times 1)}{\underline{z}_i} = \underset{(p\times q)}{U}\ \underset{(q\times 1)}{\underline{c}_i} \tag{2.C.2}$$

where $\underline{c}_i$ is the vector defined in (2.13). Then

$$\parallel \underline{z}_i \parallel_Q^2 = \underline{c}_i' U'QU\underline{c}_i = \parallel \underline{c}_i \parallel_{I_q}^2\ , \tag{2.C.3}$$

since $U'QU = I_q$.

From (2.C.1) it is clear that if $UU'Q = I_p$, (which is so only if $q = p$ or when $q < p$, the q vectors $\underline{u}_1, ..., \underline{u}_q$ form a complete basis for $\mathbb{E}^p$) then $\underline{z}_i \equiv \underline{x}_i$. Let us suppose that $q = p$. Then, since $UU'Q = I_p$, we can write

$$\| \underline{x}_i \|_Q^2 = \underline{x}_i' QUU'Q\underline{x}_i = \sum_{\alpha=1}^{p} \underline{x}_i' Q\underline{u}_\alpha \underline{u}_\alpha' Q\underline{x}_i = \sum_{\alpha=1}^{p} \| \underline{z}_i^\alpha \|_Q^2 \quad . \tag{2.C.4}$$

The expressions (2.C.2) and (2.C.3) may be further interpreted. While $\underline{z}_i$ is the projection of $\underline{x}_i$ onto $\mathcal{F}^q$ with coordinates relative to the canonical basis of $\mathbb{E}^p$, $(\underline{e}^1, ..., \underline{e}^p)$ (with $\underline{e}^j$ $(j = 1, ..., p)$ being a vector of all zeros and a 1 in the j^{th} position), $\underline{c}_i$ is the same projection of $\underline{x}_i$ onto $\mathcal{F}^q$ with coordinates relative to the basis $\underline{u}_1, ..., \underline{u}_q$, i.e. $\underline{z}_i$ is the representation of $\underline{c}_i$ on the canonical basis of $\mathbb{E}^p$ or vice-versa, $\underline{c}_i$ is the representation of $\underline{z}_i$ on the basis $\underline{u}_1, ..., \underline{u}_q$. This explains why we have

$$\underline{c}_i = U'Q\underline{x}_i$$

and

$$\underline{z}_i = U\underline{c}_i = UU'Q\underline{x}_i \ .$$

Observe that U and $U'Q$ are generalized inverses of each other in the sense that for any $q \leq p$

$$U(U'Q)U = U$$

and

$$(U'Q)U(U'Q) = U'Q$$

because for any $q \leq p$, $U'QU = I_q$.

In general we will have a situation where $q < p$ and the q vectors $\underline{u}_1, ..., \underline{u}_q$ do not form a complete basis for $\mathbb{E}^p$. In this situation we have

$$\underline{z}_i \approx \underline{x}_i \quad \text{or} \quad \underline{z}_i = \hat{\underline{x}}_i \ .$$

This may be interpreted as "$\underline{z}_i$ is close to $\underline{x}_i$", the 'closeness' depending on how 'complete' is $(\underline{u}_1, ..., \underline{u}_q)$ as a basis for $\mathbb{E}^p$.

In this case we will have

$$\| \underline{z}_i \|_Q^2 = \sum_{\alpha=1}^{q} \| \underline{z}_i^\alpha \|_Q^2$$

and since $q < p$ and $\| \underline{x}_i \|_Q^2 = \sum_{\alpha=1}^{p} \| \underline{z}_i^\alpha \|_Q^2$ and since $\underline{u}_1, ..., \underline{u}_q$ do not form a complete basis for $\mathbb{E}^p$, at least one of the norms $\| \underline{z}_i^\alpha \|_Q^2$ $(\alpha = 1, ..., q)$ will not be null for at least one $\underline{x}_i$, and we will in general have

$$\| \underline{x}_i \|_Q^2 \geq \| \underline{z}_i \|_Q^2 \tag{2.C.5}$$

i.e.

$$\| \underline{x}_i \|_Q^2 = \underline{x}_i' Q \underline{x}_i \ \geq \ \underline{x}_i' Q U U' Q \underline{x}_i \ = \| \underline{x}_i \|_{QUU'Q}^2 \ ,$$

and then also

$$\| \underline{z}_i - \underline{z}_{i'} \|_Q^2 = \| \underline{x}_i - \underline{x}_{i'} \|_{QUU'Q}^2 \leq \| \underline{x}_i - \underline{x}_{i'} \|_Q^2 \ . \tag{2.C.6}$$

We can then conclude from (2.C.5) and (2.C.6) that the projection of the cloud $N(I)$ onto $\mathcal{F}^q$, a subspace of reduced dimension, or onto a plane defined by two factorial axes $\Delta\underline{u}_\alpha$ and $\Delta\underline{u}_{\alpha'}$, with $\alpha \neq \alpha'$ and $\alpha, \alpha' \in \{1, ..., q\}$, will always be a reduction (contraction) of the original cloud $N(I)$ since from (2.C.5) and (2.C.6) we see that not only the projection points have smaller norm than the original points but also lie closer to each other.

CHAPTER 3

THE GENERALIZED CANONICAL ANALYSIS

3.1 Introduction.

Let us suppose that m groups of distinct variables have been observed on a set of n individuals. It is assumed that the k^{th} $(k = 1, ..., m)$ group of variables has p_k variables in it and that the total number of variables is $p = \sum_{k=1}^{m} p_k$. The sets of variables may be sets of indicator variables for the categories of an attribute or categorical variable, but we shall consider this situation later.

Subdivide the data matrix X of order $n \times p$ as

$$X = \left[X_1 \;\middle|\; \ldots \;\middle|\; X_k \;\middle|\; \ldots \;\middle|\; X_m \right]$$

where

$$X = \left[\; x_i^j \; \right] \qquad i = 1, \ldots, n \quad ; \quad j = 1, \ldots, p$$

and

$$X_k = \left[\; x_{ki}^j \; \right] \qquad i = 1, \ldots, n \quad ; \quad j = 1, \ldots, p_k \quad ; \quad k = 1, \ldots, m \; .$$

Pearson's correlation coefficient is a measure of linearity between two variables. Geometrically, cosine of the angle is the correlation coefficient.

Canonical correlations may then be seen as a generalization of the correlation coefficient as they measure the linear relationship between two sets of variables.

In Generalized Canonical Analysis (GCA) we consider such relations among several (m) groups or sets of variables.

In this chapter, we will consider the geometrical properties of GCA (based on Carroll's (1968) idea) and how GCA is an adequate technique to study the closeness of the subspaces spanned by each group of variables. For measuring this closeness, we use the cosines of critical angles between the subspaces. For the choice of an adequate metric Q the eigenvalues of VQ, where V and Q have been defined in Chapter 2, are shown to be the sum of the squares of such cosines. The principal or generalized canonical variables will be derived and their optimal properties studied. Also the canonical variables for each set of variables will be defined and shown to be the projections of the principal or generalized canonical variable on the subspaces spanned by the variables in each set. In the next chapter it will be shown that the product of all the possible eigenvalues that summarize the closeness among the m subspaces spanned by the m groups of variables is related to the likelihood ratio test statistic. The distributional approach of GCA and the distribution of the test statistics for some useful hypotheses will also be derived in the next chapter.

3.2 The m+1 clouds of observation-points.

Each observation i $(i = 1, \ldots, n)$ has its 'global' representation in $\mathbb{E}^p$ through the column vector $\underline{x}_i \in \mathbb{E}^p$, where $\underline{x}_i'$ is the i^{th} row of X. In addition, each observation has a 'partial' representation in each of the m subspaces $\mathbb{E}^{p_k}$ $(k = 1, \ldots, m)$ of $\mathbb{E}^p$, through the column vectors $\underline{x}_{ki} \in \mathbb{E}^{p_k}$, $\underline{x}_{ki}'$ being the i^{th} row of X_k.

To each one of the n observation-points $\underline{x}_i$ in $\mathbb{E}^p$ we attach a mass $m_i > 0$, such that $\sum_{i=1}^{n} m_i = 1$. Then we have the cloud

$$N(I) = \left\{ (\underline{x}_i, m_i),\ \underline{x}_i \in I\!\!E^p, m_i > 0,\ \sum_{i=1}^{n} m_i = 1 \mid i \in I \right\} \quad , \ I = \{1, ..., n\}$$

and we will refer to this cloud as the cloud of observation-points. But, we also have m other clouds of points $\underline{x}_{ki}$, one in each subspace $I\!\!E^{p_k}$ of $I\!\!E^p$, namely the clouds

$$N_k(I) = \left\{ (\underline{x}_{ki}, m_i),\ \underline{x}_{ki} \in I\!\!E^{p_k},\ m_i > 0,\ \sum_{i=1}^{n} m_i = 1 \mid i \in I \right\} \quad , \ k \in K = \{1, ..., m\} \ .$$

3.3 Choice and interpretation of the metric Q in $I\!\!E^p$.

Q, the matrix of the metric used in $I\!\!E^p$ (see section 2.3), may be any symmetric positive-definite matrix. However, we need to make the choice of Q according to some desirable criteria. In Generalized Canonical Analysis (GCA) we will use the metric

$$Q = bdiag(Q_k) \quad \text{with} \quad Q_k = (X_k' D X_k)^{-1} \quad (k = 1, ..., m) \ . \tag{3.1}$$

This metric may easily be seen as a generalization of the metric $Q^* = diag(1/\hat{\sigma}_j^2)$ $(j = 1, \ldots, p)$. This latter one is commonly used in several statistical methods such as Principal Components Analysis. With this metric Q^* each variable is weighted by the inverse of its sample variance, to make the global computation of the distances scale free. We will see how with the use of the metric Q in $I\!\!E^p$ each group of variables is 'weighted' by the inverse of its inertia matrix, or the estimate of its variance-covariance matrix, if some distribution is assumed for the variables. In this and the next section the meaning of the term 'weighted' will be made more clear.

In GCA we are mainly interested in the groups of variables, or rather in the subspaces generated by them (such groups of variables), and not in the variables themselves. Our primary goal is to study the relationships of the m subspaces $I\!\!E^{p_k}$, generated by the columns of each X_k $(k = 1, \ldots, m)$.

Correspondingly, in the observations space $\mathbb{E}^p$ we are not primarily interested in the distances among the 'global' representations of the observations but mainly in the distances among the m 'partial' representations of the same observation.

It makes sense to give to each group of variables a participation proportional to the number of variables in it while computing the distances among the observations. This will be achieved by using the metric $Q = bdiag(X_k' D X_k)^{-1}$ as we will see later. This metric weights the contribution of each group of variables to the total inertia of the cloud of observation-points by the inverse of their inertia matrix, or estimated variance-covariance matrix, $X_k' D X_k$.

The metric Q may be seen as a multiple Mahalanobis type metric. Its use in $\mathbb{E}^p$ implies the use in each subspace $\mathbb{E}^{p_k}$ of its Mahalanobis metric $Q_k = V_k^{-1} = (X_k' D X_k)^{-1}$.

3.4 Norm, scalar product and orthogonality in $\mathbb{E}^p$, associated with the metric Q.

Definition 1: The introduction of the metric Q allows us to define the norm of an observation vector as

$$\| \underline{x}_i \|_Q^2 = \sum_{k=1}^{m} \| \underline{x}_{ki} \|_{Q_k}^2$$

where

$$\| \underline{x}_{ki} \|_{Q_k}^2 = \underline{x}_{ki}' Q_k \, \underline{x}_{ki} \; . \; \square$$

The scalar product of any two vectors in $\mathbb{E}^p$, in particular of two observation vectors $\underline{x}_i$ and $\underline{x}_{i'}$ may also be decomposed by blocks in a similar way

$$\underline{x}_i' Q \underline{x}_{i'} = \sum_{k=1}^{m} \underline{x}_{ki}' Q_k \, \underline{x}_{ki'} \; . \qquad (3.2)$$

Any two vectors $\underline{x}_i$ and $\underline{x}_{i'}$, in $\mathbb{E}^p$, are then said to be orthogonal when (3.2) is equal to zero.

3.5 Inertia of the cloud of observation-points.

Similar to the Definition 4 in 2.5.b), the inertia of the cloud of observation-points relative to the origin is now defined as follows.

Definition 2: The inertia of cloud $N(I)$ relative to the origin $\underline{O} \in \mathbb{E}^p$ is

$$\begin{aligned} In_{\underline{O}}(I) &= \sum_{i=1}^{n} m_i \parallel \underline{x}_i \parallel_Q^2 = \sum_{i=1}^{n} m_i \ \underline{x}_i' Q \underline{x}_i = \sum_{i=1}^{n} m_i \sum_{k=1}^{m} \underline{x}_{ki}' Q_k \, \underline{x}_{ki} \\ &= tr\left(\left(\sum_{k=1}^{m} X_k Q_k X_k'\right) D\right) = tr\left(\left(\sum_{k=1}^{m} W_k\right) D\right) \\ &= tr(WD) = tr(VQ) = p \ , \end{aligned}$$

where Q is defined as in (3.1), $W_k = X_k Q_k X_k'$, $W = XQX'$, $D = diag(m_i)$ and $V = X'DX$. □

We may note that

$$In_{\underline{O}}(I) = \sum_{k=1}^{m} In_k(I) \tag{3.3}$$

where

$$In_k(I) = \sum_{i=1}^{n} m_i \parallel \underline{x}_{ki} \parallel_{Q_k}^2 = tr(V_k Q_k) = tr(I_{p_k}) = p_k \ ,$$

is the inertia relative to the origin of each cloud $N_k(I)$. This shows that each group of variables will have a contribution to the total inertia of the cloud of observation-points equal to their number of variables.

The inertia relative to a given direction in $\mathbb{E}^p$, of either the main cloud of observation-points or any of the other $m+1$ subclouds, can be defined as in 2.5 (Definition 5).

3.6 Factorial axes. Subspace of $\mathbb{E}^p$ closest to the cloud of observation-points.

The factorial axes are defined as in Chapter 2, 2.6, and are the directions $\Delta\underline{u}_\alpha$ $(\alpha = 1, \ldots, q)$ in $\mathbb{E}^p$ defined by the Q-unitary vectors $\underline{u}_\alpha$ $(\alpha = 1, \ldots, q)$, eigenvectors of VQ associated with the eigenvalues λ_α, inertias of the cloud of observation-points relative to the same axes for $\alpha = 1, \ldots, q$. The q-dimensional subspace of $\mathbb{E}^p$ closer to the cloud of observation-points will still be the subspace $\mathcal{F}^q$ spanned by the q vectors $\underline{u}_\alpha$. The vectors $\underline{u}_\alpha$ $(\alpha = 1, \ldots, q)$ will form a Q-orthonormal basis for $\mathcal{F}^q$. The subspace $\mathcal{F}^q$ will not only be the subspace of $\mathbb{E}^p$ closer to the cloud of points $N(I)$ but also the subspace of $\mathbb{E}^p$ closer to the m clouds $N_k(I)$, because if, from Lemma 5 in Chapter 2, $\mathcal{F}^q$ is the subspace of $\mathbb{E}^p$ that minimizes $In_{(\underline{u}_1,\ldots,\underline{u}_q)^\perp}(I)$, from (3.3) is the subspace of $\mathbb{E}^p$ that minimizes $\sum_{k=1}^m In_k(I) - In_{(\underline{u}_1,\ldots,\underline{u}_q)}(I)$.

We should note that nevertheless the fact that each cloud $N_k(I)$ is representable in a space $\mathbb{E}^{p_k}$ of dimension p_k (or $\leq p_k$), a number of q^* factorial axes $(q^* \leq min_{k\in K} p_k,\ K = \{1, ..., m\})$, may be not enough to obtain a good representation of the cloud of points $N(I)$. We would have $q = min_{k\in K} p_k$ if all the m clouds $N_k(I)$ were coincident.

3.7 Generalized Canonical Variables and Canonical Variables.

We should remember that, as pointed out in 3.3, the use in $\mathbb{E}^p$ of the metric $Q = bdiag(Q_k)$, where $Q_k = (X_k' D X_k)^{-1}$, induces the use of the m metrics Q_k in each subspace $\mathbb{E}^{p_k}$ of $\mathbb{E}^p$ $(k = 1, ..., m)$. With this in mind we have the following definition.

Definition 3: We will call 'global' coordinate of the i^{th} observation on the factorial axis of order α $(\alpha = 1, ..., q)$ to

$$c_i^{\alpha} = \underline{u}'_{\alpha} Q \underline{x}_i \ ,$$

since it is obtained through $\underline{x}_i$, the 'global' representation of the observation i.

We will call k^{th} $(k = 1, ..., m)$ 'partial' coordinate of the i^{th} observation on the factorial axis of order α $(\alpha = 1, ..., q)$ to

$$c_{ki}^{\alpha} = \underline{u}'_{\alpha} Q_k \, \underline{x}_{ki} \qquad (k = 1, \ldots, m) \ ,$$

since it is obtained through $\underline{x}_{ki}$, the k^{th} 'partial' representation of the i^{th} observation. $\square$

Therefore, in the subspace $\mathcal{F}^q$ of $\mathbb{E}^p$ we may obtain for each observation i what we will call a 'global' representation and m 'partial' representations.

Keeping in mind that the subspace $\mathcal{F}^q$ of $\mathbb{E}^p$ is spanned by the Q-orthogonal vectors $\underline{u}_1, ..., \underline{u}_q$, columns of the matrix U, defined in 2.8, we now use the following definition.

Definition 4: The 'global' representation of the observation i in $\mathcal{F}^q$ will be its projection on $\mathcal{F}^q$ obtained through the i^{th} row of the 'global' matrix X. It is given by the i^{th} row of the matrix C, defined by

$$C = XQU \ , \tag{3.4}$$

where U is the matrix of the q column vectors $\underline{u}_{\alpha}$. The i^{th} row of C, $\underline{c}'_i$, then satisfies

$$\underline{c}_i = U'Q\underline{x}_i \ ,$$

and has elements c_i^{α} $(\alpha = 1, \ldots, q)$.

The m 'partial' representations of the observation i in $\mathcal{F}^q$ will be obtained through the projection on $\mathcal{F}^q$, of its m 'partial' representations given by the i^{th} row of each of the m matrices X_k. Such 'partial' representations in $\mathcal{F}^q$ will then be given by the rows of the matrix C_k, defined by

$$C_k = X_k Q_k U_k \quad (k = 1, \ldots, m) . \tag{3.5}$$

Each U_k $(k = 1, ..., m)$ is that submatrix of p_k rows of U which correspond to the k^{th} group of variables, or in other words, U_k is the restriction of U to $\mathbb{E}^{p_k}$. Then the i^{th} row of C_k will be the vector

$$\underline{c}'_{ki} = (U'_k Q_k \, \underline{x}_{ki})' \ , \quad (k = 1, \ldots, m)$$

and has coordinates c^{α}_{ki}. □

The n rows of the matrix C and of the m matrices C_k correspond respectively to the projections of the clouds of observation-points $N(I)$ and $N_k(I)$ $(k = 1, \ldots, m)$ on $\mathcal{F}^q$.

From the above definitions of the matrices C and C_k $(k = 1, ..., m)$, in (3.4) and (3.5), we observe that

$$C = \sum_{k=1}^{m} C_k \tag{3.6}$$

and consequently

$$\underline{c}_i = \sum_{k=1}^{m} \underline{c}_{ki} \ .$$

Definition 5: The columns of the matrix C,

$$\underline{c}^{\alpha} = XQ\underline{u}_{\alpha} \quad \in \quad \mathbb{E}^n \quad (\alpha = 1, \ldots, q) \tag{3.7}$$

will be called Generalized Canonical Variables or Principal Canonical Variables (of order α).

Consider the columns of the matrix C_k. They are given by

$$\underline{c}^{\,\alpha}_{\,k} = X_k Q_k \, \underline{u}_{k\alpha} \quad \in \quad \mathbb{E}^n \quad (\alpha = 1, \ldots, q) \ \ (k = 1, \ldots, m) \tag{3.8}$$

where $\underline{u}_{k\alpha}$ denotes the α^{th} column of U_k, i.e. the restriction of $\underline{u}_{\alpha}$ to $\mathbb{E}^{p_k}$. To the vectors $\underline{c}^{\alpha}_k$ we will call Canonical Variables (of order α). □

<u>Theorem 1</u>: For each α ($\alpha = 1, \ldots, q$), the vectors $\underline{c}_k^\alpha$ are proportional to the projection of $\underline{c}^\alpha$ on the subspace spanned by the columns of X_k, the proportionality constant being the inverse of the eigenvalue or inertia corresponding to the factorial axis of order α.

Proof: Note that

$$VQ = \begin{bmatrix} X_1'DXQ \\ \cdots \\ X_k'DXQ \\ \cdots \\ X_m'DXQ \end{bmatrix} .$$

Then, from the partitioning of (2.7), by blocks of p_k rows, each vector $\underline{u}_{k\alpha}$ ($k = 1, ..., m$) used in Definition 5, will satisfy

$$X_k'DXQ\underline{u}_\alpha = \lambda_\alpha \underline{u}_{k\alpha}$$

or, using (2.14) or (3.7)

$$\underline{u}_{k\alpha} = \frac{1}{\lambda_\alpha} X_k'DXQ\underline{u}_\alpha = \frac{1}{\lambda_\alpha} X_k'D\underline{c}^\alpha \ . \tag{3.9}$$

Therefore, from (3.8), and using (3.9) and (3.1)

$$\begin{aligned} \underline{c}_k^\alpha &= X_kQ_k\underline{u}_{k\alpha} \\ &= \frac{1}{\lambda_\alpha} X_kQ_kX_k'D\underline{c}^\alpha \\ &= \frac{1}{\lambda_\alpha} X_k(X_k'DX_k)^{-1}X_k'D\underline{c}^\alpha \\ &= \frac{1}{\lambda_\alpha} P_k\underline{c}^\alpha \end{aligned} \tag{3.10}$$

where $P_k = X_k(X_k'DX_k)^{-1}X_k'D$ is the D-orthogonal projector on $\mathbb{E}_k$, the subspace of $\mathbb{E}^n$ spanned by the columns of X_k. ∎

While each generalized canonical variable $\underline{c}^\alpha$ is a linear combination of all the $p = \sum_{k=1}^m p_k$ variables, each $\underline{c}_k^\alpha \in \mathbb{E}_k \subset \mathbb{E}^n$ is the D-orthogonal projection of $\underline{c}^\alpha$

on the corresponding subspace $\mathbb{E}_k$ of $\mathbb{E}^n$, divided by λ_α, the squared norm of $\underline{c}^\alpha$. This way, each canonical variable $\underline{c}_k^\alpha$ will be a linear combination of the variables represented by the columns of each X_k.

From the above definitions of the matrices C and C_k $(k = 1, \dots, m)$ in (3.4) and (3.5), and also of the vectors $\underline{c}^\alpha$ and $\underline{c}_k^\alpha$ $(k = 1, ..., m)$ in (3.7) and (3.8) and yet from (3.6) we can readily state the following Lemma.

Lemma 1:

$$\underline{c}^\alpha = \sum_{k=1}^{m} \underline{c}_k^\alpha \ . \ \blacksquare$$

The Generalized Canonical Variables $\underline{c}^\alpha$ are, for each α $(\alpha = 1, \dots, q)$, the variables that lie closest, "on an average", to all the m groups of variables (columns of the m submatrices X_k), i.e. closest "on an average" to all the m subspaces $\mathbb{E}_k$ spanned by the columns of each X_k. By 'closest "on an average"' we mean what is expressed in the following Theorem.

Theorem 2: For each order α $(\alpha = 1, \dots, q)$, $\underline{c}^\alpha$ is the variable that maximizes

$$\sum_{k=1}^{m} \cos^2 \theta(\underline{c}^\alpha,\ \mathbb{E}_k) = \sum_{k=1}^{m} \cos^2 \theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) \ , \tag{3.11}$$

where $\theta(.,.)$ represents the angle between the two arguments. ($\blacksquare$)

The proof of this theorem will be given later.

Really to prove Theorem 2 is equivalent to show that

$$\sum_{k=1}^{m} \cos^2 \theta(\underline{c}^\alpha,\ \mathbb{E}_k) \ = \ \lambda_\alpha \ = \| \ \underline{c}^\alpha \ \|_D^2 \qquad (\alpha = 1, \dots, q) \ ,$$

where λ_α is the quantity maximized at each step α $(\alpha = 1, \dots, q)$.

Since the multiple correlation coefficient of $\underline{c}^\alpha$ with the variables in each subspace $\mathbb{E}_k$ will be nothing but the square of the cosine of the angle between $\underline{c}^\alpha$ and $\mathbb{E}_k$ (i.e. between $\underline{c}^\alpha$ and its D-orthogonal projection on $\mathbb{E}_k$), we then have the following corollary.

Corollary 1: For each α ($\alpha = 1, \ldots, q$), the variables $\underline{c}^\alpha$ are the variables that maximize the sum of the squares of their multiple correlation coefficients with the variables in each of the m groups. ∎

The following is an interesting result since it relates the squared norm of the vectors $\underline{c}_k^\alpha$ with the angle between them and the corresponding vector $\underline{c}^\alpha$.

Lemma 2:

$$\| \underline{c}_k^\alpha \|_D^2 = \frac{1}{\lambda_\alpha}(\underline{c}^\alpha)' D \underline{c}_k^\alpha \;=\; \frac{1}{\lambda_\alpha} \cos^2 \theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) \ .$$

Proof: Since $\underline{c}_k^\alpha \in I\!\!E^n$, the norm $\| \underline{c}_k^\alpha \|^2$ is really $\| \underline{c}_k^\alpha \|_D^2$ since D is the metric used in $I\!\!E^n$. Thus

$$\begin{aligned}
\| \underline{c}_k^\alpha \|_D^2 &= (\underline{c}_k^\alpha)' D \underline{c}_k^\alpha \\
&= \underline{u}_{k\alpha}' Q_k X_k' D X_k Q_k \, \underline{u}_{k\alpha} && \big(\text{from } (3.8)\big) \\
&= \underline{u}_{k\alpha}' Q_k \, \underline{u}_{k\alpha} && \big(\text{from } (3.1)\big) \\
&= (\underline{c}^\alpha)' D X_k \frac{1}{\lambda_\alpha} (X_k' D X_k)^{-1} \frac{1}{\lambda_\alpha} X_k' D \underline{c}^\alpha && \big(\text{from } (3.9)\big) \\
&= \frac{1}{\lambda_\alpha^2} (\underline{c}^\alpha)' D X_k (X_k' D X_k)^{-1} X_k' D \underline{c}^\alpha \\
&= \frac{1}{\lambda_\alpha} (\underline{c}^\alpha)' D \underline{c}_k^\alpha \ . && \big(\text{from } (3.10)\big) \qquad (3.12)
\end{aligned}$$

But, on the other hand

$$\begin{aligned}
\cos^2 \theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) &= \frac{((\underline{c}^\alpha)' D \underline{c}_k^\alpha)^2}{\| \underline{c}^\alpha \|_D^2 \, \| \underline{c}_k^\alpha \|_D^2} \\
&= \frac{((\underline{c}^\alpha)' D \underline{c}_k^\alpha)^2}{\lambda_\alpha \frac{1}{\lambda_\alpha} (\underline{c}^\alpha)' D \underline{c}_k^\alpha}
\end{aligned}$$

from (3.12) above and also from the fact that

$$\| \underline{c}^\alpha \|_D^2 = (\underline{c}^\alpha)' D \underline{c}^\alpha = \underline{u}' Q X' D X Q \underline{u}_\alpha = \lambda_\alpha \ .$$

Then finally

$$\cos^2 \theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) = (\underline{c}^\alpha)' D \underline{c}_k^\alpha \ ,$$

or

$$\| \underline{c}_k^\alpha \|_D^2 = \frac{1}{\lambda_\alpha} \cos^2 \theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha)\ .\ \blacksquare$$

Lemma 3:

$$\sum_{k=1}^{m} \| \underline{c}_k^\alpha \|_D^2 = 1\ .$$

Proof: Using Lemma 1 and an intermediary result from Lemma 2, we can write

$$\begin{aligned}\sum_{k=1}^{m} \| \underline{c}_k^\alpha \|_D^2 &= \sum_{k=1}^{m} \frac{1}{\lambda_\alpha}(\underline{c}^\alpha)' D \underline{c}_k^\alpha \;=\; \frac{1}{\lambda_\alpha}(\underline{c}^\alpha)' D \left(\sum_{k=1}^{m} \underline{c}_k^\alpha \right) \\ &= \frac{1}{\lambda_\alpha}(\underline{c}^\alpha)' D \underline{c}^\alpha \;=\; \frac{1}{\lambda_\alpha} \| \underline{c}^\alpha \|_D^2 \;=\; \frac{1}{\lambda_\alpha}\lambda_\alpha \;=\; 1\ .\ \blacksquare\end{aligned}$$

Now we are in a position to prove Theorem 2, as promised earlier.

Proof of Theorem 2:

From Lemma 2 we can write

$$\begin{aligned}\textstyle\sum_{k=1}^{m} \cos^2 \theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) &= \lambda_\alpha \textstyle\sum_{k=1}^{m} \| \underline{c}_k^\alpha \|_D^2 \\ &= \lambda_\alpha\ , \qquad\qquad \text{(using Lemma 3)}\end{aligned} \tag{3.13}$$

where λ_α is the quantity maximized in each step α ($\alpha = 1, \ldots, q$). $\blacksquare$

Thus, in each step α ($\alpha = 1, \ldots, q$) the variable $\underline{c}^\alpha$, (eigenvector of $WD = XQX'D$ associated with the eigenvalue λ_α) is the variable that maximizes the sum of the cosines of its angles with each of the subspaces $I\!E_k$. This means that

$$\lambda_1 \;=\; \sum_{k=1}^{m} \cos^2 \theta(\underline{c}^1, I\!E_k) \;=\; \max_{\underline{a} \in I\!E^n} \sum_{k=1}^{m} \cos^2 \theta(\underline{a}, I\!E_k)$$

and that

$$\lambda_2 \;=\; \sum_{k=1}^{m} \cos^2 \theta(\underline{c}^2, I\!E_k) \;=\; \max_{\substack{\underline{a} \in I\!E^n \\ \underline{a}' D \underline{c}^1 = 0}} \sum_{k=1}^{m} \cos^2 \theta(\underline{a}, I\!E_k)$$

and so on, for $\lambda_3, \ldots, \lambda_q$, taking into account the constraints of orthogonality of the vectors $\underline{c}^\alpha$.

Definition 6:

To λ_α we will call generalized canonical correlation of order α. □

From Theorem 2 and Lemma 2 we have the following Corollary.

Corollary 2:

$$\| \underline{c}_k^\alpha \|_D^2 = \frac{\cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha)}{\sum_{k=1}^m \cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha)}\ . \ \blacksquare$$

We now summarize some of the most interesting properties of the vectors $\underline{c}^\alpha$ and $\underline{c}_k^\alpha$ $(\alpha = 1, \ldots, q)$.

1) $\underline{c}^\alpha = \sum_{k=1}^m \underline{c}_k^\alpha$
2) $\| \underline{c}^\alpha \|_D^2 = \lambda_\alpha = \sum_{k=1}^m \cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha)$
3) $\underline{c}^{\alpha\prime} D \underline{c}_k^\alpha = \cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha)$
4) $\| \underline{c}_k^\alpha \|_D^2 = \frac{1}{\lambda_\alpha}\underline{c}^{\alpha\prime} D \underline{c}_k^\alpha = \frac{1}{\lambda_\alpha}\cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) = \frac{\cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha)}{\sum_{k=1}^m \cos^2\theta(\underline{c}^\alpha, \underline{c}_k^\alpha)}$
5) $\sum_{k=1}^m \| \underline{c}_k^\alpha \|_D^2 = 1$.

We also state now some results about the eigenvalues λ_α. Directly from (3.13), the last result in the proof of Theorem 2.

Corollary 3:

$$0 \le \lambda_\alpha \le m \qquad (\alpha = 1, \ldots, q)\ . \ \blacksquare$$

Theorem 3: If all the subspaces $\mathbb{E}_k$ coincide then $\lambda_\alpha = m$, $(\alpha = 1, \ldots, s = rank(V_{kk}))$, and conversely.

Proof: If the m subspaces $\mathbb{E}_k$ coincide, we have

$$\cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) = 1 \quad , \quad \forall k = 1, \ldots, m\ .$$

Then, what we have to show is that

$$\forall \alpha = 1, ..., q: \quad \cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) = 1 \quad (\forall k = 1, ..., m) \iff \lambda_\alpha = m\ .$$

The proof of

$$\cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) = 1 \quad (\forall k = 1,...,m) \implies \lambda_\alpha = m \qquad (\forall \alpha = 1,...,q)\ ,$$

follows from

$$\cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) = 1 \quad (\forall k = 1,...,m) \implies \sum_{k=1}^{m} \cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) = \lambda_\alpha = m \quad (\forall \alpha = 1,...,q)\ ,$$

using Theorem 2.

Now we prove

$$\forall \alpha = 1,...,q: \qquad \lambda_\alpha = m \implies \cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) = 1\ , \quad (\forall k = 1,...,m)\ .$$

Using Lemma 2,

$$\forall \alpha = 1,...,q: \qquad \lambda_\alpha = m \implies \| \underline{c}_k^\alpha \|_D^2 = \frac{1}{m}\cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha)$$

but using Lemma 3, since $\sum_{k=1}^{m} \| \underline{c}_k^\alpha \|_D^2 = 1$, then we can write

$$\sum_{k=1}^{m} \frac{1}{m}\cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) = \frac{1}{m}\sum_{k=1}^{m}\cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) = 1\ ,$$

which implies

$$\cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) = 1 \qquad \forall k = 1,\dots,m$$

because $\max_{\forall \alpha,k}\ \cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) = 1$ and we need $\sum_{k=1}^{m}\cos^2\theta(\underline{c}^\alpha,\ \underline{c}_k^\alpha) = m$.

But since, $\sum_{\alpha=1}^{p}\lambda_\alpha = tr(VQ) = p$, all the $s = rank(V_{kk})$ eigenvalues λ_α are then equal to m. (If we assume all the matrices V_{kk} to be positive-definite then we do not have any λ_α eigenvalues equal to zero, but if one or more of the matrices V_{kk} are only semipositive-definite then we will have some of the eigenvalues λ_α equal to zero, which really should not be taken into account.) ■

If only $s^* < s$ dimensions are coincident, then $\lambda_1 = \ldots = \lambda_{s^*} = m > \lambda_{s^*+1} > \ldots > \lambda_s$.

<u>Theorem 4</u>: If all the subspaces $\mathbb{E}_k$ are mutually orthogonal then $\lambda_\alpha = 1$ ($\forall \alpha = 1, \ldots, p$), and conversely.

Proof: Mutual orthogonality of all the m subspaces $\mathbb{E}_k$ is equivalent to orthogonality of all $\underline{c}_k^\alpha$ in k (i.e., for $k = 1, \ldots, m$) for a given α. So, all we have to show is that for $\alpha = 1, \ldots, p$,

$$\text{orthogonality of all } \underline{c}_k^\alpha \text{ in } k \text{ (for a given } \alpha) \iff \lambda_\alpha = 1.$$

We first prove that

$$\text{orthogonality of all } \underline{c}_k^\alpha \text{ in } k \text{ for a given } \alpha \implies \lambda_\alpha = 1 \text{ (for that given } \alpha).$$

From Lemma 1 and Lemma 2 we can write

$$\begin{aligned} \| \underline{c}_k^\alpha \|_D^2 &= \frac{1}{\lambda_\alpha} (\underline{c}^\alpha)' D \underline{c}_k^\alpha = \frac{1}{\lambda_\alpha} \left(\sum_{k=1}^m \underline{c}_k^\alpha \right)' D \underline{c}_k^\alpha \\ &= \frac{1}{\lambda_\alpha} \| \underline{c}_k^\alpha \|_D^2 \qquad \text{(by orthogonality of the } \underline{c}_k^\alpha \text{ in } k) \ , \end{aligned}$$

which implies $\lambda_\alpha = 1$, for that α.

Now we prove

$$\lambda_\alpha = 1 (\text{ for a given } \alpha) \implies \text{orthogonality of all } \underline{c}_k^\alpha \text{ in } k \text{ (for that given } \alpha) \ .$$

Using Lemma 3 and $\| \underline{c}^\alpha \|_D^2 = \lambda_\alpha$, we can write, for a given α,

$$\lambda_\alpha = 1 \implies \| \underline{c}^\alpha \|_D^2 = 1 = \sum_{k=1}^m \| \underline{c}_k^\alpha \|_D^2$$

or, using Lemma 1,

$$\sum_{k=1}^m \| \underline{c}_k^\alpha \|_D^2 = \left\| \sum_{k=1}^m \underline{c}_k^\alpha \right\|_D^2 .$$

This implies that the $\underline{c}_k^\alpha$ are orthogonal in k, using a multidimensional version of Pythagoras' Theorem.

If all the m subspaces $\mathbb{E}_k$ are mutually orthogonal, then the above two implications are verified for all the dimensions or directions in $\mathbb{E}^n$, but since $\sum_{\alpha=1}^p \lambda_\alpha = p$ (we assume $n > p$) and then $\forall \alpha, \lambda_\alpha = 1$, we would have p eigenvalues λ_α, all equal to 1. ∎

If only a few dimensions are orthogonal, then we will have only the corresponding eigenvalues equal to 1.

3.8 Another look at the Generalized Canonical Analysis: the dual analysis.

The dual analysis is an alternative way of approaching this problem from the variables space rather than the observations space.

Now we look for the variables in $I\!E^n$ that maximize the sum of their multiple correlation coefficient with the variables in the m groups of variables.

Based on the results in section 2.13 of Chapter 2, we restate the generic results obtained there, now applied to this particular analysis, under the form of the following two definitions.

Definition 7: The factorial axes in the dual analysis are the directions $\Delta\underline{v}^\alpha$ in $I\!E^n$, where $\underline{v}^\alpha$ is the D-unitary eigenvector of WD, i.e.

$$WD\underline{v}^\alpha = \lambda_\alpha \underline{v}^\alpha$$

subject to the restrictions

$$\underline{v}^{\alpha\prime} D\underline{v}^\alpha = 1 \quad \text{and} \quad \underline{v}^{\alpha\prime} D\underline{v}^{\alpha-i} = 0 \quad (\alpha = 1, ..., q) \ \ (i = 1, ..., \alpha - 1) \tag{3.14}$$

where $W = XQX'$, or

$$WD = XQX'D = \sum_{k=1}^{m} \left(X_k Q_k X_k' D \right) = \sum_{k=1}^{m} W_k D = \sum_{k=1}^{m} P_k$$

and $P_k = W_k D = X_k Q_k X_k' D = X_k (X_k' D X_k)^{-1} X_k' D$ is the D-orthogonal projector on $I\!E_k$, as defined in (3.10). The vectors $\underline{v}^\alpha$ verify

$$\underline{v}^\alpha = \frac{1}{\lambda_\alpha} \underline{c}^\alpha \ . \ \square$$

Definition 8: The vectors $\underline{v}^\alpha$ will be the D-unitary vectors that maximize for each α,

$$\begin{aligned} In_{\underline{v}^\alpha}(J) &= \underline{v}^{\alpha\prime} DWD\underline{v}^\alpha \\ &= \underline{v}^{\alpha\prime} D \left(\sum_{k=1}^{m} W_k D \right) \underline{v}^\alpha = \underline{v}^{\alpha\prime} D \left(\sum_{k=1}^{m} P_k \right) \underline{v}^\alpha \ . \ \square \end{aligned} \tag{3.15}$$

Theorem 5: To maximize (3.15) subject to (3.14) is equivalent to maximizing

$$\sum_{k=1}^{m} \cos^2 \theta(\underline{v}^\alpha,\ \mathbb{E}_k) \ . \tag{3.16}$$

Proof: To maximize (3.15) subject to (3.14) is equivalent to maximizing

$$\frac{(\underline{v}^\alpha)' DWD\underline{v}^\alpha}{(\underline{v}^\alpha)' D\underline{v}^\alpha} = \frac{(\underline{v}^\alpha)' D \left(\sum_{k=1}^{m} P_k \right) \underline{v}^\alpha}{(\underline{v}^\alpha)' D\underline{v}^\alpha} = \sum_{k=1}^{m} \frac{(\underline{v}^\alpha)' DP_k\underline{v}^\alpha}{(\underline{v}^\alpha)' D\underline{v}^\alpha}$$

where

$$\frac{(\underline{v}^\alpha)' DP_k\underline{v}^\alpha}{(\underline{v}^\alpha)' D\underline{v}^\alpha}$$

is the square of the cosine of the angle between $\underline{v}^\alpha$ and its D-orthogonal projection on $\mathbb{E}_k$. Observe that

$$\frac{(\underline{v}^\alpha)' DP_k\underline{v}^\alpha}{(\underline{v}^\alpha)' D\underline{v}^\alpha} = \cos^2 \theta(\underline{v}^\alpha,\ P_k\underline{v}^\alpha) \ ,$$

because

$$\begin{aligned} \cos^2 \theta(\underline{v}^\alpha,\ P_k\underline{v}^\alpha) &= \frac{((\underline{v}^\alpha)' DP_k\underline{v}^\alpha)^2}{(\underline{v}^\alpha)' D\underline{v}^\alpha \ (\underline{v}^\alpha)' P_k' DP_k\underline{v}^\alpha} \\ &= \frac{((\underline{v}^\alpha)' DP_k\underline{v}^\alpha)^2}{(\underline{v}^\alpha)' D\underline{v}^\alpha \ (\underline{v}^\alpha)' DP_k'\underline{v}^\alpha} \\ &= \frac{(\underline{v}^\alpha)' DP_k\underline{v}^\alpha}{(\underline{v}^\alpha)' D\underline{v}^\alpha} \end{aligned}$$

since

$$\begin{aligned} P_k' DP_k &= DX_k(X_k' DX_k)^{-1} X_k' DX_k (X_k' DX_k)^{-1} X_k' D = \\ &= DX_k(X_k' DX_k)^{-1} X_k' D = DP_k \end{aligned}$$

as expected from the usual idempotent property of any projector. ■

The variables $\underline{v}^{\alpha}$ will then be the variables that maximize the sum over k $(k = 1, ..., m)$ of the square of the cosine of the angle between $\underline{v}^{\alpha}$ and each of the m subspaces $\mathbb{E}_k$. Given the equivalence between cosine and correlation we may then say that, for each order α the vectors $\underline{v}^{\alpha}$ are the ones that maximize the sum over k of the square of the multiple correlation coefficients of $\underline{v}^{\alpha}$ with the variables in each set of variables $\underline{x}_{(k)}$ $(k = 1, ..., m)$.

Since (3.16) is really equivalent to (3.11), in Theorem 2, we see that, for each order α $(\alpha = 1, ..., q)$, the vectors $\underline{c}^{\alpha}$ (or $\underline{v}^{\alpha}$) may be seen as the principal components of the m vectors $\underline{c}_k^{\alpha}$ (or $P_k\underline{v}^{\alpha}$). This 'property' was also noticed by Kettenring (1971) but, in this generalized analysis it is really not an intrinsic property of the vectors $\underline{c}^{\alpha}$ (or $\underline{v}^{\alpha}$) but only a consequence of the way the vectors $\underline{c}_k^{\alpha}$ (or $P_k\underline{v}^{\alpha}$) are defined (see (3.8) and (3.10)). The vectors $\underline{c}_k^{\alpha}$ defined either by (3.8) or (3.10) are obtained after the vectors $\underline{u}_{\alpha}$ or the corresponding vectors $\underline{c}^{\alpha}$ have been obtained. When each set of variables has one only variable then the GCA reduces to the Principal Components Analysis (Principal Components Analysis on the correlation matrix we should note) and then we may see the above as an intrinsic property of the vectors $\underline{c}^{\alpha}$ which are then the principal components.

3.9 Usefulness of the approach outlined.

In sections 3.2–3.8 of this chapter, we provided an abstract analysis of relationships among variables in general. This type of multivariate analysis is carried out in more or less similar ways in Principal Components Analysis, Discriminant Analysis, Correspondence Analysis and Canonical Correlation Analysis, corresponding to different structures of the data matrix X and consequently to different metrics Q. All these analysis have some similarities and some specific details and objectives. The abstract approach provides a unified mathematical background for these analyses. We will now turn to more specific aspects. For this we need to assume that the data matrix X is obtained as a sample from some suitable multivariate statistical distribution (and the obvious choice is the popular multivariate normal distribution).

3.10 Note about the use of generalized inverses in Q.

So far we have implicitly assumed that the matrices $V_{kk} = X_k' D X_k$ are of full rank. Such assumption was necessary to be able to define the diagonal blocks of Q, blocks $Q_k = V_{kk}^{-1}$.

However V_{kk} may be singular when there are hidden linear relationships in the k^{th} set of variables.

Let us assume that

$$rank(V_{kk}) = q_k < p_k \ .$$

Usually the rank deficiency $p_k - q_k$ is small. One would deal with this problem by deleting rows and columns of V_{kk}, or equivalently the corresponding columns from X_k. However this loses the determinacy of the problem since it would not then be clear which variables could preferably be deleted from V_{kk}.

A better approach is to use a generalized inverse of V_{kk} (Rao and Mitra, 1971; Boullion and Odell, 1971). Let V_{kk}^- be a generalized inverse of V_{kk} satisfying

$$V_{kk}^- V_{kk} V_{kk}^- = V_{kk}^- \tag{3.17}$$

and

$$V_{kk} V_{kk}^- V_{kk} = V_{kk} \ . \tag{3.18}$$

There is one fine way to get such a good generalized inverse. Such a V_{kk}^- may be obtained from the eigenvalue-eigenvector decomposition of V_{kk}.

Let

$$V_{kk}\,\underline{u}^*_{k\alpha} = \lambda_\alpha\,\underline{u}^*_{k\alpha}$$

with

$$\underline{u}^{*\prime}_{k\alpha}\underline{u}^*_{k\alpha} = 1 \qquad \text{and} \qquad \underline{u}^{*\prime}_{k\alpha}u^*_{k(\alpha-i)} = 0 \qquad (\alpha = 1, ..., q_k) \;\; (i = 1, ..., \alpha - 1) \ ,$$

or, alternatively,

$$V_{kk}U^*_k = U^*_k\Lambda^*$$

where $U^*_k = \left[\underline{u}_{k1}|\underline{u}_{k2}|...|\underline{u}_{kq_k}\right]$ and $\Lambda^* = diag(\lambda_1, \lambda_2, ..., \lambda_{q_k})$. Since the q_k eigenvectors $\underline{u}^*_{k\alpha}$ form an orthonormal basis for $I\!E^{p_k}$, since they correspond to all the non-null eigenvalues of V_{kk}, then we may write (see also the discussion in Appendix 2.C)

$$V_{kk} = U^*_k\Lambda^*U^{*'}_k \ . \tag{3.19}$$

Then from (3.19) we define

$$V^-_{kk} = U^*_k\Lambda^{*-1}U^{*'}_k \ . \tag{3.20}$$

It is easy to see that this V^-_{kk} satisfies (3.17) and (3.18).

Then our D-orthogonal projectors may be defined as

$$P_k = X_kV^-_{kk}X'_kD = X_kU^*_k\Lambda^{*-1}U^{*'}_kX'_kD \ . \tag{3.21}$$

That they are still the desired projectors may be easily seen from basic notions in measure theory and matrix applications. If we consider any matrix in (3.21) as an application matrix, we have

$$\begin{array}{lllll} X_k & : & I\!E^{p^*_k} \longrightarrow I\!E^n & & (\text{and then } \; X'_k \; : \; I\!E^{n^*} \longrightarrow I\!E^{p_k}) \\ U^*_k & : & I\!E^{q_k} \longrightarrow I\!E^{p_k} & & (\text{and then } \; U^{*'}_k \; : \; I\!E^{p_k} \longrightarrow I\!E^{q_k}) \\ \Lambda^* & : & I\!E^{q_k} \longrightarrow I\!E^{q_k} & & (\text{and then } \; \Lambda^{*-1} \; : \; I\!E^{q_k} \longrightarrow I\!E^{q_k}) \\ D & : & I\!E^n \longrightarrow I\!E^{n*} & . & \end{array}$$

Then we have

$$V^-_{kk} = I_{p_k}U^*_k\Lambda^{*-1}U^{*'}_k \;:\; I\!E^{p_k} \to I\!E^{q_k} \to I\!E^{q_k} \to I\!E^{p_k} \to I\!E^{p_k} \to I\!E^{p^*_k} \;\equiv\; I\!E^{p_k} \to I\!E^{p^*_k} \ ,$$

the expected result since now V_{kk}^{-} takes the place of the metric in $I\!E^{p_k}$. (We should note that since $I_{p_k} : I\!E^{p_k} \rightarrow I\!E^{p_k^*}$ was the metric used in $I\!E^{p_k}$ in the eigenvalue-eigenvector decomposition of V_{kk} we should really write $V_{kk} = U_k^* \Lambda_k^* U_k^{*'} I'_{p_k}$ and $V_{kk}^{-} = I_{p_k} U_k^* \Lambda_k^{*-1} U_k^{*'}$ what in matricial terms comes to the same expressions as the ones used but it makes a difference when writing the application matrices.)

Then

$$P_k = X_k U_k^* \Lambda^{*-1} U_k^{*'} X_k' D \;:\; I\!E^n \longrightarrow I\!E^n \;.$$

And that P_k defined this way is still idempotent is easily seen from

$$\begin{aligned} P_k P_k &= X_k U_k^* \Lambda^{*-1} U_k^{*'} X_k' D X_k U_k^* \Lambda^{*-1} U_k^{*'} X_k' D \\ &= X_k U_k^* \Lambda^{*-1} U_k^{*'} U_k^* \Lambda^* U_k^{*'} U_k^* \Lambda^{*-1} U_k^{*'} X_k' D \\ &= X_k U_k^* \Lambda^{*-1} U_k^{*'} X_k' D = P_k \;. \end{aligned}$$

Thus whenever V_{kk}^{-1} does not exist we will use V_{kk}^{-} given by (3.20) instead.

3.11 GCA in the case $m = 2$: the usual Canonical Analysis.

In the case $m = 2$ the expression

$$VQ\underline{u}_\alpha \;=\; \lambda_\alpha \underline{u}_\alpha$$

reduces to

$$\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \begin{bmatrix} V_{11}^{-1} & 0 \\ 0 & V_{22}^{-1} \end{bmatrix} \begin{bmatrix} \underline{u}_{1\alpha} \\ \underline{u}_{2\alpha} \end{bmatrix} \;=\; \lambda_\alpha \begin{bmatrix} \underline{u}_{1\alpha} \\ \underline{u}_{2\alpha} \end{bmatrix}$$

which implies

$$\begin{cases} \underline{u}_{1\alpha} + V_{12} V_{22}^{-1} \underline{u}_{2\alpha} \;=\; \lambda_\alpha \underline{u}_{1\alpha} \\ V_{21} V_{11}^{-1} \underline{u}_{1\alpha} + \underline{u}_{2\alpha} \;=\; \lambda_\alpha \underline{u}_{2\alpha} \;, \end{cases}$$

or that

$$\begin{cases} V_{12}V_{22}^{-1}\underline{u}_{2\alpha} &= (\lambda_\alpha - 1)\underline{u}_{1\alpha} \\ V_{21}V_{11}^{-1}\underline{u}_{1\alpha} &= (\lambda_\alpha - 1)\underline{u}_{2\alpha} \ . \end{cases} \tag{3.22}$$

Then

$$\begin{cases} V_{12}V_{22}^{-1}V_{21}V_{11}^{-1}\,\underline{u}_{1\alpha} &= (\lambda_\alpha - 1)^2\,\underline{u}_{1\alpha} \\ V_{21}V_{11}^{-1}V_{12}V_{22}^{-1}\,\underline{u}_{2\alpha} &= (\lambda_\alpha - 1)^2\,\underline{u}_{2\alpha} \ . \end{cases} \tag{3.23}$$

These are the standard relations in the usual Canonical Analysis of two vector variables or two sets of variables where $(\lambda_\alpha - 1)^2$ are the square of the sample canonical correlations (Kshirsagar, 1972).

A little algebra will show that the vectors $\underline{u}_{1\alpha}$ and $\underline{u}_{2\alpha}$ in (3.23) are now orthogonal in α as well as the corresponding vectors $\underline{c}_1^\alpha$ and $\underline{c}_2^\alpha$. This means, that in this case where $m = 2$, and opposite to what happens when $m > 2$,

$$\underline{u}_{1\alpha}'Q_1\,\underline{u}_{1(\alpha-i)} \;=\; \underline{u}_{2\alpha}'Q_2\,\underline{u}_{2(\alpha-i)} \;=\; 0 \qquad (\alpha = 1, ..., q) \;\; (i = 1, ..., \alpha)$$

as well as

$$(\underline{c}_1^\alpha)'D\underline{c}_1^{\alpha-i} \;=\; (\underline{c}_2^\alpha)'D\underline{c}_2^{\alpha-i} \;=\; 0 \qquad (\alpha = 1, ..., q) \;\; (i = 1, ..., \alpha) \ .$$

If we denote by μ_α^2 the square of the sample canonical correlation of order α we then have the following relationship

$$\mu_\alpha^2 \;=\; (\lambda_\alpha - 1)^2 \;\Rightarrow\; \pm\mu_\alpha = \lambda_\alpha - 1 \;\Rightarrow\; \begin{cases} \lambda_\alpha = 1 + \mu_\alpha \\ \lambda_{\alpha'} = 1 - \mu_\alpha \end{cases}$$

what shows that if for example $p_2 < p_1$ and all the p_2 sample canonical correlations are different from zero then we will have p_2 eigenvalues $\lambda_\alpha = 1 + \mu_\alpha$ $(\alpha = 1, ..., p_2)$, other p_2 eigenvalues $\lambda_{\alpha'} = 1 - \mu_\alpha$ $(\alpha' = p_1 + 1, ..., p_2 + p_1)$ and $p_1 - p_2$ structural eigenvalues $\lambda_\alpha = 1$ $(\alpha = p_2 + 1, ..., p_1)$.

3.12 A glimpse at estimation.

Though estimation is not the main topic of this dissertation, we will consider this issue briefly, as it will be useful to understand the models introduced in Chapter 4 and the meaning of the vectors $\underline{a}^{\alpha}$ in Chapter 2, 2.7.

Consider the Canonical Analysis model

$$\left\{\begin{array}{rcl} \beta'_{2(1)}\underline{x}_{(2)} & = & \Psi\beta'_{1(2)}\underline{x}_{(1)} + \underline{\mathcal{E}}_2 \\ \beta'_{1(2)}\underline{x}_{(1)} & = & \Psi\beta'_{2(1)}\underline{x}_{(2)} + \underline{\mathcal{E}}_1 \end{array}\right. \tag{3.24}$$

where $\underline{x}_{(1)}$ and $\underline{x}_{(2)}$ are two vectors of respectively p_1 and p_2 variables and $\beta_{1(2)}$ $(p_1 \times q)$ and $\beta_{2(1)}$ $(p_2 \times q)$ are two parameter matrices corresponding to the q population canonical correlations taken into account in the model, Ψ is a diagonal matrix of the first q $(q \leq min(p_1, p_2))$ population canonical correlations, and $\underline{\mathcal{E}}_1$ and $\underline{\mathcal{E}}_2$ are two random error vectors with distributions

$$N_q(\underline{0}, I_q - \Psi^2) \ .$$

From (3.24), when estimating the parameter matrices involved in and using sample quantities we get the following equations

$$\left\{\begin{array}{rcl} X_2\hat{\beta}_{2(1)} & = & X_1\hat{\beta}_{1(2)}\hat{\Psi} \\ X_1\hat{\beta}_{1(2)} & = & X_2\hat{\beta}_{2(1)}\hat{\Psi} \ . \end{array}\right. \tag{3.25}$$

We may notice that premultiplying the first equation in (3.25) by $P_1 = X_1Q_1X_1'D$ and the second by $P_2 = X_2Q_2X_2'D$ (and taking note that $Q_1 = (X_1'DX_1)^{-1} = V_{11}^{-1}$ and $Q_2 = (X_2'DX_2)^{-1} = V_{22}^{-1}$) we get

$$\left\{\begin{array}{rcl} P_1X_2\hat{\beta}_{2(1)} & = & X_1\hat{\beta}_{1(2)}\hat{\Psi} \\ P_2X_1\hat{\beta}_{1(2)} & = & X_2\hat{\beta}_{2(1)}\hat{\Psi} \ . \end{array}\right. \tag{3.26}$$

But from (3.22), premultiplying the first equation by X_1Q_1 and the second one by X_2Q_2 we obtain

$$\left\{\begin{array}{rcl} X_1Q_1V_{12}Q_2U_2 & = & X_1Q_1U_1\mathcal{M}^{1/2} \\ X_2Q_2V_{21}Q_1U_1 & = & X_2Q_2U_2\mathcal{M}^{1/2} \end{array}\right. \tag{3.27}$$

where $\mathcal{M}^{1/2} = diag(\sqrt{\mu_\alpha})$ $(\alpha = 1, ..., q)$ where $\mu_\alpha = (\lambda_\alpha - 1)^2$. But if we note that $X_1 Q_1 V_{12} = P_1 X_2$, $X_2 Q_2 V_{21} = P_2 X_1$ and if we take

$$\begin{aligned} \hat{\beta}_{2(1)} &= Q_2 U_2 \\ \hat{\beta}_{1(2)} &= Q_1 U_1 \end{aligned} \tag{3.28}$$

and

$$\hat{\Psi} = \mathcal{M}^{1/2} \tag{3.29}$$

then (3.27) holds the equations (3.26). We may notice that the columns of the matrices $\hat{\beta}_{2(1)} = Q_2 U_2$ and $\hat{\beta}_{1(2)} = Q_1 U_1$ are the vectors $\underline{a}^\alpha$ $(\alpha = 1, ..., q)$ defined in 2.7 (Chapter 2). Equations (3.28) and (3.29) will define the estimates for our Canonical Analysis model in (3.24).

The distribution of the estimates $\hat{\beta}_{2(1)}$ and $\hat{\beta}_{1(2)}$ has been studied, see for example Kshirsagar (1972).

From (3.28) we may write

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_{1(2)} \\ \hat{\beta}_{2(1)} \end{bmatrix} = \begin{bmatrix} Q_1 U_1 \\ Q_2 U_2 \end{bmatrix} = QU$$

where $Q = bdiag(Q_1, Q_2)$ and $U = \begin{bmatrix} U_1' & U_2' \end{bmatrix}'$.

When we consider the overall GCA model without subdividing it in submodels the interpretation becomes more complex when we try to relate the model to the m sets of variables but we may then think of it as a model for the variance-covariance matrix of the $p = \sum_{k=1}^{m} p_k$ variables and we may then still consider the matrix QU as a parameter matrix. Further study in this area is desirable.

CHAPTER 4

GENERALIZED CANONICAL ANALYSIS - A DISTRIBUTIONAL APPROACH

4.1 Introduction.

In Chapter 3 we provided an abstract analysis of the relationships among m groups of variables in general. This type of multivariate analysis is carried out in more or less similar ways in Principal Components Analysis, Discriminant Analysis, Correspondence Analysis and Canonical Correlation Analysis. All these analyses have some similarities and some specific details and objectives. The abstract approach provides a kind of unified mathematical background for these analyses. We will now turn to more specific aspects. For this we need to assume that the data matrix X is obtained as a sample from some suitable multivariate statistical distribution and the obvious choice is the popular multivariate normal distribution.

Let

$$\underline{x} = \left[\ \underline{x}'_{(1)}, \ \ \underline{x}'_{(2)}, \ \ \ldots \ \ , \ \ \underline{x}'_{(k)}, \ \ \ldots \ \ , \ \ \underline{x}'_{(m)} \ \right]' \sim N_p(\underline{\mu}, \ \Sigma) \ , \tag{4.1}$$

then

$$\underline{x}_{(k)} \ \sim \ N_{p_k}(\underline{\mu}_k, \ \Sigma_{kk}) \qquad (k = 1, \ldots, m)$$

where

$$\underline{\mu} = \left[\ \underline{\mu}_1', \ \ \underline{\mu}_2', \ \ \dots \ , \ \ \underline{\mu}_k', \ \ \dots \ , \ \ \underline{\mu}_m' \ \right]' \ ,$$

and

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1k} & \dots & \Sigma_{1m} \\ \Sigma_{21} & \Sigma_{22} & \dots & \Sigma_{2k} & \dots & \Sigma_{2m} \\ \vdots & \vdots & & \vdots & & \vdots \\ \Sigma_{k1} & \Sigma_{k2} & \dots & \Sigma_{kk} & \dots & \Sigma_{km} \\ \vdots & \vdots & & \vdots & & \vdots \\ \Sigma_{m1} & \Sigma_{m2} & \dots & \Sigma_{mk} & \dots & \Sigma_{mm} \end{bmatrix} . \tag{4.2}$$

In this chapter we will also see that the product of all the eigenvalues λ_α of VQ, namely

$$|VQ| = \prod_{\alpha=1}^{p} \lambda_\alpha \ ,$$

where V, Q and λ_α are the same as in Chapter 3, is the $(2/n)^{\text{th}}$ power of the Likelihood Ratio Test statistic for testing the hypothesis

$$H_0 \ : \ \underline{x}_{(k)} \text{ is independent of } \underline{x}_{(k')} \quad \text{for} \quad k \neq k' \ , \quad k, k' \in \{1, ..., m\} \ , \tag{4.3}$$

[this hypothesis can also be written as

$$H_0 \ : \ \Sigma_{kk'} = 0 \ , \forall k \neq k' \ , \quad k, k' \in \{1, ..., m\}$$

or as

$$H_0 \ : \ \Sigma = bdiag(\Sigma_{kk}) \quad (k = 1, ..., m) \ ]$$

against

$$H_1 \ : \ \text{there exists } \Sigma_{kk'} \neq 0 \ \text{ for } \ k \neq k' \qquad (k, k' \in \{1, ..., m\}) \ .$$

An asymptotic approximation to the distribution of this statistic will be derived in Chapter 5.

The above hypothesis H_0 will be split into meaningful subhypotheses. Test statistics will be derived for these subhypotheses and their distributions will be shown to be related to the Wilks' lambda distribution.

We will then proceed to show the relationship of the GCA model with a number of well known multivariate linear models, and other models not studied in the literature, so far.

4.2 Some notation concerning the normal and Wishart distributions.

We will assume throughout this chapter that $\underline{x}$ is a vector of p components, with a multivariate normal distribution as specified by (4.1) and (4.2).

Let X^* (an $n{\times}p$ matrix) denote a random sample of size n from this distribution. Then each row of X^* has mean $\underline{\mu}'$ and the rows are independently distributed and have a common variance-covariance matrix Σ. This will be denoted by

$$(vec(X^*)) \sim N_{np}\left(vec(M),\ \Sigma \otimes I_n\right) \ ,$$

(where $vec(X^*)$ stands for a vector obtained by stacking the columns of X^* one below the other) or even as

$$X^* \sim N_{np}(M,\ \Sigma \otimes I_n)$$

sometimes used as an abbreviation, with $M = E_{n1}\underline{\mu}'$, where E_{ab} denotes an $a{\times}b$ matrix of unit elements.

The $p{\times}p$ matrix $B^{**} = X^{*\prime}X^*$ is then said to have a non-central Wishart distribution with n degrees of freedom and parameter matrix Σ. The density function of B^{**} is then

$$f(B^{**}) \quad = \quad \frac{|B^{**}|^{(n-p-1)/2}\ {}_0F_1\Big(\frac{n}{2};\ \frac{1}{4}\Sigma^{-1}M'M\Sigma^{-1}B^{**}\Big)}{2^{np/2}\ \Gamma_p\Big(\frac{n}{2}\Big)\ \ |\Sigma|^{n/2}}\ e^{-\frac{1}{2}\ tr\left[\Sigma^{-1}(B^{**}-M'M)\right]}$$

$$(B^{**} > 0) \tag{4.4}$$

where

$${}_0F_1\left(\frac{n}{2};\ \frac{1}{4}\Sigma^{-1}M'M\Sigma^{-1}B^{**}\right) = \frac{\Gamma_m\Big(\frac{n}{2}\Big)}{2^m\pi^{np/2}} \int_{H_1\in V_{m,n}} e^{tr\Sigma^{-1}M'H_1T}\ H_1'\ dH_1$$

is an hypergeometric function of matrix argument and where $V_{m,n}$ is the Stiefel manifold of $n{\times}p$ matrices with orthonormal columns (James, 1954, 1955, 1964; see also Constantine, 1963). This will be denoted by

$$B^{**} \sim W_p'(n,\ \Sigma,\ M)\ .$$

Now let

$$X = X^* - E_{n1}E_{1n}DX^* = (I - E_{nn}D)X^*$$

be the centered data matrix for the masses m_i $(i = 1, ..., n)$, i.e. the data matrix obtained from X^* by subtracting the sample means $\underline{\overline{x}} = E_{1n}DX^*$, where we use $D = diag(m_i)$.

Then $B = X'X$ is said to have a central Wishart distribution with $n-1$ degrees of freedom. Its distribution will then still be given by (4.4), but with n replaced by $n-1$ and M by $0_{n\times p}$ so that then ${}_0F_1\Big(\frac{n}{2};\ \frac{1}{4}\Sigma^{-1}M'M\Sigma^{-1}B\Big) = 1$. This then will be denoted by

$$B \sim W_p(n-1,\ \Sigma)\ .$$

4.3 The MLE's of $\underline{\mu}$ and Σ for a random sample with arbitrary weights attached to the observations.

Suppose that we have a random sample of size n from the distribution (4.1). Then the likelihood function of the sample observations is

$$L = \frac{1}{(2\pi)^{np/2}\ |\Sigma|^{n/2}}\ e^{-\frac{1}{2}tr\left(\Sigma^{-1}(X^*-E_{n1}\underline{\mu}')'(X^*-E_{n1}\underline{\mu}')\right)}\ .$$

The exponent in this likelihood is really the sum of Mahalanobis distances between each observation and the assumed model. Each observation is given a weight equal to nm_i which is 1 since each observation has the same mass $m_i = \frac{1}{n}$ $(i = 1, ..., n)$. The MLE's (Maximum Likelihood Estimates) of $\underline{\mu}$ and Σ are well known. They are

$$\hat{\underline{\mu}} = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_1$$

and

$$\hat{\Sigma} = \frac{1}{n}A \qquad \text{where} \qquad A = \sum_{i=1}^{n}(\underline{x}_i - \underline{\mu})(\underline{x}_i - \underline{\mu})'\ .$$

However, we may choose, for some extraneous reasons, as explained in Chapter 2, to give a different mass m_i to each observation, thus giving a weight nm_i $(m_1 > 0,\ \sum_{i=1}^{n} m_i = 1)$ to each of the terms $(\underline{x}_i^* - \underline{\mu})'\Sigma^{-1}(\underline{x}_i^* - \underline{\mu})$ in the exponent, for some extraneous reasons.

We now consider the MLE's of $\underline{\mu}$ and Σ in this case. But first we will define the Likelihood of a random sample with arbitrary weights.

<u>Definition 1</u>: The likelihood of a random sample of size n from the multivariate normal distribution in (4.1), when an arbitrary mass m_i $(m_i > 0,\ \sum_{i=1}^{n} m_i = 1)$ is assigned to the i^{th} observation, is defined as

$$L^* \ = \ \frac{1}{(2\pi)^{np/2}\ |\Sigma|^{n/2}}\ e^{-\frac{1}{2}n\ tr\left(\Sigma^{-1}(X^*-E_{n1}\underline{\mu}')'D(X^*-E_{n1}\underline{\mu}')\right)}\ , \qquad (4.5)$$

where $D = diag(m_i)$ and X^* is the (non-centered) matrix of the random-sample of size n from the distribution of $\underline{x}$. □

Strictly speaking L^* cannot be called a likelihood because L^* is not the density of X^* and the constant in L^* is not such that the integral of L^* over the space of X^* is 1. Therefore, if the above definition of likelihood, when weights are used, is not agreeable, then what we call as the MLE's should be interpreted as weighted least squares estimates.

We have then the following result.

Theorem 1: The MLE's of $\underline{\mu}$ and Σ, in case the weights m_i are used, are

$$\hat{\underline{\mu}} = \overline{\underline{x}} = X^{*\prime} D E_{n1}$$

and

$$\hat{\Sigma} = V = X' D X \qquad \text{with} \qquad X = (I - E_{n1} E_{1n} D) X^* \ .$$

Proof:

Let $\overline{\underline{x}}$ be the weighted mean vector for the weights m_i, i.e. let

$$\overline{\underline{x}} = (E_{1n} D X^*)' \ .$$

Then it is easy to see that

$$\left(X^* - E_{n1}\underline{\mu}'\right)' D \left(X^* - E_{n1}\underline{\mu}'\right) \ = \ V + \left(\overline{\underline{x}} - \underline{\mu}\right)\left(\overline{\underline{x}} - \underline{\mu}\right)' \ .$$

So that we can write the likelihood (4.5) as

$$L^* \ = \ K \ |\Sigma|^{-n/2} \ e^{-\frac{1}{2} n \ tr\left(\Sigma^{-1} V + \Sigma^{-1} (\overline{\underline{x}} - \underline{\mu})(\overline{\underline{x}} - \underline{\mu})'\right)} \ .$$

Then it can be seen that the MLE's of $\underline{\mu}$ and Σ are

$$\hat{\underline{\mu}} = \overline{\underline{x}} = X^{*\prime} D E_{n1}$$

and

$$\hat{\Sigma} = V \ .$$

The proof is similar to the one used by Muirhead (1982) in his Theorem 3.1.5.

<u>Result 1</u>:

$$\mathcal{E}(V) \;=\; (1-l)\,\Sigma \qquad \text{with} \qquad l = E_{1n}D^2E_{n1} = \sum_{i=1}^{n} m_i^2 \;.$$

The proof follows directly from the definition. ∎

4.4 The distribution of V, its submatrices and their functions.

Let $V = X'DX$ be defined as in Theorem 1. Then we have the following Lemma.

<u>Lemma 1</u>: The distribution of $V = X'DX$ is approximately $W_p\left(\frac{(1-l)^2}{g},\ \frac{g}{1-l}\Sigma\right)$, what we will denote as

$$V \overset{app}{\sim} W_p\left(\frac{(1-l)^2}{g},\ \frac{g}{1-l}\Sigma\right) , \tag{4.6}$$

where

$$g = tr(A^2)$$

with

$$A = (I_n - E_{nn}D)'D(I_n - E_{nn}D) \;,$$

and

$$l = E_{1n}D^2E_{n1} \;,$$

and where we also assumed $E_{1n}DE_{n1} = 1$.

The approximation is such that the two first moments of V agree with the two first moments of the Wishart distribution.

(Note that g is also given by

$$g = tr(A^2) = tr(D^2) - 2m + l^2$$

where

$$l = E_{1n}D^2E_{n1} \quad \text{and} \quad m = E_{1n}D^3E_{n1} \;.)$$

Proof:

Since X^* is a random sample from $N_p(\underline{\mu}, \Sigma)$, $X^{*\prime}X^*$ has a non-central Wishart distribution,

$$X^{*\prime}X^* \sim W_p'(n,\ \Sigma,\ E_{n1}\underline{\mu}') \ .$$

Then we may use a result of Neudecker (1986) that the variance-covariance matrix of $vec(V) = vec(X^{*\prime}AX^*)$, when $X^{*\prime}X^*$ has a non-central Wishart distribution $W_p'(n,\ \Sigma,\ M)$ and A is symmetric is

$$\mathcal{D}(vec(V)) = (I_{p^2} + K_{pp})\left[(trA^2)(\Sigma \otimes \Sigma) + M'A^2M \otimes \Sigma + \Sigma \otimes M'A^2M\right]$$

where $\mathcal{D}$ stands for 'dispersion matrix' and $M = E_{n1}\underline{\mu}'$ is the non-centrality parameter from the Wishart distribution of $X^{*\prime}X^*$. In our case

$$A = (I_n - E_{nn}D)'D(I_n - E_{nn}D)$$

clearly symmetric.

Let

$$\begin{aligned} k &= E_{1n}DE_{n1} \\ l &= E_{1n}D^2E_{n1} \\ m &= E_{1n}D^3E_{n1} \ , \end{aligned}$$

i.e., k, l and m will be respectively the sum of the elements in D, D^2 and D^3. We then have, after some simplification

$$A = D(I_n + (k-2)E_{nn}D)$$

and

$$A^2 = D^2 + (k-2)D^2E_{nn}D + (k-2)DE_{nn}D^2 + (k-2)^2lDE_{nn}D$$

and then

$$tr(A^2) = tr(D^2) + 2(k-2)m + (k-2)^2l^2 \ ,$$

and

$$M'A^2M = (l + 2k(k-2)l + k^2(k-2)^2l)\underline{\mu}\ \underline{\mu}' \ .$$

If we assume $k = 1$ then

$$A = D(I_n - E_{nn}D) \ ,$$

$$tr(A^2) = tr(D^2) - 2m + l^2$$

and

$$l + 2k(k-2)l + k^2(k-2)^2 l = l - 2l + l = 0$$

so that

$$M'A^2M = 0_{(p \times p)} \ .$$

This situation $k = E_{1n}DE_{n1} = 1$ is a quite plausible one even in a more general setting when D is not diagonal, and is clearly a superset of the case $D = diag(m_i)$ with $\sum_{i=1}^{n} m_i = 1$.

Using the result of Neudecker (1986),

$$\mathcal{D}(vec(V)) = (trA^2)(I_{p^2} + K_{pp})(\Sigma \otimes \Sigma) \ .$$

From Result 1,

$$\mathcal{E}(V) = (1 - l) \ \Sigma \ .$$

We also know that (see for example Kshirsagar (1972, 1989)) if

$$V^* \sim W_p\left(\frac{a_1^2}{a_2}, \ \frac{a_2}{a_1}\Gamma\right)$$

then

$$\mathcal{E}(V^*) = a_1 \ \Gamma \ ,$$

and

$$\mathcal{D}(vec(V^*)) = a_2(I_{p^2} + K_{pp})(\Gamma \otimes \Gamma) \ .$$

Then if we let

$$a_1 = 1 - l$$

$$a_2 = g = tr(A^2) = tr(D^2) - 2m + l^2 \ ,$$

we obtain

$$V \overset{app}{\sim} W_p\left(\frac{(1-l)^2}{g},\ \frac{g}{1-l}\Sigma\right) ,$$

in the sense that the two first moments of V agree with the two first moments of the Wishart distribution. ■

We may notice that when $D = \frac{1}{n}I_n$ then V is equal to $\frac{n-1}{n}$ times the usual sample variance-covariance matrix and is distributed as $W_p(n-1,\ \ \frac{1}{n}\Sigma)$, which in this case is the exact distribution of V.

Now we will examine some details of the distributions of submatrices and functions of submatrices of a Wishart matrix.

<u>Lemma 2</u>: Let

$$V^* \sim W_p(f,\ \Gamma)$$

where

$$\Gamma = bdiag(\Gamma_{kk}) \qquad (\text{with } \Gamma_{kk}\ p_k{\times}p_k) \qquad (k = 1, ..., m) \ .$$

Let V^* be partitioned as

$$V^* = \begin{array}{c} \begin{array}{ccccccc} p_1 & p_2 & p_3 & & p_k & & p_m \end{array} \\ \left[\begin{array}{c|cccccc} V^*_{11} & V^*_{12} & V^*_{13} & \cdots & V^*_{1k} & \cdots & V^*_{1m} \\ \hline V^*_{21} & & & & & & \\ V^*_{31} & & & & & & \\ \vdots & & & V^*_2 & & & \\ V^*_{k1} & & & & & & \\ \vdots & & & & & & \\ V^*_{m1} & & & & & & \end{array}\right] \end{array} , \qquad p = \sum_{k=1}^{m} p_k$$

and let Γ be partitioned similarly.

Then

i) $V_{11.2}^* = V_{11}^* - A_{12} \; \sim \; W_{p_1}\,(f - (p - p_1),\; \Gamma_{11})$

ii) $V_2^* \; \sim \; W_{p-p_1}(f,\; \Gamma_2)$

iii) $\left(vec\left(\left[V_{12}^* \; V_{13}^* \ldots V_{1k}^* \ldots V_{1m}^* \right] \right) \,\middle|\, V_2^* \right) \; \sim \; N_{p_1(p-p_1)}(\underline{0}_{p_1(p-p_1)},\; \Gamma_{11} \otimes V_2^*)$

iv) $A_{12} = \left[V_{12}^* \; V_{13}^* \ldots V_{1k}^* \ldots V_{1m}^* \right] V_2^{*-1} \left[V_{12}^* V_{13}^* \ldots V_{1k}^* \ldots V_{1m}^* \right]' \; \sim \; W_{p_1}(p - p_1,\; \Gamma_{11})$

v) $V_{11}^* \; \sim \; W_{p_1}(f,\; \Gamma_{11})$

vi) V_{11}^* and V_2^* are independent. Also $V_2^*, V_{11.2}^*$ and A_{12} are independently distributed.

Proof:

Note that

$$
\begin{aligned}
|V^*| \;&=\; |V_2^*| \left| V_{11}^* - [V_{12}^* \; V_{13}^* \ldots V_{1k}^* \ldots V_{1m}^*] \; V_2^{*-1} \; [V_{12}^* \; V_{13}^* \ldots V_{1k}^* \ldots V_{1m}^*]' \right| \\
&=\; |V_2^*| \; |V_{11.2}^*|
\end{aligned}
$$

and

$$
|\;\Gamma\;| \;=\; |\Gamma_{11}| \; |\Gamma_2| \;,
$$

where Γ_2 is a submatrix of Γ in exactly the same way as V_2^* is of V^*.

Then we may transform from V^* to $V_{11.2}^*,\; V_2^*,\; B_{12}$, where

$$
B_{12} = \left[\; V_{12}^* \; V_{13}^* \ldots V_{1k}^* \ldots V_{1m}^* \; \right] \;.
$$

Since

$$
dV^* = dV_{11}^* \; dV_2^* \; dB_{12}
$$

and

$$
J(V_{11}^* \rightarrow V_{11.2}^*) = |I_{p_1}| = 1 \;,
$$

then

$$
dV^* = dV_{11.2}^* \; dV_2^* \; dB_{12} \;.
$$

Given that

$$\begin{aligned} tr(\Gamma^{-1}V^*) &= tr(\Gamma_{11}^{-1}V_{11}^*) + tr(\Gamma_2^{-1}V_2^*) \\ &= tr(\Gamma_{11}^{-1}(V_{11.2}^* + B_{12}V_2^{*-1}B_{21})) + tr(\Gamma_2^{-1}V_2^*) \\ &= tr(\Gamma_{11}^{-1}V_{11.2}^*) + tr(\Gamma_{11}^{-1}B_{12}V_2^{*-1}B_{21}) + tr(\Gamma_2^{-1}V_2^*) \ , \end{aligned}$$

it follows immediately, from (4.4), that

$$f(V_{11.2}^*, V_2^*, B_{12})\ dV_{11.2}^*\ dV_2^*\ dB_{12} =$$

$$= \frac{|V_{11.2}^*|^{(f-p-1)/2}\ |V_2^*|^{(f-p-1)/2}}{2^{fp/2}\ \Gamma_p\left(\frac{f}{2}\right) |\Gamma_{11}|^{f/2}\ |\Gamma_2|^{f/2}}\ e^{-\frac{1}{2}tr\left(\Gamma_{11}^{-1}V_{11.2}^* + \Gamma_{11}^{-1}B_{12}V_2^{*-1}B_{21} + \Gamma_2^{-1}V_2^*\right)}\ dV_{11.2}^*\ dV_2^*\ dB_{12}$$

$$= \frac{|V_{11.2}^*|^{(f-p+p_1-p_1-1)/2}}{2^{(f-p+p_1)p_1/2}\ \Gamma_{p_1}\left(\frac{f-p+p_1}{2}\right) |\Gamma_{11}|^{(f-p+p_1)/2}}\ e^{-\frac{1}{2}tr\left(\Gamma_{11}^{-1}V_{11.2}^*\right)}\ dV_{11.2}^*$$

$$\cdot\ \frac{|V_2^*|^{(f-p+p_1-1)/2}}{2^{f(p-p_1)/2}\ \Gamma_{p-p_1}\left(\frac{f}{2}\right) |\Gamma_2|^{f/2}}\ e^{-\frac{1}{2}tr\left(\Gamma_2^{-1}V_2^*\right)}\ dV_2^*$$

$$\cdot\ \frac{1}{(2\pi)^{p_1(p-p_1)/2}\ |V_2^*|^{p_1/2}\ |\Gamma_{11}|^{(p-p_1)/2}}\ e^{-\frac{1}{2}tr\left(\Gamma_{11}^{-1}B_{12}V_2^{*-1}B_{21}\right)}\ dB_{12}\ .$$

What shows that

i) $V_{11.2}^* \sim W_{p_1}(f-p+p_1,\ \Gamma_{11})$

ii) $V_2^* \sim W_{p-p_1}(f,\ \Gamma_2)$

iii) $\left(vec(B_{12})|V_2^*\right) \sim N_{p_1(p-p_1)}(\underline{0}_{p_1(p-p_1)},\ \Gamma_{11}\otimes V_2^*)$

and they are all independent.

But then iii) implies that $vec\left(B_{12}V_2^{*-1/2}\right) \sim N_{p_1(p-p_1)}(\underline{0}_{p_1(p-p_1)},\ \Gamma_{11}\otimes I_{p-p_1})$ or that then $A_{12} = B_{12}V_2^{*-1}B_{21} \sim W_{p_1}(p-p_1, \Gamma_{11})$ is itself independent of $V_{11.2}^*$ and also of V_2^*. And then also

$$V_{11}^*\left(= V_{11.2}^* + A_{12}\right) \sim W_{p_1}(f-p+p_1+p-p_1,\ \Gamma_{11}) \equiv W_{p_1}(f,\ \Gamma_{11})$$

still independent of V_2^*.

Really the independence of V_{11}^* and V_2^* follows directly from the fact that Γ is block diagonal. ■

In the above proof, the following Lemma has been used.

<u>Lemma 3</u>:

$$\begin{aligned}
&\left(vec(V_{kk'})\middle|V_{k'k'}\right) \sim N_{p_k p_{k'}}(\underline{0}_{p_k p_{k'}},\ \Gamma_{kk}\otimes V_{k'k'}) \Rightarrow \\
&\Rightarrow\ vec\left(V_{kk'}V_{k'k'}^{-1/2}\right) \sim N_{p_{k'}}(\underline{0}_{p_k p_{k'}},\ \Gamma_{kk}\otimes I_{p_{k'}}) \\
&\Rightarrow\ V_{kk'}V_{k'k'}^{-1}V_{k'k} \sim W_{p_k}(p_{k'},\ \Gamma_{kk})\ .
\end{aligned}$$

The Lemma is well known in Multivariate Analysis (Kshirsagar, 1972). ■

<u>Lemma 4</u>: Let V^* be the same as in Lemma 2, i.e. let

$$V^* \ \sim\ W_p(f,\ \Gamma)\ ,$$

where

$$\Gamma = bdiag(\Gamma_{kk}) \qquad (\text{with } \Gamma_{kk}\ p_k{\times}p_k) \qquad (k=1,...,m)\ .$$

Let V^* be partitioned as

$$V^* = \begin{array}{c} \begin{array}{ccccccc} p_1 & p_2 & p_3 & \cdots & p_k & \cdots & p_m \end{array} \\ \left[\begin{array}{ccccccc}
V_{11}^* & V_{12}^* & V_{13}^* & \cdots & V_{1k}^* & \cdots & V_{1m}^* \\
V_{21}^* & V_{22}^* & V_{23}^* & \cdots & V_{2k}^* & \cdots & V_{2m}^* \\
V_{31}^* & V_{32}^* & V_{33}^* & \cdots & V_{3k}^* & \cdots & V_{3m}^* \\
\vdots & \vdots & \vdots & & \vdots & & \vdots \\
V_{k1}^* & V_{k2}^* & V_{k3}^* & \cdots & V_{kk}^* & \cdots & V_{km}^* \\
\vdots & \vdots & \vdots & & \vdots & & \vdots \\
V_{m1}^* & V_{m2}^* & V_{m3}^* & \cdots & V_{mk}^* & \cdots & V_{mm}^*
\end{array}\right] \end{array} \begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_k \\ \vdots \\ p_m \end{array}$$

and let

$$V_{kk'.k''}^* = V_{kk'}^* - V_{kk''}^* V_{k''k''}^{*-1} V_{k''k'}^*\ ,$$

$$V^*_{k(l,...,m)} \;=\; \left[V^*_{kl}\; V^*_{k,l+1}\; V^*_{k,l+2}\; \ldots\; V^*_{km}\right] \;=\; V^{*'}_{(l,...,m)k} \qquad m \geq l \;,$$

$$V^*_l \;=\; V^*_{(l,...,m)(l,...,m)} \;=\; \begin{bmatrix} V^*_{l(l,...,m)} \\ V^*_{l+1(l,...,m)} \\ \vdots \\ V^*_{m(l,...,m)} \end{bmatrix} \qquad (l = 1,...,m) \quad \text{with} \quad V^*_1 = V^*$$

and

$$V^*_{kk'.(l,...,m)} \;=\; V^*_{kk'} - V^*_{k(l,...,m)} V^{*-1}_{(l,...,m)(l,...,m)} V^*_{(l,...,m)k'} \;.$$

Then, for $k = 1, ..., m$

i) $V^*_{kk} \;\sim\; W_{p_k}(f,\; \Gamma_{kk})$

ii) $\left(vec\left(V^*_{k(k+1,...,m)}\right)\Big| V^*_{(k+1,...,m)(k+1,...,m)}\right)$
$\sim\; N_{p_k(p_{k+1}+...+p_m)}\left(\underline{0}_{p_k(p_{k+1}+...+p_m)},\; \Gamma_{kk} \otimes V^*_{(k+1,...,m)(k+1,...,m)}\right)$

iii) $\left(vec(V^*_{kk'})|V^*_{k'k'}\right) \;\sim\; N_{p_k p_{k'}}(\underline{0}_{p_k p_{k'}},\; \Gamma_{kk} \otimes V^*_{k'k'})$ $\qquad (k \neq k')$

iv) $V^*_{k(k+1,...,m)} V^{*-1}_{(k+1,...,m)(k+1,...,m)} V^*_{(k+1,...,m)k}$
$\sim\; W_{p_k}\left(p_{k+1} + ... + p_m,\; \Gamma_{kk}\right)$

v) $V^*_{kk} - V^*_{k(k+1,...,m)} V^{*-1}_{(k+1,...,m)(k+1,...,m)} V^*_{(k+1,...,m)k}$
$\sim\; W_{p_k}\left(f - (p_{k+1} + ... + p_m),\; \Gamma_{kk}\right)$

vi) $V^*_{kk'} V^{*-1}_{k'k'} V^*_{k'k} \;\sim\; W_{p_k}(p_{k'},\; \Gamma_{kk})$ $\qquad (k \neq k')$

vii) V^*_{kk} and $V^*_{(k+1,...,m)(k+1,...,m)}$ are independent,
$V^*_{kk.(k+1,...,m)}$, $V^*_{k(k+1,...,m)} V^{*-1}_{(k+1,...,m)(k+1,...,m)} V^*_{(k+1,...,m)k}$
and $V^*_{(k+1,...,m)(k+1,...,m)}$ are also independent

viii) V^*_{kk} and $V^*_{k'k'}$ are independent,
V^*_{kk} , $V^*_{k'k'}$ and $V^*_{kk'} V^{*-1}_{k'k'} V^*_{k'k}$ are also independent, $\qquad$ for $k \neq k'$
and $V^*_{kk.k'}$, $V^*_{k'k'}$ and $V^*_{kk'} V^{*-1}_{k'k'} V^*_{k'k}$ are also independent

ix) $V^*_{kk.k'} = V^*_{kk} - V^*_{kk'} V^{*-1}_{k'k'} V^*_{k'k} \;\sim\; W_{p_k}(f - p_{k'},\; \Gamma_{kk})$ $\qquad (k \neq k')$

x) $\left(vec(V^*_{kk'.k''})|V^*_{k'k'.k''}\right) \;\sim\; N_{p_k p_{k'}}(\underline{0}_{p_k p_{k'}},\; \Gamma_{kk} \otimes V^*_{k'k'.k''})$ $\qquad k \neq k' \neq k''$

xi) $V^*_{kk'.k''} V^{*-1}_{k'k'.k''} V^*_{k'k.k''} \;\sim\; W_{p_k}(p_{k'},\; \Gamma_{kk})$ $\qquad k \neq k' \neq k''$

xii) $V^*_{kk.k''}$ and $V^*_{k'k'.k''}$ are independent, for $k \neq k'$,

$V^*_{kk.k''} - V^*_{kk'.k''} V^{*-1}_{k'k'.k''} V^*_{k'k.k''}$, $V^*_{k'k'.k''}$ and $V^*_{kk'.k''} V^{*-1}_{k'k'.k''} V^*_{k'k.k''}$ $\quad k \neq k''$

are also independent . $\quad k' \neq k''$

Proof of i):

Let us assume a partition of V^* as in Lemma 2. Then using i) and ii) from Lemma 2 and by induction, successively partitioning V^*_l ($l = 1, ..., m$) and Γ_l in a similar way as we did in Lemma 2 with V^* and Γ we obtain

$$V^*_{kk} \sim W_{p_k}(f,\ \Gamma_{kk}) \qquad k = 1, \ldots, m \ .$$

Proof of ii):

Again successively partitioning V^* and Γ as in i) above, and using iii) of Lemma 2 (by induction) we obtain

$$\begin{aligned}
&\left(vec([V^*_{12}\ V^*_{13} \ldots V^*_{1k} \ldots V^*_{1m}]) | V^*_2 \right) \sim N_{p_1(p-p_1)}(\underline{0}_{p_1(p-p_1)},\ \Gamma_{11} \otimes V^*_2) \\
&\left(vec([V^*_{23} \ldots V^*_{2k} \ldots V^*_{2m}]) | V^*_3 \right) \sim N_{p_2(p-p_1-p_2)}(\underline{0}_{p_2(p-p_1-p_2)},\ \Gamma_{22} \otimes V^*_3) \\
&\qquad \vdots \\
&\left(vec(V^*_{m-1,m-1}) | V^*_{mm} \right) \sim N_{p_{m-1}p_m}(\underline{0}_{p_{m-1}p_m},\ \Gamma_{m-1,m-1} \otimes V^*_{mm}) \ .
\end{aligned}$$

Proof of iii):

Directly from ii) above, it is a known result for multivariate normal distributions that

$$\left(vec([V^*_{12}\ V^*_{13} \ldots V^*_{1k} \ldots V^*_{1m}]) \mid V^*_2 \right) \sim N_{p_1(p-p_1)}(\underline{0}_{p_1(p-p_1)},\ \Gamma_{11} \otimes V^*_2) \Rightarrow$$

$$\begin{aligned}
\Rightarrow\ &\left(vec(V^*_{12}) \mid V^*_{22} \right) \sim N_{p_1p_2}(\underline{0}_{p_1p_2},\ \Gamma_{11} \otimes V^*_{22}) \\
&\left(vec(V^*_{13}) \mid V^*_{33} \right) \sim N_{p_1p_3}(\underline{0}_{p_1p_3},\ \Gamma_{11} \otimes V^*_{33}) \\
&\qquad \vdots \\
&\left(vec(V^*_{1m}) \mid V^*_{mm} \right) \sim N_{p_1p_m}(\underline{0}_{p_1p_m},\ \Gamma_{11} \otimes V^*_{mm})
\end{aligned}$$

and similarly for the other submatrices $V^*_{k(k+1,\ldots,m)}$ $(k = 3, \ldots, m-1)$ of V^*. These lead to

$$\left(vec(V^*_{kk'}) \mid V^*_{k'k'} \right) \sim N_{p_k p_{k'}}(\underline{0}_{p_k p_{k'}},\ \Gamma_{kk} \otimes V^*_{k'k'}) \qquad (\text{for } k' > k)\ .$$

Then, using a similar sequence of partitions of V^* and Γ but beginning with V^*_{mm}, $V^*_{(1,\ldots,m)(1,\ldots,m)}$ and $V^*_{m(1,\ldots,m)}$ and then partitioning $V^*_{(1,\ldots,m)(1,\ldots,m)}$ successively in a similar way, we obtain a similar result for $k > k'$.

Proof of iv):

Directly from ii), using Lemma 3.

Proof of v):

Once again using a sequence of partitions of V^* and Γ as in the proof of i) above and using v) of Lemma 2.

Proof of vi):

Directly from iii), using Lemma 3.

Proof of vii):

Again using a sequence of partitions of V^* and Γ as in the proof of i) and using vi) of Lemma 2 we get the independence of V^*_{kk}, $V^*_{(k+1,\ldots,m)(k+1,\ldots,m)}$ and $V^*_{k(k+1,\ldots,m)}V^{*-1}_{(k+1,\ldots,m)(k+1,\ldots,m)}V^*_{(k+1,\ldots,m)k}$, and also the independence of $V^*_{kk.(k+1,\ldots,m)}$, $V^*_{(k+1,\ldots,m)(k+1,\ldots,m)}$ and $V^*_{k(k+1,\ldots,m)}V^{*-1}_{k+1,\ldots,m)(k+1,\ldots,m)}V^*_{(k+1,\ldots,m)k}$.

Proof of viii):
Let

$$S_{p_k p_{k'}} = \begin{array}{c} \begin{array}{ccccccccc} p_1 & p_2 & \cdots & p_k & \cdots & p_{k'} & \cdots & p_m \end{array} \\ \left[\begin{array}{ccccccccc} 0 & 0 & \ldots & I_{p_k} & \ldots & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & \ldots & I_{p_{k'}} & \ldots & 0 \end{array} \right] \end{array} \begin{array}{c} \\ p_k \\ p_{k'} \end{array}$$

(for $k, k' \in \{1, \ldots, m\}$ and where the order of k and k' may be reverted), where zeros denote null matrices of the denoted orders.

Then, from

$$V^* \sim W_p(f,\ \Gamma) ,$$

$$S_{p_k p_{k'}} V^* S'_{p_k p_{k'}} \sim W_p(f,\ S_{p_k p_{k'}} \Gamma S'_{p_k p_{k'}})$$

where

$$S_{p_k p_{k'}} V^* S'_{p_k p_{k'}} = \begin{bmatrix} V^*_{kk} & V^*_{kk'} \\ V^*_{k'k} & V^*_{k'k'} \end{bmatrix}$$

and

$$S_{p_k p_{k'}} \Gamma S'_{p_k p_{k'}} = \begin{bmatrix} \Gamma_{kk} & 0 \\ 0 & \Gamma_{k'k'} \end{bmatrix}$$

so that for any $k \neq k'$

$$\begin{bmatrix} V^*_{kk} & V^*_{kk'} \\ V^*_{k'k} & V^*_{k'k'} \end{bmatrix} \sim W_{p_k+p_{k'}} \left(f, \begin{bmatrix} \Gamma_{kk} & 0 \\ 0 & \Gamma_{k'k'} \end{bmatrix} \right) . \tag{4.7}$$

But then application of vi) from Lemma 2 leads us immediately to the conclusion that V^*_{kk}, $V^*_{k'k'}$ and $(V^*_{kk'}|V^*_{k'k'})$ are all independent and thus so are V^*_{kk}, $V^*_{k'k'}$ and $V^*_{kk'}V^{*-1}_{k'k'}V^*_{k'k}$.

Proof of ix):

Directly from i), vi) and the above proven independence between V^*_{kk} and $V^*_{kk'}V^{*-1}_{k'k'}V^*_{k'k}$, or using v) from Lemma 2 in (4.7) above.

Proof of x):

In iii) make the transformation

$$V^*_{kk'} \longrightarrow V^*_{kk'.k''} = V^*_{kk'} - V^*_{kk''}V^{*-1}_{k''k''}V^*_{k''k'} .$$

The Jacobian of the above transformation is 1 and thus, from iii) $(V^*_{kk'.k''}|V^*_{k'k'})$ has the same distribution as $(V^*_{kk'}|V^*_{k'k'})$.

Then make the transformation

$$V^*_{k'k'} \longrightarrow V^*_{k'k'.k''} = V^*_{k'k'} - V^*_{k'k''} V^{*-1}_{k''k''} V^*_{k''k'} \ .$$

The Jacobian of the transformation is again 1, and thus

$$(vec(V^*_{kk'.k''})|V^*_{k'k'.k''}) \ \sim \ N_{p_k p_{k'}} \left(\underline{0}_{p_k p_{k'}},\ \Gamma_{kk} \otimes V^*_{k'k'.k''} \right) \ .$$

Proof of xi):

Directly from x), using Lemma 3.

Proof of xii):

Make the transformation

$$\begin{bmatrix} V^*_{kk} & V^*_{kk'} \\ V^*_{k'k} & V^*_{k'k'} \end{bmatrix} \longrightarrow \begin{bmatrix} V^*_{kk.k''} & V^*_{kk'.k''} \\ V^*_{k'k'.k''} & V^*_{k'k'.k''} \end{bmatrix} = \begin{bmatrix} V^*_{kk} & V^*_{kk'} \\ V^*_{k'k} & V^*_{k'k'} \end{bmatrix} - \begin{bmatrix} V^*_{kk''} \\ V^*_{k'k''} \end{bmatrix} V^{*-1}_{k''k''} \begin{bmatrix} V^*_{kk''} \\ V^*_{k'k''} \end{bmatrix}' .$$

Then, from (4.7), using viii) above we obtain

$$\begin{bmatrix} V^*_{kk.k''} & V^*_{kk'.k''} \\ V^*_{k'k.k''} & V^*_{k'k'.k''} \end{bmatrix} \sim W_{p_k+p_{k'}} \left(f - p_{k''},\ \begin{bmatrix} \Gamma_{kk} & 0 \\ 0 & \Gamma_{k'k'} \end{bmatrix} \right) .$$

Applying vi) from Lemma 2 we obtain the desired result. ∎

We should note that in x), xi) and xii) above, k'' may represent not only one of the m sets of variables but really any combination of those sets as long as it is not included in k'' the variables from either set k or set k'.

Since the distribution of $V = X'DX$ is not exactly Wishart but rather, as given by (4.6),

$$V \ \overset{app}{\sim} \ W_p \left(\frac{(1-l)^2}{g},\ \frac{g}{1-l} \Sigma \right) \ ,$$

then the results of Lemmas 2 and 4 hold with f replaced by $\frac{(1-l)^2}{g}$, Γ by $\frac{g}{1-l}\Sigma$ and '$\sim$' by ' $\overset{app}{\sim}$ ', where the approximation is in the sense stated in Lemma 1.

4.5 The choice of the metric Q.

As in Chapter 3, we now consider,

$$Q = bdiag\Big(X_k' D X_k \Big)^{-1}, \qquad (k = 1, ..., m)$$

as our choice for the metric in $I\!E^p$.

If $\underline{x}$ has the normal distribution (4.1), then, from Theorem 1 in 4.3, Q is the inverse of the MLE of Σ, under the hypothesis of independence of the m sets of p_k variables. With $\underline{x}$ partitioned as in (4.1), and a corresponding a partition of V (as in Lemma 4), we have

$$|VQ| = \frac{|V|}{\prod_{k=1}^{m} |V_{kk}|} = \prod_{\alpha=1}^{p} \lambda_\alpha \ .$$

Observe that $|VQ|$ is the $(2/n)^{\text{th}}$ power of the Likelihood Ratio statistic to test the null hypothesis of independence of the m sets of variables. We will refer to $|VQ|$ as the generalized Wilks' Λ to test the above null hypothesis. Also with the above choice of Q GCA may be seen as a generalization of the usual Canonical Analysis, and as a generalization of other multivariate methods like Multivariate Analysis of Variance (MANOVA), Multivariate Analysis of Covariance (MANCOVA), Discriminant Analysis, Principal Components Analysis and Reciprocal Averaging Methods or Correspondence Analysis. All such methods and some other methods related to these are particular cases of GCA (Generalized Canonical Analysis). Therefore GCA is very useful as a statistical technique and model and has interesting practical applications.

4.6 Definition of the Wilks' lambda criterion and its distribution.

Definition 2: (Kshirsagar, 1972, 1983) We shall say that a random variable Λ has the Wilks' lambda distribution with parameters f, p and q and we shall denote this by

$$\Lambda \ \sim \ \Lambda(f,p,q) \equiv \Lambda(f,q,p)$$

if

$$\Lambda \ = \ \frac{|A|}{|A+B|}$$

where

$$A \ \sim \ W_p(f-q,\ \Sigma)$$

and

$$B \ \sim \ W_p(q,\ \Sigma)$$

and are independent. Usually A and B are called 'Error' and 'Hypothesis' matrices, respectively, because under certain assumptions and models A has a Wishart distribution but B has a Wishart distribution only when a certain hypothesis is true.

Alternatively we may say that the random variable Λ has the Wilks' lambda distribution with parameters $f,\ p,\ q$ if its h^{th} moment is

$$\mathcal{E}(\Lambda^h) \ = \ \prod_{i=1}^{p} \frac{\Gamma\left(\frac{f+1-i}{2}\right)\ \Gamma\left(\frac{f-q+1-i}{2}+h\right)}{\Gamma\left(\frac{f+1-i}{2}+h\right)\ \Gamma\left(\frac{f-q+1-i}{2}\right)} \ = \ \frac{\Gamma_p\left(\frac{f}{2}\right)\ \Gamma_p\left(\frac{f-q}{2}+h\right)}{\Gamma_p\left(\frac{f}{2}+h\right)\ \Gamma_p\left(\frac{f-q}{2}\right)} \ .$$

These are also the moments of

$$\prod_{i=1}^{p} t_i^2$$

where

$$t_i^2 \ \sim \ Beta\left(\frac{f-q+1-i}{2},\ \frac{q}{2}\right) \quad , \quad (i=1,...,p)$$

are independent. If we interchange p and q, the moments remain the same and so, these are also the moments of

$$\prod_{i=1}^{q} t_i^{*2}$$

where

$$t_i^{*2} \sim Beta\left(\frac{f-p+1-i}{2},\ \frac{p}{2}\right) \quad , \quad (i=1,...,q)$$

are independent.

Since the random variable Λ is defined over a finite range, the moments define the distribution uniquely and thus $\Lambda(f,p,q)$ and $\Lambda(f,q,p)$ denote the same distribution. $\square$

If A and B in the above definition are not exact Wishart matrices but are only approximately so, then the corresponding Wilks' Λ will also be an approximate one in the same sense. Specifically if

$$V \overset{app}{\sim} W_p\left(\frac{(1-l)^2}{g},\ \frac{g}{1-l}\Sigma\right)$$

and A and B are independent functions of submatrices of V then the distribution of $|A|/|A+B|$ will not be an exact Wilks' lambda distribution but only approximately so, in the sense that the matrices involved are not exactly Wishart but their first and second moments coincide with the two first moments of the associated Wishart distributions. For the sake of brevity, we will call the distribution of Λ in many cases as Wilks' lambda though strictly speaking it is only an approximate distribution because the matrices involved have only approximate Wishart distributions. However, it will always be clear from the context what the expressions 'Wilks' lambda distribution' or 'Wilks' lambda' mean.

4.7 The generalization of the Wilks' lambda.

The case $m=2$ corresponds to the usual Canonical Analysis. Λ=$|VQ|$ will be the usual Wilks' lambda, with

$$\begin{aligned}\Lambda = |VQ| &= \frac{|V|}{|V_{11}|\ |V_{22}|} \\ &= \frac{|V_{11} - V_{12}V_{22}^{-1}V_{21}|}{|V_{11}|} = \frac{|A|}{|A+B|}\end{aligned}$$

where

$$\begin{aligned}A &= V_{11} - V_{12}V_{22}^{-1}V_{21} \\ B &= V_{12}V_{22}^{-1}V_{21}\ .\end{aligned}$$

However, in the more general case, i.e. when $m > 2$, $\Lambda = |VQ|$ will be a generalization of the Wilks' lambda for two sets, and will be a power of the likelihood ratio test statistic to test the hypothesis H_0 that the m sets of variables are all independent or 'H_0: $\Sigma_{kk'} = 0$ for all $k \neq k'$'. It is our objective to show how in this more general case $\Lambda = |VQ|$ can be written as the product of $m(m-1)/2$ quantities that themselves are Wilks' lambdas (but are not independent), or alternatively, as the product of $m-1$ independent Wilks' lambda distributed variables.

First we analyze the particular cases $m = 3$ and $m = 4$ and then derive a more general result, for any m. These particular results enable us to obtain more insight in the interpretation of the generalized Wilks' lambda.

<u>Result 2</u>: When $m = 3$ we have

$$\Lambda = |VQ| = \frac{|\ V\ |}{|V_{11}|\ |V_{22}|\ |V_{33}|}$$

where

$$|\ V\ | = |V_{33}|\ |V_{22.3}|\ |V_{11.3} - V_{12.3}V_{22.3}^{-1}V_{21.3}|$$

with

$$V_{kk'.k''} = V_{kk'} - V_{kk''}V_{k''k''}^{-1}V_{k''k'} = V'_{k'k.k''}\ .$$

Furthermore, let,

$$V_{(kk.k'').(k'k'.k'')} = V_{kk.k''} - V_{kk'.k''}V_{k'k'.k''}^{-1}V_{k'k.k''}\ .$$

Then we may write,

$$\Lambda \;=\; |VQ| \;=\; \frac{|V_{22.3}\,|}{|V_{22}|}\;\frac{|V_{(11.3).(22.3)}|}{|V_{11}|} \;=\; \frac{|V_{22.3}|}{V_{22}}\;\frac{|V_{11.3}|}{|V_{11}|}\;\frac{V_{(11.3).(22.3)}}{|V_{11.3}|}\;.$$

Let

$$\begin{aligned}
A &= V_{22.3}\\
B &= V_{23}V_{33}^{-1}V_{32}\\
C &= V_{11.3}\\
D &= V_{13}V_{33}^{-1}V_{31}\\
E &= V_{(11.3).(22.3)} \;=\; V_{11.3} - V_{12.3}V_{22.3}^{-1}V_{21.3}\\
F &= V_{12.3}V_{22.3}^{-1}V_{21.3}\\
G &= D + F\;.
\end{aligned}$$

Then

$$\Lambda \;=\; |VQ| \;=\; \frac{|A|}{|A+B|}\;\frac{|E|}{|E+G|} \;=\; \frac{|A|}{|A+B|}\;\frac{|C|}{|C+D|}\;\frac{|E|}{|E+F|}\;. \tag{4.8}$$

It will be shown that all the factors on the right sides of (4.8) are Wilks' lambdas. We may write (4.8) as

$$\Lambda \;=\; \Lambda_{(2)(3)}\;\Lambda_{(1)(23)} \;=\; \Lambda_{(2)(3)}\;\Lambda_{(1)(3)}\;\Lambda_{(1.3)(2.3)} \tag{4.9}$$

where $\Lambda_{(i)(j)}$ stands for the Wilks' lambda for the test of independence of sets i and j, while $\Lambda_{(i.j)(k.j)}$ represents the Wilks' lambda for the test of independence between the appropriate residual vectors.

Result 3: For $m = 4$, and using the same notation as in Result 2, we may write

$$\Lambda \;=\; |VQ| \;=\; \frac{|\;V\;|}{|V_{11}|\;|V_{22}|\;|V_{33}|\;|V_{44}|}$$

where

$$|V| \;=\; |V_{44}|\;|V_{33.4}|\;|W|\;|Z - XW^{-1}X'|$$

with

$$\begin{aligned}
W &= V_{22.4} - V_{23.4}V_{33.4}^{-1}V_{32.4} &&= V_{(22.4).(33.4)} \\
Z &= V_{11.4} - V_{13.4}V_{33.4}^{-1}V_{31.4} &&= V_{(11.4).(33.4)} \\
X &= V_{12.4} - V_{13.4}V_{33.4}^{-1}V_{32.4} &&= V_{(12.4).(33.4)} \ .
\end{aligned}$$

Then

$$\begin{aligned}
\Lambda \ = \ |VQ| &= \frac{|V_{33.4}\ |}{V_{33}|} \ \frac{|\ W\ |}{|V_{22}|} \ \frac{|Z - XW^{-1}X'|}{|V_{11}|} \\
&= \frac{|V_{11.4}|}{|V_{11}|} \ \frac{|V_{22.4}|}{|V_{22}|} \ \frac{|V_{33.4}|}{|V_{33}|} \ \frac{|\ W\ |}{|V_{22.4}|} \ \frac{|Z - XW^{-1}X'|}{|\ Z\ |} \ .
\end{aligned}$$

Let

$$\begin{aligned}
A &= V_{11.4} \\
B &= V_{14}V_{44}^{-1}V_{41} \\
C &= V_{22.4} \\
D &= V_{24}V_{44}^{-1}V_{42} \\
E &= V_{33.4} \\
F &= V_{34}V_{44}^{-1}V_{43} \\
G &= W \ = \ V_{(22.4).(33.4)} \\
H &= V_{23.4}V_{33.4}^{-1}V_{32.4} \\
I &= Z \ = \ V_{(11.4).(33.4)} \\
J &= V_{13.4}V_{33.4}^{-1}V_{31.4} \\
K &= Z - XW^{-1}X' \\
L &= XW^{-1}X' \ .
\end{aligned}$$

Then Λ may be written as

$$\begin{aligned}
\Lambda &= \frac{|E|}{|E+F|} \ \frac{|G|}{|G+(H+D)|} \ \frac{|K|}{|K+(L+B)|} \\
&= \frac{|A|}{|A+B|} \ \frac{|C|}{|C+D|} \ \frac{|E|}{|E+F|} \ \frac{|G|}{|G+H|} \ \frac{|I|}{|I+J|} \ \frac{|K|}{|K+L|}
\end{aligned}$$

or

$$\begin{aligned}\Lambda &= \Lambda_{(3)(4)}\ \Lambda_{(2)(34)}\ \Lambda_{(1)(234)} \\ &= \Lambda_{(1)(4)}\ \Lambda_{(2)(4)}\ \Lambda_{(3)(4)}\ \Lambda_{(22.4)(33.4)}\ \Lambda_{(11.4)(33.4)}\ \Lambda_{((11.4).(33.4))((22.4).(33.4))}\ ,\end{aligned} \tag{4.10}$$

where in the first equality the three independent factors represent the Wilks' lambda statistics corresponding to the canonical analysis of sets 3 versus 4, 2 versus 3 and 4 together, and 1 versus 2, 3 and 4 together, and in the second equality the first three factors represent the Wilks' lambda statistics corresponding to the canonical analysis between the fourth set of variables and the first, second and third sets of variables, respectively, the following three factors representing the Wilks' lambda statistics corresponding to the canonical analysis between the residuals specified by the subscripts.

We may see from (4.9) and (4.10) that the number of independent factors is $m-1$ and the number of non-independent factors is $m(m-1)/2$. All the factors in (4.9) and (4.10) are Wilks' lambda distributed variables. However, some are independent and some are not, as it will be shown in the following two Theorems.

For $m=3$ groups of variables let the following notation hold

$$V = \begin{array}{c} \begin{array}{ccc} p_1 & p_2 & p_3 \end{array} \\ \left[\begin{array}{ccc} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{array}\right] \end{array} \begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \end{array} \qquad \text{with} \quad p = \textstyle\sum_{k=1}^{m} p_k \quad (m=3)$$

where

$$V \overset{app}{\sim} W_p(f,\ \Gamma)\ ,$$

with

$$f = \frac{(1-l)^2}{g} \quad , \quad \Gamma = \frac{g}{1-l}\Sigma\ ,$$

where l and g are defined as in Lemma 1 in section 4.4.

Consider the null hypothesis

$$H_0: \quad \Gamma = \Gamma_0 = bdiag(\Gamma_{11},\ \Gamma_{22},\ \Gamma_{33})\ ,$$

for a partition of Γ corresponding to the above partition of V.

Furthermore let the notation relative to $V_{kk'.k''}$ and $V_{(kk.k'').(k'k'.k'')}$ used in Result 2 hold. Then we state the following Theorem.

Theorem 2: Let

$$V \overset{app}{\sim} W_p(f,\ \Gamma)\ .$$

Then, for $m = 3$, when the null hypothesis (4.3) is true

i) $V_{33} \overset{app}{\sim} W_{p_3}(f,\ \Gamma_{33})$

ii) $V_{22.3} \overset{app}{\sim} W_{p_2}(f - p_3,\ \Gamma_{22})$

iii) $V_{(11.3).(22.3)} = V_{11.3} - V_{12.3}V_{22.3}^{-1}V_{21.3} \overset{app}{\sim} W_{p_1}(f - p_3 - p_2,\ \Gamma_{11})$

iv) $(V_{12.3} \mid V_{22.3}) \overset{app}{\sim} N_{p_1}(\ 0,\ \Gamma_{11} \otimes V_{22.3})$

v) $(V_{13} \mid V_{33}) \overset{app}{\sim} N_{p_1}(\ 0,\ \Gamma_{11} \otimes V_{33})$

vi) $(V_{23} \mid V_{33}) \overset{app}{\sim} N_{p_2}(\ 0,\ \Gamma_{22} \otimes V_{33})$

and they are all independent.

Furthermore, as shown in Result 2, we may write

$$\Lambda = |VQ| = \frac{|V_{22} - V_{23}V_{33}^{-1}V_{32}|}{|V_{22}|} \left(\frac{|V_{11} - V_{13}V_{33}^{-1}V_{31}|}{|V_{11}|} \ \frac{|V_{(11.3).(22.3)}|}{|V_{11} - V_{13}V_{33}^{-1}V_{31}|} \right)$$

where all the three factors are (approximate) Wilks' lambdas, but the factors in the parenthesis are not independent.

But, then we may also write

$$\Lambda = |VQ| = \frac{|V_{22} - V_{23}V_{33}^{-1}V_{32}|}{|V_{22}|} \ \frac{|V_{(11.3).(22.3)}|}{|V_{11}|}\ ,$$

where (under H_0) both factors have independent (approximate) Wilks' lambda distributions.

Proof: Though the distributions i) through vi) may be obtained from Lemma 4, they may be also directly proved from the Wishart distribution of V.

In

$$V \overset{app}{\sim} W_p(f,\ \Gamma)$$

we make the following transformations:

$$\begin{aligned} V_{11} &\to V_{(11.3).(22.3)} \\ V_{22} &\to V_{22.3} \\ V_{33} &\to V_{33} \\ V_{12} &\to V_{12.3} \\ V_{13} &\to V_{13} \\ V_{23} &\to V_{23}\ . \end{aligned}$$

The Jacobian of transformation is 1. Since

$$|V| = |V_{33}|\ |V_{22} - V_{23}V_{33}^{-1}V_{32}|\ |V_{(11.3).(22.3)}|$$

and

$$\begin{aligned} tr(\Gamma^{-1}V) &= tr(\Gamma_{11}^{-1}V_{11}) + tr(\Gamma_{22}^{-1}V_{22}) + tr(\Gamma_{33}^{-1}V_{33}) \\ &= tr(\Gamma_{33}^{-1}V_{33}) + tr(\Gamma_{22}^{-1}V_{22.3}) + tr(\Gamma_{22}^{-1}V_{23}V_{33}^{-1}V_{32}) \\ &\quad + tr(\Gamma_{11}^{-1}V_{(11.3).(22.3)}) + tr(\Gamma_{11}^{-1}V_{12.3}V_{22.3}^{-1}V_{21.3}) \\ &\quad + tr(\Gamma_{11}^{-1}V_{13}V_{33}^{-1}V_{31})\ , \end{aligned}$$

$$f(V_{(11.3).(22.3)}, V_{22.3}, V_{33}, V_{12.3}, V_{13}, V_{23})\ dV_{(11.3).(22.3)}\ dV_{22.3}\ dV_{33}\ dV_{12.3}\ dV_{13}\ dV_{23} =$$

$$= \frac{|V_{(11.3).(22.3)}|^{(f-p-1)/2}\ |V_{22.3}|^{(f-p-1)/2}\ |V_{33}|^{(f-p-1)/2}}{2^{fp/2}\ \Gamma_p\left(\frac{f}{2}\right)\ |\Gamma_{11}|^{f/2}\ |\Gamma_{22}|^{f/2}\ |\Gamma_{33}|^{f/2}} \cdot$$

$$\cdot\ e^{tr-\frac{1}{2}(\Gamma_{33}^{-1}V_{33}+\Gamma_{22}^{-1}V_{22.3}+\Gamma_{22}^{-1}V_{23}V_{33}^{-1}V_{32}+\Gamma_{11}^{-1}V_{(11.3).(22.3)}+\Gamma_{11}^{-1}V_{12.3}V_{22.3}^{-1}V_{21.3}+\Gamma_{11}^{-1}V_{13}V_{33}^{-1}V_{31})}$$

$$dV_{(11.3).(22.3)}\ dV_{22.3}\ dV_{33}\ dV_{12.3}\ dV_{13}\ dV_{23}$$

$$= \frac{|V_{33}|^{(f-p_3-1)/2}}{2^{fp_3/2}\ \Gamma_{p_3}\left(\frac{f}{2}\right)\ |\Gamma_{33}|^{f/2}}\ e^{tr-\frac{1}{2}\Gamma_{33}^{-1}V_{33}}\ \ dV_{33}$$

$$\cdot \; \frac{|V_{22.3}|^{(f-p_3-p_2-1)/2}}{2^{(f-p_3)p_2/2}\ \Gamma_{p_2}\left(\frac{f-p_3}{2}\right)\ |\Gamma_{22}|^{(f-p_3)/2}}\ e^{tr-\frac{1}{2}\Gamma_{22}^{-1}V_{22.3}}\quad dV_{22.3}$$

$$\cdot \; \frac{|V_{(11.3).(22.3)}|^{((f-p_3-p_2)-p_1-1)/2}}{2^{(f-p_3-p_2)p_1/2}\ \Gamma_{p_1}\left(\frac{f-p_3-p_2}{2}\right)\ |\Gamma_{11}|^{(f-p_3-p_2)/2}}\ e^{tr-\frac{1}{2}\Gamma_{11}^{-1}V_{(11.3).(22.3)}}\quad dV_{(11.3).(22.3)}$$

$$\cdot \; \frac{1}{(2\pi)^{p_1p_2/2}\ |\Gamma_{11}|^{p_2/2}\ |V_{22.3}|^{p_1/2}}\ e^{tr-\frac{1}{2}\Gamma_{11}^{-1}V_{12.3}V_{22.3}^{-1}V_{21.3}}\quad dV_{12.3}$$

$$\cdot \; \frac{1}{(2\pi)^{p_1p_3/2}\ |\Gamma_{11}|^{p_3/2}\ |V_{33}|^{p_1/2}}\ e^{tr-\frac{1}{2}\Gamma_{11}^{-1}V_{13}V_{33}^{-1}V_{31}}\quad dV_{13}$$

$$\cdot \; \frac{1}{(2\pi)^{p_2p_3/2}\ |\Gamma_{22}|^{p_3/2}\ |V_{33}|^{p_2/2}}\ e^{tr-\frac{1}{2}\Gamma_{22}^{-1}V_{23}V_{33}^{-1}V_{32}}\quad dV_{23}$$

which proves i) through vi) and the independence of the distributions.

In the above we used

$$\begin{aligned}
\Gamma_p\left(\frac{f}{2}\right) &= \pi^{p(p-1)/4}\ \prod_{i=1}^{p}\Gamma\left[\frac{1}{2}(f-i+1)\right] \\
&= \pi^{p_3(p_3-1)/4}\ \prod_{i=1}^{p_3}\Gamma\left[\frac{1}{2}(f-i+1)\right] \\
&\quad \pi^{p_2(p_2-1)/4}\ \prod_{i=1}^{p_2}\Gamma\left[\frac{1}{2}(f-p_3-i+1)\right] \\
&\quad \pi^{p_1(p_1-1)/4}\ \prod_{i=1}^{p_1}\Gamma\left[\frac{1}{2}(f-p_3-p_2-i+1)\right] \\
&\quad \pi^{p_1p_2/2}\ \pi^{p_1p_3/2}\ \pi^{p_2p_3/2} \\
&= \Gamma_{p_3}\left(\frac{f}{2}\right)\ \Gamma_{p_2}\left(\frac{f-p_3}{2}\right)\ \Gamma_{p_1}\left(\frac{f-p_3-p_2}{2}\right)\ \pi^{p_1p_2/2}\ \pi^{p_1p_3/2}\ \pi^{p_2p_3/2}\ .
\end{aligned}$$

That $\frac{|V_{22.3}|}{|V_{22}|}$ and $\frac{|V_{11.3}|}{|V_{11}|}$ are Wilks' lambdas is a known fact that may be confirmed through i), vi) and viii) in Lemma 4. They are respectively (approximately) distributed as $\Lambda(f,p_2,p_3)$ and $\Lambda(f,p_1,p_3)$. Then, from iv) and Lemma 3, $V_{12.3}V_{22.3}^{-1}V_{21.3}$ is (approximately) $W_{p_1}(p_2,\Gamma_{11})$ and is independent of $V_{(11.3).(22.3)}$ that is (approximately) distributed as $W_{p_1}(f-p_3-p_2,\Gamma_{11})$ so that $\frac{|V_{(11.3).(22.3)}|}{|V_{11.3}|}$ is a Wilks' lambda since we may write

$$V_{(11.3).(22.3)} = V_{11.3} - V_{12.3}V_{22.3}^{-1}V_{21.3}\ .$$

Therefore $\frac{|V_{(11.3).(22.3)}|}{|V_{11.3}|}$ is approximately distributed as $\Lambda(f-p_3, p_1, p_2)$.

But, then $\frac{|V_{11.3}|}{|V_{11}|}$ is clearly not independent of $\frac{|V_{(11.3).(22.3)}|}{|V_{11.3}|}$. However, we may write $|VQ|$ as

$$|VQ| = \frac{|V_{22} - V_{23}V_{33}^{-1}V_{32}|}{|V_{22}|} \ \frac{|V_{(11.3).(22.3)}|}{|V_{11}|}$$

where both factors are now independent Wilks' lambdas.

$\frac{|V_{(11.3).(22.3)}|}{|V_{11}|}$ is a Wilks' lambda because, from iii),

$$V_{(11.3).(22.3)} = V_{11} - (V_{13}V_{33}^{-1}V_{31} + V_{12.3}V_{22.3}^{-1}V_{21.3})$$

is (approximately) distributed as $W_{p_1}(f-p_2-p_3, \Gamma_{11})$, and from v), iv) and Lemma 3, $V_{13}V_{33}^{-1}V_{31}$ is (approximately) $W_{p_1}(p_3, \Gamma_{11})$ independent of $V_{12.3}V_{22.3}^{-1}V_{21.3}$ which is (approximately) $W_{p_1}(p_2, \Gamma_{11})$, and both are independent of $V_{(11.3).(22.3)}$ as seen from iii), iv) and v). Thus $\frac{|V_{(11.3).(22.3)}|}{|V_{11}|}$ is (approximately) distributed as $\Lambda(f, p_1, p_2+p_3)$, and is independent of $\frac{|V_{22.3}|}{|V_{22}|}$ since from ii) and iii) $V_{22.3}$ and $V_{(11.3).(22.3)}$ are independent and from vii) in Lemma 4 V_{11} is independent of V_{22} and $V_{22.3}$, and V_{22}, that is a submatrix of $V_{(2,3)(2,3)}$, is independent of $V_{(11.3).(22.3)}$, since with a little algebra we can show that $V_{(11.3).(22.3)} = V_{11.(2,3)}$. ∎

For $m = 4$ groups of variables suppose that

$$V = \begin{array}{c} \begin{array}{cccc} p_1 & p_2 & p_3 & p_4 \end{array} \\ \begin{bmatrix} V_{11} & V_{12} & V_{13} & V_{14} \\ V_{21} & V_{22} & V_{23} & V_{24} \\ V_{31} & V_{32} & V_{33} & V_{34} \\ V_{41} & V_{42} & V_{43} & V_{44} \end{bmatrix} \end{array} \begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{array} \qquad \text{with} \quad p = \textstyle\sum_{k=1}^{m} p_k \quad (m=4)$$

with

$$V \overset{app}{\sim} W_p(f,\ \Gamma)$$

where under H_0, the null hypothesis (4.3),

$$\Gamma = \Gamma_0 = \ bdiag\ (\Gamma_{11},\ \Gamma_{22},\ \Gamma_{33},\ \Gamma_{44})\ .$$

Furthermore let

$$\begin{aligned}
Z &= V_{11} - V_{14}V_{44}^{-1}V_{41} - (V_{13} - V_{14}V_{44}^{-1}V_{43})(V_{33} - V_{34}V_{44}^{-1}V_{43})^{-1}(V_{31} - V_{34}V_{44}^{-1}V_{41}) \\
X &= V_{12} - V_{14}V_{44}^{-1}V_{42} - (V_{13} - V_{14}V_{44}^{-1}V_{43})(V_{33} - V_{34}V_{44}^{-1}V_{43})^{-1}(V_{32} - V_{34}V_{44}^{-1}V_{42}) \\
X' &= V_{21} - V_{24}V_{44}^{-1}V_{41} - (V_{23} - V_{24}V_{44}^{-1}V_{43})(V_{33} - V_{34}V_{44}^{-1}V_{43})^{-1}(V_{31} - V_{34}V_{44}^{-1}V_{41}) \\
W &= V_{22} - V_{24}V_{44}^{-1}V_{42} - (V_{23} - V_{24}V_{44}^{-1}V_{43})(V_{33} - V_{34}V_{44}^{-1}V_{43})^{-1}(V_{32} - V_{34}V_{44}^{-1}V_{42})
\end{aligned}$$

and

$$V_{kk'.k''} = V_{kk'} - V_{kk''}V_{k''k''}^{-1}V_{k''k'} = V'_{k'k.k''} \ .$$

Then the following Theorem holds.

Theorem 3: Let

$$V \overset{app}{\sim} W_p(f,\ \Gamma)\ .$$

Then, under H_0, for $m = 4$

$$\begin{aligned}
&\text{i)} && V_{44} \overset{app}{\sim} W_{p_4}(f,\ \Gamma_{44}) \\
&\text{ii)} && V_{33.4} \overset{app}{\sim} W_{p_3}(f - p_4,\ \Gamma_{33}) \\
&\text{iii)} && W \overset{app}{\sim} W_{p_2}(f - p_4 - p_3,\ \Gamma_{22}) \\
&\text{iv)} && (Z - XW^{-1}X') \overset{app}{\sim} W_{p_1}(f - p_4 - p_3 - p_2,\ \Gamma_{11}) \\
&\text{v)} && (X \mid W) \overset{app}{\sim} N_{p_1}(0,\ \Gamma_{11} \otimes W) \\
&\text{vi)} && (V_{13.4} \mid V_{33.4}) \overset{app}{\sim} N_{p_1}(0,\ \Gamma_{11} \otimes V_{33.4}) \\
&\text{vii)} && (V_{23.4} \mid V_{33.4}) \overset{app}{\sim} N_{p_2}(0,\ \Gamma_{22} \otimes V_{33.4}) \\
&\text{viii)} && (V_{14} \mid V_{44}) \overset{app}{\sim} N_{p_1}(0,\ \Gamma_{11} \otimes V_{44}) \\
&\text{ix)} && (V_{24} \mid V_{44}) \overset{app}{\sim} N_{p_2}(0,\ \Gamma_{22} \otimes V_{44}) \\
&\text{x)} && (V_{34} \mid V_{44}) \overset{app}{\sim} N_{p_3}(0,\ \Gamma_{33} \otimes V_{44})
\end{aligned}$$

and they are all independent.

Furthermore, as shown in Result 3, we may write

$$|VQ| = \frac{|V_{33} - V_{34}V_{44}^{-1}V_{43}|}{|V_{33}|} \left(\frac{|V_{22} - V_{24}V_{44}^{-1}V_{42}|}{|V_{22}|} \ \frac{|\ W\ |}{|V_{22} - V_{24}V_{44}^{-1}V_{42}|} \right)$$

$$\left(\frac{|Z - XW^{-1}X'|}{|Z|} \ \frac{|\ Z\ |}{|V_{11} - V_{14}V_{44}^{-1}V_{41}|} \ \frac{|V_{11} - V_{14}V_{44}^{-1}V_{41}|}{|V_{11}|} \right)$$

where all the six factors are (approximate) Wilks' lambdas, those which are not independent are shown inside the same parenthesis.

We may also write

$$|VQ| = \frac{|V_{33} - V_{34}V_{44}^{-1}V_{43}|}{|V_{33}|} \; \frac{|\; W \;|}{|V_{22}|} \; \frac{|Z - XW^{-1}X'|}{|V_{11}|} \tag{4.11}$$

where (under H_0) all the three factors are now independent (approximate) Wilks' lambdas.

Proof:

In the approximate Wishart distribution of V we make the transformations:

$$\begin{aligned}
V_{11} &\rightarrow (Z - XW^{-1}X') = (V_{11.4} - V_{13.4}V_{33.4}^{-1}V_{31.4}) - (V_{12.4} - V_{13.4}V_{33.4}^{-1}V_{32.4}) \\
&\qquad\qquad (V_{22.4} - V_{23.4}V_{33.4}^{-1}V_{32.4})^{-1}(V_{21.4} - V_{23.4}V_{33.4}^{-1}V_{31.4}) \\
V_{22} &\rightarrow W = V_{22.4} - V_{23.4}V_{33.4}^{-1}V_{32.4} \; (= V_{22} - V_{24}V_{44}^{-1}V_{42} - V_{23.4}V_{33.4}^{-1}V_{32.4}) \\
V_{33} &\rightarrow V_{33.4} \\
V_{44} &\rightarrow V_{44} \\
V_{12} &\rightarrow X = V_{12.4} - V_{13.4}V_{33.4}^{-1}V_{32.4} \\
V_{13} &\rightarrow V_{13.4} \\
V_{23} &\rightarrow V_{23.4} \\
V_{14} &\rightarrow V_{14} \\
V_{24} &\rightarrow V_{24} \\
V_{34} &\rightarrow V_{34} \; .
\end{aligned}$$

Given that all the Jacobians for the above transformations are equal to 1 and we may also write

$$\begin{aligned}
tr(\Gamma^{-1}V) &= tr(\Gamma_{11}^{-1}V_{11}) + tr(\Gamma_{22}^{-1}V_{22}) + tr(\Gamma_{33}^{-1}V_{33}) + tr(\Gamma_{44}^{-1}V_{44}) \\
&= tr(\Gamma_{44}^{-1}V_{44}) + tr(\Gamma_{33}^{-1}V_{33.4}) + tr(\Gamma_{33}^{-1}V_{34}V_{44}^{-1}V_{43}) \\
&\quad + tr(\Gamma_{22}^{-1}W) + tr(\Gamma_{22}^{-1}V_{23.4}V_{33.4}^{-1}V_{32.4}) + tr(\Gamma_{22}^{-1}V_{24}V_{44}^{-1}V_{42}) \\
&\quad + tr(\Gamma_{11}^{-1}(Z - XW^{-1}X')) + tr(\Gamma_{11}^{-1}XW^{-1}X') + tr(\Gamma_{11}^{-1}V_{13.4}V_{33.4}^{-1}V_{31.4}) \\
&\quad + tr(\Gamma_{11}^{-1}V_{14}V_{44}^{-1}V_{41}) \; ,
\end{aligned}$$

and

$$|V| = |V_{44}|\ |V_{33} - V_{34}V_{44}^{-1}V_{43}|\ |W|\ |Z - XW^{-1}X'|$$

then

$$f(Z-XW^{-1}X',\ W,\ V_{33.4},\ V_{44},\ X,\ V_{13.4},\ V_{14},\ V_{23.4},\ V_{24},\ V_{34})\ d(Z-XW^{-1}X')\ dW$$

$$dV_{33.4}\ dV_{44}\ dX\ dV_{13.4}\ dV_{14}\ dV_{23.4}\ dV_{24}\ dV_{34} =$$

$$= \frac{|V_{44}|^{(f-p-1)/2}\ |V_{33.4}|^{(f-p-1)/2}\ |W|^{(f-p-1)/2}\ |Z-XW^{-1}X'|^{(f-p-1)/2}}{2^{fp/2}\ \Gamma_p\left(\frac{f}{2}\right)|\Gamma_{11}|^{f/2}|\Gamma_{22}|^{f/2}|\Gamma_{33}|^{f/2}|\Gamma_{44}|^{f/2}} \cdot$$

$$\cdot\ exp\left[tr - \frac{1}{2}(\Gamma_{44}^{-1}V_{44} + \Gamma_{33}^{-1}V_{33.4} + \Gamma_{33}^{-1}V_{34}V_{44}^{-1}V_{43} + \Gamma_{22}^{-1}W + \Gamma_{22}^{-1}V_{23.4}V_{33.4}^{-1}V_{32.4} + \right.$$

$$+\Gamma_{22}^{-1}V_{24}V_{44}^{-1}V_{42} + \Gamma_{11}^{-1}(Z - XW^{-1}X') + \Gamma_{11}^{-1}XW^{-1}X' + \Gamma_{11}^{-1}V_{13.4}V_{33.4}^{-1}V_{31.4} +$$

$$\left. +\Gamma_{11}^{-1}V_{14}V_{44}^{-1}V_{41}\right]$$

$$d(Z - XW^{-1}X')\ dW\ dV_{33.4}\ dV_{44}\ dX\ dV_{13.4}\ dV_{14}\ dV_{23.4}\ dV_{24}\ dV_{34}$$

$$= \frac{|V_{44}|^{(f-p_4-1)/2}}{2^{fp_4/2}\ \Gamma_{p_4}\left(\frac{f}{2}\right)\ |\Gamma_{44}|^{f/2}}\ e^{tr-\frac{1}{2}\Gamma_{44}^{-1}V_{44}}\quad dV_{44}$$

$$\cdot\ \frac{|V_{33.4}|^{(f-p_4-p_3-1)/2}}{2^{(f-p_4)p_3/2}\ \Gamma_{p_3}\left(\frac{f-p_4}{2}\right)\ |\Gamma_{33}|^{(f-p_4)/2}}\ e^{tr-\frac{1}{2}\Gamma_{33}^{-1}V_{33.4}}\quad dV_{33.4}$$

$$\cdot\ \frac{|\ W\ |^{(f-p_4-p_3-p_2-1)/2}}{2^{(f-p_4-p_3)p_2/2}\ \Gamma_{p_2}\left(\frac{f-p_4-p_3}{2}\right)\ |\Gamma_{22}|^{(f-p_4-p_3)/2}}\ e^{tr-\frac{1}{2}\Gamma_{22}^{-1}W}\quad dW$$

$$\cdot\ \frac{|Z-XW^{-1}X'|^{(f-p_4-p_3-p_2-p_1-1)/2}}{2^{(f-p_4-p_3-p_2)p_1/2}\ \Gamma_{p_1}\left(\frac{f-p_4-p_3-p_2}{2}\right)\ |\Gamma_{11}|^{(f-p_4-p_3-p_2)/2}}$$

$$e^{tr-\frac{1}{2}\Gamma_{11}^{-1}(Z-XW^{-1}X')}\quad d(Z-XW^{-1}X')$$

$$\cdot\ \frac{1}{(2\pi)^{p_1p_3/2}\ |\Gamma_{11}|^{p_3/2}\ |V_{33.4}|^{p_1/2}}\ e^{tr-\frac{1}{2}\Gamma_{11}^{-1}V_{13.4}V_{33.4}^{-1}V_{31.4}}\quad dV_{13.4}$$

$$\cdot\ \frac{1}{(2\pi)^{p_1p_4/2}\ |\Gamma_{11}|^{p_4/2}\ |V_{44}|^{p_1/2}}\ e^{tr-\frac{1}{2}\Gamma_{11}^{-1}V_{14}V_{44}^{-1}V_{41}}\quad dV_{14}$$

$$\cdot \frac{1}{(2\pi)^{p_2p_3/2}\ |\Gamma_{22}|^{p_3/2}\ |V_{33.4}|^{p_2/2}}\ e^{tr - \frac{1}{2}\Gamma_{22}^{-1}V_{23.4}V_{33.4}^{-1}V_{32.4}}\ \ dV_{23.4}$$

$$\cdot \frac{1}{(2\pi)^{p_2p_4/2}\ |\Gamma_{22}|^{p_4/2}\ |V_{44}|^{p_2/2}}\ e^{tr - \frac{1}{2}\Gamma_{22}^{-1}V_{24}V_{44}^{-1}V_{42}}\ \ dV_{24}$$

$$\cdot \frac{1}{(2\pi)^{p_3p_4/2}\ |\Gamma_{33}|^{p_4/2}\ |V_{44}|^{p_3/2}}\ e^{tr - \frac{1}{2}\Gamma_{33}^{-1}V_{34}V_{44}^{-1}V_{43}}\ \ dV_{34}$$

$$\cdot \frac{1}{(2\pi)^{p_1p_2/2}\ |\Gamma_{11}|^{p_2/2}\ |\ W\ |^{p_1/2}}\ e^{tr - \frac{1}{2}\Gamma_{11}^{-1}XW^{-1}X'}\ \ dX\ ,$$

what proves i) through x) and the independence of the distributions there referenced. In the above we used

$$\begin{aligned}
\Gamma_p\left(\frac{f}{2}\right) &= \pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left[\tfrac{1}{2}(f-i+1)\right] \\
&= \pi^{p_4(p_4-1)/4}\prod_{i=1}^{p_4}\Gamma\left[\tfrac{1}{2}(f-i+1)\right] \\
&\quad \pi^{p_3(p_3-1)/4}\prod_{i=1}^{p_3}\Gamma\left[\tfrac{1}{2}(f-p_4-i+1)\right] \\
&\quad \pi^{p_2(p_2-1)/4}\prod_{i=1}^{p_2}\Gamma\left[\tfrac{1}{2}(f-p_4-p_3-i+1)\right] \\
&\quad \pi^{p_1(p_1-1)/4}\prod_{i=1}^{p_1}\Gamma\left[\tfrac{1}{2}(f-p_4-p_3-p_2-i+1)\right] \\
&\quad \pi^{p_1p_2/2}\ \pi^{p_1p_3/2}\ \pi^{p_1p_4/2}\ \pi^{p_2p_3/2}\ \pi^{p_2p_4/2}\ \pi^{p_3p_4/2} \\
&= \Gamma_{p_4}\left(\frac{f}{2}\right)\ \Gamma_{p_3}\left(\frac{f-p_4}{2}\right)\ \Gamma_{p_2}\left(\frac{f-p_4-p_3}{2}\right)\ \Gamma_{p_1}\left(\frac{f-p_4-p_3-p_2}{2}\right) \\
&\qquad \pi^{p_1p_2/2}\ \pi^{p_1p_3/2}\ \pi^{p_1p_4/2}\ \pi^{p_2p_3/2}\ \pi^{p_2p_4/2}\ \pi^{p_3p_4/2}
\end{aligned}$$

Then we can show that the six factors of the generalized Wilks' lambda are themselves all Wilks' lambdas.

That $\frac{|V_{33.4}|}{|V_{33}|}$, $\frac{V_{22.4}|}{|V_{22}|}$ and $\frac{|V_{11.4}|}{|V_{11}|}$ are Wilks' lambdas is a known fact that may be confirmed through i), vi) and viii) from Lemma 4. Then, from vii) and Lemma 3, $V_{23.4}V_{33.4}^{-1}V_{32.4}$ is (approximately) $W_{p_2}(p_3, \Gamma_{22})$ and is independent of W that from iii) is (approximately) $W_{p_2}(f-p_4-p_3, \Gamma_{22})$, so that $\frac{|\ W\ |}{|V_{22.4}|}$ has the approximate Wilks' lambda distribution $\Lambda(f-p_4, p_2, p_3)$ as $W = V_{22.4} - V_{23.4}V_{33.4}^{-1}V_{32.4}$. Also, from v) and Lemma 3 we know that $XW^{-1}X'$ is (approximately) distributed as $W_{p_1}(p_2, \Gamma_{11})$ independent of $Z - XW^{-1}X'$ that from iv) is (approximately) distributed as $W_{p_1}(f-p_4-p_3-p_2,\Gamma_{11})$, so that $\frac{|Z-XW^{-1}X'|}{|Z|}$ has the approximate Wilks' lambda distribution $\Lambda(f-p_4-p_3,p_1,p_2)$. Also $\frac{|Z|}{|V_{11.4}|}$ is a Wilks' lambda since we may write $Z=V_{11.4}-V_{13.4}V_{33.4}^{-1}V_{31.4}$ and from vi) and Lemma 3 we know that $V_{13.4}V_{33.4}^{-1}V_{31.4}$ is (approximately) distributed as $W_{p_1}(p_3,\Gamma_{11})$ independent of Z, since we may write $Z=(Z-XW^{-1}X')+XW^{-1}X'$, where from iv) and v) $(Z-XW^{-1}X')$ and $XW^{-1}X'$ are (approximately) distributed as $W_{p_1}(f-p_4-p_3-p_2,\Gamma_{11})$ and $W_{p_1}(p_2,\Gamma_{11})$ respectively, both independent of $V_{13.4}V_{33.4}^{-1}V_{31.4}$ through the independence of the distributions in iv), v) and vi). Thus $\frac{|Z|}{|V_{11.4}|}$ has the (approximate) $\Lambda(f-p_4,p_1,p_2)$ distribution.

But, then $\frac{|V_{22.4}|}{|V_{22}|}$ is clearly not independent of $\frac{|W|}{|V_{22.4}|}$, nor are the three Wilks' lambdas $\frac{|Z-XW^{-1}X'|}{|Z|}$, $\frac{|Z|}{|V_{11.4}|}$ and $\frac{|V_{11.4}|}{|V_{11}|}$. But then we may write $|VQ|$ as

$$|VQ|=\frac{|V_{33}-V_{34}V_{44}^{-1}V_{43}|}{|V_{33}|}\;\frac{|W|}{|V_{22}|}\;\frac{|Z-XW^{-1}X'|}{|V_{11}|}$$

where all the three factors are now independent Wilks' lambdas.

That $\frac{|W|}{|V_{22}|}$ is a Wilks' lambda may be shown through the fact that $W=V_{22}-(V_{24}V_{44}^{-1}V_{42}+V_{23.4}V_{33.4}^{-1}V_{32.4})$ where, from ix) and Lemma 3, $V_{24}V_{44}^{-1}V_{42}$ is (approximately) $W_{p_2}(p_4,\Gamma_{22})$ independent of $V_{23.4}V_{33.4}^{-1}V_{32.4}$ which from vii) is (approximately) $W_{p_2}(p_3,\Gamma_{22})$ and both are independent of W because of the independence of the distributions in iii), vii) and ix). Then $(V_{24}V_{44}^{-1}V_{42}+V_{23.4}V_{33.4}^{-1}V_{32.4})$ is (approximately) distributed as $W_{p_2}(p_3+p_4,\Gamma_{22})$ and is independent of W. Thus $\frac{|W|}{|V_{22}|}$ is approximately distributed as $\Lambda(f,p_2,p_3+p_4)$.

$\frac{|Z-XW^{-1}X'|}{|V_{11}|}$ is also a Wilks' lambda. This may be shown by the fact that we may write $Z-XW^{-1}X'=V_{11}-(V_{14}V_{44}^{-1}V_{41}+V_{13.4}V_{33.4}^{-1}V_{31.4}+XW^{-1}X')$ where from viii), vi), v) and iv) we know that $V_{14}V_{44}^{-1}V_{41}$ is (approximately) distributed as $W_{p_1}(p_4,\Gamma_{11})$, $V_{13.4}V_{33.4}^{-1}V_{31.4}$ is (approximately) distributed as $W_{p_1}(p_3,\Gamma_{11})$ and $XW^{-1}X'$ is (approximately) distributed as $W_{p_1}(p_2,\Gamma_{11})$, all independent among them and independent of $Z-XW^{-1}X'$, so that $(V_{14}V_{44}^{-1}V_{41}+V_{13.4}V_{33.4}^{-1}V_{31.4}+XW^{-1}X')$ is (approximately) $W_{p_1}(p_2+p_3+p_4,\Gamma_{11})$ and is independent of $Z-XW^{-1}X'$ and then $\frac{|Z-XW^{-1}X'|}{|V_{11}|}$ is approximately distributed as $\Lambda(f,p_1,p_2+p_3+p_4)$.

But from ii), iii) and iv) follows the independence of $V_{33.4}$, W and $Z-XW^{-1}X'$, and from Lemma 4, the independence of V_{11}, V_{22} and V_{33}, under H_0. Besides, with a little algebra and using the identity

$$V_{kk.k'}^{-1} \;=\; V_{kk}^{-1}+V_{kk}^{-1}V_{kk'}V_{k'k'.k}^{-1}V_{k'k}V_{kk}^{-1}$$

for $k\neq k'$, we may see that

$$Z-XW^{-1}X' \;=\; V_{11}-V_{1(2,3,4)}V_{(2,3,4)(2,3,4)}^{-1}V_{(2,3,4)1} \;=\; V_{11.(2,3,4)}$$

and

$$W \;=\; V_{22}-V_{2(3,4)}V_{(3,4)(3,4)}^{-1}V_{(3,4)2} \;=\; V_{22.(3,4)}\ .$$

Then, we may see that V_{11} is independent of W and also of $V_{33.4}$ since from vii) in Lemma 4 V_{11} is independent of $V_{(2,3,4)(2,3,4)}$ and as such independent of any function of any submatrices of $V_{(2,3,4)(2,3,4)}$. For a similar reason V_{22} is independent of $V_{33.4}$. Also, V_{22} is independent of $Z-XW^{-1}X'$ since V_{22} is a submatrix of $V_{(2,3,4)(2,3,4)}$ that still from vii) in Lemma 4 is independent of $Z-XW^{-1}X'=V_{11.(2,3,4)}$. By similar arguments also V_{33} is independent of W and $Z-XW^{-1}X'$, since V_{33} is a submatrix of $V_{(3,4)(3,4)}$ that yet from vii) in Lemma 4, is independent of $V_{22.(3,4)}=W$, and V_{33} is also a submatrix of $V_{(2,3,4)(2,3,4)}$ that from vii) in Lemma 4 is independent of $V_{11.(2,3,4)}=Z-XW^{-1}X'$. Therefore, the three Wilks' lambdas in (4.11) are independent. ∎

These Theorems show how the generalized Wilks' lambda may be written as a product of independent Wilks' lambdas under H_0. The above two Theorems for $m=3$ and $m=4$ may easily be generalized for any m, through the following Theorem.

First let us consider the following three types of (nested) null hypotheses (for a partition of Σ as in (4.2) and using a notation similar to the one used in Lemma 4):

$H_0 : \Sigma_{ij} = 0 \quad , \quad i = 1, ..., m-1;\ j = i+1, ..., m$

$\Big(\Leftrightarrow H_0 :$ the m sets of variables $\underline{x}_{(k)}$ $(k = 1, ..., m)$ are all independent$\Big)$;

$H_0^{(k)} : \Sigma_{k(k+1,...,m)} = 0 \quad (k = 1, ..., m-1)$

$\Big(\Leftrightarrow H_0^{(k)} :$ the set of variables $\underline{x}_{(k)}$ is independent of all the sets $\underline{x}_{(k+1)}$ through $\underline{x}_{(m)}$; $\quad (k = 1, ..., m-1)\ \Big)$

$$H_0^{i(k)} : \Sigma_{ki} \Big| \bigwedge_{\substack{l=k+1 \\ l \neq i}}^{m} \Sigma_{kl} \quad (i > k) \quad (k = 1, ..., m-1)\ (i = k+1, ..., m)$$

$\Big(\Leftrightarrow H_0^{i(k)} : \underline{x}_{(k)}$ and $\underline{x}_{(i)}$ are independent $(i > k)$, conditionally on all the sets $\underline{x}_{(k+1)}$ through $\underline{x}_{(m)}$ (but not $\underline{x}_{(i)}$); $(k = 1, ..., m-1)\ (i = k+1, ..., m)\Big)$.

We may note that if H_0 holds then $H_0^{(k)}$ holds for all $k = 1, ..., m-1$, and if $H_0^{(k)}$ holds (for a given k) then $H_0^{i(k)}$ holds for all $i = k+1, ..., m$, and conversely, i.e., if $H_0^{(k)}$ holds for all $k = 1, ..., m-1$ then H_0 holds and if (for a given k) $H_0^{i(k)}$ holds for all $i = k+1, ..., m$ then $H_0^{(k)}$ holds (for that given k).

Theorem 4: We may write

$$\Lambda = |VQ| = \prod_{k=1}^{m-1} \frac{|V_{kk} - V_{k(k+1,...,m)} V^{-1}_{(k+1,...,m)(k+1,...,m)} V_{(k+1,...,m)k}|}{|V_{kk}|} = \prod_{k=1}^{m-1} \Lambda_{k(k+1,...,m)} \tag{4.12}$$

where, under $H_0^{(k)}$, $\Lambda_{k(k+1,...,m)}$ $(k = 1, ..., m-1)$ has the approximate Wilks' lambda distribution

$$\Lambda_{k(k+1,...,m)} \overset{app}{\sim} \Lambda\left(\frac{(1-l)^2}{g},\ p_k,\ (p_{k+1} + ... + p_m)\right) \qquad (i = 1, ..., m-1)\ .$$

Under H_0 (the null hypothesis (4.3)), $\Lambda_{k(k+1,...,m)}$ $(k = 1, ..., m-1)$ are independent random variables.

Proof: If we split V as we did with V^* in Lemmas 2 and 4 and then $V_{(2,...,m)(2,...,m)}$

again in a similar way and if we continue this method, we will see that

$$| V | = |V_{(2,\ldots,m)(2,\ldots,m)}| \; |V_{11} - V_{1(2,\ldots,m)}V^{-1}_{(2,\ldots,m)(2,\ldots,m)}V_{(2,\ldots,m)1}|$$

where

$$|V_{(2,\ldots,m)(2,\ldots,m)}| = |V_{(3,\ldots,m)(3,\ldots,m)}| \; |V_{22} - V_{2(3,\ldots,m)}V^{-1}_{(3,\ldots,m)(3,\ldots,m)}V_{(3,\ldots,m)2}|$$

and so on, till finally we have,

$$|V_{(m-1,m)(m-1,m)}| = |V_{mm}| \; |V_{m-1,m-1} - V_{m-1,m}V^{-1}_{mm}V_{m,m-1}| \; .$$

We may then write

$$| V | = \left[\prod_{k=1}^{m-1} |V_{kk} - V_{k(k+1,\ldots,m)}V^{-1}_{(k+1,\ldots,m)(k+1,\ldots,m)}V_{(k+1,\ldots,m)k}| \right] |V_{mm}| \; .$$

Since $Q = bdiag\left(V^{-1}_{kk} \right)$ $(k = 1, \ldots, m)$,

$$|VQ| = \prod_{k=1}^{m-1} \frac{|V_{kk} - V_{k(k+1,\ldots,m)}V^{-1}_{(k+1,\ldots,m)(k+1,\ldots,m)}V_{(k+1,\ldots,m)k}|}{|V_{kk}|} \; .$$

Under $H_0^{(k)}$,

$$\Lambda_{k(k+1,\ldots,m)} = \frac{|V_{kk} - V_{k(k+1,\ldots,m)}V^{-1}_{(k+1,\ldots,m)(k+1,\ldots,m)}V_{(k+1,\ldots,m)k}|}{|V_{kk}|}$$

has an approximate Wilks' lambda distribution. This follows from Lemmas 1, 2 and 4. From Lemma 1 and i) in Lemma 4 we know that

$$V_{kk} \overset{app}{\sim} W_{p_k}\left(\frac{(1-l)^2}{g}, \; \frac{g}{1-l}\Sigma_{kk} \right) \; .$$

We should now notice that, for a given k, vii) in Lemma 4 holds not only under H_0, but also under $H_0^{(k)}$, so that from vii) in Lemma 4 we see that (under $H_0^{(k)}$)

$$V_{kk.(k+1,\ldots,m)} = V_{kk} - V_{k(k+1,\ldots,m)}V^{-1}_{(k+1,\ldots,m)(k+1,\ldots,m)}V_{(k+1,\ldots,m)k}$$

is independent of

$$V_{k(k+1,\ldots,m)}V^{-1}_{(k+1,\ldots,m)(k+1,\ldots,m)}V_{(k+1,\ldots,m)k} \; .$$

From Lemma 1 and iv) and v) in Lemma 4, we know that

$$V_{kk.(k+1,...,m)} \overset{app}{\sim} W_{p_k}\left(\frac{(1-l)^2}{g} - (p_{k+1}+...+p_m),\ \frac{g}{1-l}\Sigma_{kk}\right)$$

and

$$V_{k(k+1,...,m)}V^{-1}_{(k+1,...,m)(k+1,...,m)}V_{(k+1,...,m)k} \overset{app}{\sim} W_{p_k}\left((p_{k+1}+...+p_m),\ \frac{g}{1-l}\Sigma_{kk}\right) ,$$

so that $\Lambda_{k(k+1,...,m)}$ has, under $H_0^{(k)}$, the approximate Wilks' lambda distribution

$$\Lambda_{k(k+1,...,m)} \overset{app}{\sim} \Lambda\left(\frac{(1-l)^2}{g},\ p_k,\ (p_{k+1}+...+p_m)\right) .$$

But then using the notation in Lemma 4,

$$\Lambda_{k(k+1,...,m)} = \frac{|V_{kk.(k+1,...,m)}|}{|V_{kk}|} \qquad \text{and} \qquad \Lambda_{k+l(k+l+1,...,m)} = \frac{|V_{k+l,k+l.(k+l+1,...,m)}|}{|V_{k+l,k+l}|} ,$$

where, for $l = 1,...,m-i-1$, from vii) in Lemma 4, if $H_0^{(k)}$ holds, $V_{kk.(k+1,...,m)}$ is independent of $V_{(k+1,...,m)(k+1,...,m)}$, and as such, of any function of any submatrices in $V_{(k+1,...,m)(k+1,...,m)}$ and as such of $V_{k+l,k+l.(k+l+1,...,m)}$ and also of $V_{k+l,k+l}$. Still from vii) in Lemma 4, V_{kk} is independent of $V_{k+l,k+l}$ and $V_{k+l,k+l.(k+l+1,...,m)}$. Thus, any two of such Wilks' lambdas are independent. But then, if H_0 holds, i.e. if $H_0^{(k)}$ holds for all $k = 1,...,m-1$, any two Wilks' lambdas in (4.12) are then independent. ∎

Let us define

$$\Lambda_{ki|(k+1,...,i-1,i+1,...,m)} = \frac{\Lambda_{k(k+1,...,m)}}{\Lambda_{k(k+1,...,i-1,i+1,...,m)}} \tag{4.13}$$

as the conditional Wilks' lambda corresponding to the test of $H_0^{i(k)}$, i.e., to the test of significance of the canonical correlations between $\underline{x}_{(k)}$ and $\underline{x}_{(i)}$ given that sets $\underline{x}_{(k+1)}$ through $\underline{x}_{(i-1)}$ and $\underline{x}_{(i+1)}$ through $\underline{x}_{(m)}$ are in the model, with $m > i > k$ for some ordering of the sets $\underline{x}_{(1)}$ through $\underline{x}_{(m)}$.

We use the notation defined earlier to Theorem 4, and let

$$V_{k(k+1,...,i-1,i+1,...,m,i)} = \left[V_{k(k+1,...,i-1,i+1,...,m)} \; V_{ki} \right] .$$

Consider the following definition.

Definition 3: (Sequenced set)

We call a set a 'sequenced set', if its elements always appear in some predetermined order. Any subset of a sequenced set is also a sequenced set. □

In the following theorem we will show that, under $H_0^{i(k)}$, $\Lambda_{ki|(k+1,...,i-1,i+1,...,m)}$ has a Wilks' lambda distribution.

Theorem 5: Under any of the above null hypotheses, H_0, $H_0^{(k)}$ or $H_0^{i(k)}$,

$$\Lambda_{ki|(k+1,...,i-1,i+1,...,m)} = \frac{\Lambda_{k(k+1,...,m)}}{\Lambda_{k(k+1,...,i-1,i+1,...,m)}}$$

$$\stackrel{app}{\sim} \quad \Lambda\left(\tfrac{(1-l)^2}{g} - (p_{k+1} + ... + p_{i-1} + p_{i+1} + ... + p_m),\ p_k,\ p_i\right) .$$

Proof: The determinantal form of $\Lambda_{k(k+1,...,m)}$ was given in Theorem 4. Reordering the sets of variables we may write

$$\Lambda_{k(k+1,...,m)} = \frac{\left|V_{kk} - V_{k(k+1,...,i-1,i+1,...,m,i)} V^{-1}_{(k+1,...,i-1,i+1,...,m)(k+1,...,i-1,i+1,...,m,i)} V_{(k+1,...,i-1,i+1,...,m,i)}\right|}{|V_{kk}|} .$$

Let

$$u = \{i\} \quad , \quad t = \{k+1, ..., i-1, i+1, ..., m\}$$

be two sequenced sets of indexes.

Then

$$\begin{aligned} V^{-1}_{(k+1,...,i-1,i+1,...,m,i)(k+1,...,i-1,i+1,...,m,i)} &= V^{-1}_{(t,u)(t,u)} \\ &= \begin{bmatrix} V_{tt} & V_{tu} \\ V_{ut} & V_{uu} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} V^{-1}_{tt.u} & -V_{tt}^{-1} V_{tu} V^{-1}_{uu.t} \\ -V_{uu}^{-1} V_{ut} V^{-1}_{tt.u} & V^{-1}_{uu.t} \end{bmatrix} . \end{aligned}$$

Then

$$V_{k(t,u)}V^{-1}_{(t,u)(t,u)}V_{(t,u)k} = [V_{kt}\ V_{ku}]\begin{bmatrix} V^{-1}_{tt.u} & -V^{-1}_{tt}V_{tu}V^{-1}_{uu.t} \\ -V^{-1}_{uu}V_{ut}V^{-1}_{tt.u} & V^{-1}_{uu.t}\end{bmatrix}\begin{bmatrix} V_{tk} \\ V_{uk}\end{bmatrix}$$

$$= \left[V_{kt}V^{-1}_{tt.u} - V_{ku}V^{-1}_{uu}V_{ut}V^{-1}_{tt.u} \middle| -V_{kt}V^{-1}_{tt}V_{tu}V^{-1}_{uu.t} + V_{ku}V^{-1}_{uu.t}\right]\begin{bmatrix} V_{tk} \\ V_{uk}\end{bmatrix}$$

$$= V_{kt}V^{-1}_{tt.u}V_{tk} - V_{ku}V^{-1}_{uu}V_{ut}V^{-1}_{tt.u}V_{tk} - V_{kt}V^{-1}_{tt}V_{tu}V^{-1}_{uu.t}V_{uk} + V_{ku}V^{-1}_{uu.t}V_{uk}$$

$\Big($then using

$$\begin{aligned} V^{-1}_{tt.u} &= (V_{tt} - V_{tu}V^{-1}_{uu}V_{ut})^{-1} = \\ &= V^{-1}_{tt} + V^{-1}_{tt}V_{tu}V^{-1}_{uu.t}V_{ut}V^{-1}_{tt} \end{aligned}$$

(see the proof of viii) in Lemma 4)$\Big)$

$$= V_{ku}V^{-1}_{uu.t}V_{uk} + V_{kt}V^{-1}_{tt}V_{tk} + V_{kt}V^{-1}_{tt}V_{tu}V^{-1}_{uu.t}V^{-1}_{tt}V_{tk} -$$

$$-V_{ku}V^{-1}_{uu}V_{ut}V^{-1}_{tt}V_{tk} - V_{ku}V^{-1}_{uu}V_{ut}V^{-1}_{tt}V_{tu}V^{-1}_{uu.t}V_{ut}V^{-1}_{tt}V_{tk} -$$

$$-V_{kt}V^{-1}_{tt}V_{tu}V^{-1}_{uu.t}V_{uk}$$

$$= V_{kt}V^{-1}_{tt}V_{tk} + (V_{ku} - V_{kt}V^{-1}_{tt}V_{tu})V^{-1}_{uu.t}V_{uk} -$$

$$-V_{ku}V^{-1}_{uu}V_{ut}V^{-1}_{tt}V_{tu}V^{-1}_{uu.t}V_{ut}V^{-1}_{tt}V_{tk} +$$

$$+V_{kt}V^{-1}_{tt}V_{tu}V^{-1}_{uu.t}V_{ut}V^{-1}_{tt}V_{tk} -$$

$$-V_{ku}V^{-1}_{uu}V_{uu.t}V^{-1}_{uu.t}V_{ut}V^{-1}_{tt}V_{tk}$$

$$= V_{kt}V^{-1}_{tt}V_{tk} + (V_{ku} - V_{kt}V^{-1}_{tt}V_{tu})V^{-1}_{uu.t}V_{uk} -$$

$$-V_{ku}V^{-1}_{uu}V_{ut}V^{-1}_{tt}V_{tu}V^{-1}_{uu.t}V_{ut}V^{-1}_{tt}V_{tk} +$$

$$+V_{kt}V^{-1}_{tt}V_{tu}V^{-1}_{uu.t}V_{ut}V^{-1}_{tt}V_{tk} -$$

$$-V_{ku}V^{-1}_{uu.t}V_{ut}V^{-1}_{tt}V_{tk} +$$

$$+V_{ku}V_{uu}^{-1}V_{ut}V_{tt}^{-1}V_{tu}V_{uu.t}^{-1}V_{ut}V_{tt}^{-1}V_{tk}$$

$$= V_{kt}V_{tt}^{-1}V_{tk} + V_{ku.t}V_{uu.t}^{-1}V_{uk} - (V_{ku} - V_{kt}V_{tt}^{-1}V_{tu})V_{uu.t}^{-1}V_{ut}V_{tt}^{-1}V_{tk}$$

$$= V_{kt}V_{tt}^{-1}V_{tk} + V_{ku.t}V_{uu.t}^{-1}(V_{uk} - V_{ut}V_{tt}^{-1}V_{tk})$$

$$= V_{kt}V_{tt}^{-1}V_{tk} + V_{ku.t}V_{uu.t}^{-1}V_{uk.t}$$

so that we may write

$$\Lambda_{k(k+1,\ldots,m)} = \frac{|V_{kk.t} - V_{ku.t}V_{uu.t}^{-1}V_{uk.t}|}{|V_{kk}|} \ .$$

Then, since

$$\Lambda_{k(k+1,\ldots,i-1,i+1,\ldots,m)} = \Lambda_{kt} = \frac{|V_{kk} - V_{kt}V_{tt}^{-1}V_{tk}|}{|V_{kk}|} = \frac{|V_{kk.t}|}{|V_{kk}|}$$

we have

$$\Lambda_{ki|(k+1,\ldots,i-1,i+1,\ldots,m)} = \Lambda_{ku|t} = \frac{|V_{kk.t} - V_{ku.t}V_{uu.t}^{-1}V_{uk.t}|}{|V_{kk.t}|}$$

where, from the definitions of $u = \{i\}$ and $t = \{k+1, \ldots, i-1, i+1, \ldots, m\}$, from Lemma 1 and from v) in Lemma 4

$$V_{kk.t} \overset{app}{\sim} W_{p_k}\big((h - (p_{k+1} + \ldots + p_{i-1} + p_{i+1} + \ldots + p_m),\ b\Sigma_{kk}\big) \qquad (4.14)$$

$$\left(\text{with } h = \frac{(1-l)^2}{g}\ ,\quad b = \frac{g}{1-l}\right)\ ,$$

and from xi) in Lemma 4

$$V_{ku.t}V_{uu.t}^{-1}V_{uk.t} \overset{app}{\sim} W_{p_k}\left(p_i,\ b\Sigma_{kk}\right)\ .$$

Then, from xii) in Lemma 4, under H_0, $V_{kk.t}$ and $V_{ku.t}V_{uu.t}^{-1}V_{uk.t}$ are independent and thus

$$\left(V_{kk.t} - V_{ku.t}V_{uu.t}^{-1}V_{uk.t}\right) \overset{app}{\sim} W_{p_k}\left(h - \sum_{k=i+1}^{m} p_k,\ b\Sigma_{kk}\right) . \tag{4.15}$$

But then, again from xii) in Lemma 4, under H_0, $V_{kk.t} - V_{ku.t}V_{uu.t}^{-1}V_{uk.t}$ and $V_{ku.t}V_{uu.t}^{-1}V_{uk.t}$ are independent and thus, under H_0, $\Lambda_{ku|t}$ has the approximate Wilks' lambda distribution

$$\Lambda_{ku|t} = \Lambda_{ki|(k+1,\dots,i-1,i+1,\dots,m)} \overset{app}{\sim} \Lambda\Big(h - (p_{k+1} + \dots + p_{i-1} + p_{i+1} + \dots + p_m),\ p_k,\ p_i\Big). \tag{4.16}$$

But we may notice that if $H_0^{(k)}$ holds, then $\underline{x}_{(k)}$ is independent of $\underline{x}_{(i)}$, since $m > i > k$. Thus, from Lemma 1 and Lemma 2

$$\begin{bmatrix} V_{kk} & V_{ku} \\ V_{uk} & V_{uu} \end{bmatrix} \overset{app}{\sim} W_{p_k+p_i}\left(h,\ \begin{bmatrix} b\Sigma_{kk} & 0 \\ 0 & b\Sigma_{uu} \end{bmatrix}\right) .$$

Then, making the transformation

$$\begin{bmatrix} V_{kk} & V_{ku} \\ V_{uk} & V_{uu} \end{bmatrix} \longrightarrow \begin{bmatrix} V_{kk.t} & V_{ku.t} \\ V_{uk.t} & V_{uu.t} \end{bmatrix}$$

we obtain (see the proof of xii) in Lemma 4)

$$\begin{bmatrix} V_{kk.t} & V_{ku.t} \\ V_{uk.t} & V_{uu.t} \end{bmatrix}$$

$$\overset{app}{\sim} W_{p_k+p_i}\left(h - (p_{k+1} + \dots + p_{i-1} + p_{i+1} + \dots + p_m),\ \begin{bmatrix} b\Sigma_{kk} & 0 \\ 0 & b\Sigma_{uu} \end{bmatrix}\right) . \tag{4.17}$$

But then we may also notice that if $H_0^{i(k)}$ holds, then, from Lemmas 1 and 2, we can directly obtain the joint distribution of $V_{kk.t}$, $V_{ku.t}$ and $V_{uu.t}$ as given in (4.17), from the notion of conditional canonical correlations between sets $\underline{x}_{(k)}$ and $\underline{x}_{(i)}$ given the sets indexed in $t = \{k+1, ..., i-1, i+1, ..., m\}$.

Thus, (4.17) holds under both $H_0^{(k)}$ and $H_0^{i(k)}$. Then applying i), v) and vi) from Lemma 2 or i), viii) and ix) from Lemma 4 in (4.17) we obtain the above distributions in (4.14) and (4.15) as well as the independence of $V_{kk.t} - V_{ku.t}V_{uu.t}^{-1}V_{uk.t}$ and $V_{ku.t}V_{uu.t}^{-1}V_{uk.t}$ under both $H_0^{(k)}$ and $H_0^{i(k)}$ and thus the distribution (4.16) holds also under $H_0^{(k)}$ and $H_0^{i(k)}$. ∎

4.8 The Generalized Canonical Analysis and the iterative modeling approach.

In the early years of progress in the statistical science, modeling was seen as a one step approach. However, nowadays modeling is viewed as an iterative procedure. Revising and rebuilding is done sequentially till a model that serves the purpose of prediction and explanation is finally obtained.

The GCA approach is such an iterative model and occurs naturally when dealing with models which are particular cases of the GCA.

Suppose we wish to test the null hypothesis (4.3), i.e.

$$H_0 \;:\; \underline{x}_k \text{ is independent of } \underline{x}_{k'} \text{ for any } k \neq k',\; k, k' \in \{1, \ldots, m\}\;,$$

versus

$$\begin{aligned} & H_1 \;:\; \text{there is at least one } \underline{x}_{k'} \text{ that is not independent of } \underline{x}_k \; (k \neq k') \\ \Leftrightarrow\; & H_1 \;:\; \text{there exists } \Sigma_{kk'} \neq 0 \quad \text{for} \quad k \neq k'\;. \end{aligned}$$

As seen earlier, the generalized Wilks' lambda from the GCA model may be written, (under the null hypothesis), as a product of $m-1$ independent Wilks' lambdas, that refer to the analysis of only two sets of variables.

We shall begin by laying down some guidelines about model building and model fitting.

Models are generally based on some hypothesis. For example, it doesn't make much sense to write a regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

unless the variable y is correlated with the variables x_1 and x_2 and this hypothesis has been preliminarily tested. One common way to build models in Multiple Regression is through the use of the backward elimination procedure, where one fits all the regressor variables first and then removes those that in the presence of the others do not contribute significantly. We shall use a similar technique with the GCA type models.

In GCA we deal with sets of variables rather than a single vector of variables. Many times, all sets of variables are to be analyzed simultaneously, there being no such distinction as 'dependent' and 'independent' set of variables. Anyway, we may still think about modeling one or more sets as (linear) functions of the other sets with the idea of getting the best fit for the variables in that set by a linear function of the variables in the other sets. We will call the set being modeled as the 'predictand' set and the others as predictor sets. Any of the sets can thus be a 'predictand' set.

As shown above in Theorem 4 in section 4.7, under the null hypothesis

$$H_0 \ : \ \Sigma_{kj} = 0, \quad k = 1, ..., m-1; \ j = k+1, ..., m \ ,$$

the generalized Wilks' lambda is the product of $m-1$ independent Wilks' lambdas. These Wilks' lambdas provide tests of hypotheses of the type

$$H_0^{(k)} \ : \ \Sigma_{k(k+1,...,m)} = 0 \qquad \text{versus} \qquad H_1^{(k)} \ : \ \Sigma_{k(k+1,...,m)} \neq 0$$

(for $k = 1, ..., m-1$) and where for the sake of mathematical completeness we assume that $H_0^{(1)}$ is empty.

We may then write

$$H_0 = \bigcap_{k=1}^{m-1} H_0^{(k)} \ .$$

The 'best' prediction in the least squares sense of $\underline{x}_k$, using $\underline{x}_{k+1}, ..., \underline{x}_m$ is given by the model

$$\begin{aligned} \mathcal{M}\left(\underline{x}_{(k)} \middle| H_1^{(k)}\right) \ : \ \beta_{k(k)}\underline{x}_{(k)} \ &= \ \Psi_{k(k+1,...,m)}\beta_{k(k+1,...,m)}\left(\underline{x}'_{(k+1)}, ..., \underline{x}'_{(m)}\right)' + \underline{\epsilon}_k \\ &= \ \Psi_{k(k+1,...,m)}\left(\beta_{k(k+1)}\underline{x}_{(k+1)} + ... + \beta_{k(m)}\underline{x}_{(m)}\right) + \underline{\epsilon}_k \end{aligned}$$

where $\beta_{k(l)}$ $(l = k, ..., m)$ are $p_k^* \times p_l$ parameter matrices with $\beta_{k(k+1,...,m)} = \left[\beta_{k+1}\middle|...\middle|\beta_m\right]$, $\Psi_{k(k+1,...,m)}$ is the $p_k^* \times p_k^*$ diagonal matrix of the nonnull canonical correlations between $\underline{x}_{(k)}$ and $\left(\underline{x}_{(k+1)}, ..., \underline{x}_{(m)}\right)$ and $\underline{\epsilon}_k$ is $N_{p_k^*}(0, \ (I_{p_k^*} - \Psi^2_{k(k+1,...,m)}))$, with $p_k^* \leq min(p_k, p_{k+1} + ... + p_m)$.

The overall GCA model under H_1 may then be written as

$$\mathcal{M}(\underline{x}_{(1)}, \underline{x}_{(2)}, ..., \underline{x}_{(m)}|H_1) = \bigwedge_{k=1}^{m-1} \mathcal{M}(\underline{x}_{(k)}|H_1^{(k)}) \ \equiv$$

$$\begin{aligned} \equiv \ \ & \beta_{1(1)}\underline{x}_{(1)} = \Psi_{1(2,...,m)}\beta_{1(2,...,m)}\left(\underline{x}'_{(2)}, \ldots, \underline{x}'_{(m)}\right)' + \underline{\epsilon}_1 \quad \bigwedge \\ & \beta_{2(2)}\underline{x}_{(2)} = \Psi_{2(3,...,m)}\beta_{2(3,...,m)}\left(\underline{x}'_{(3)}, \ldots, \underline{x}'_{(m)}\right)' + \underline{\epsilon}_2 \quad \bigwedge \ \ \cdots \ \ \bigwedge \\ & \beta_{k(k)}\underline{x}_{(k)} = \Psi_{k(k+1,...,m)}\beta_{k(k+1,...,m)}\left(\underline{x}'_{(k+1)}, \ldots, \underline{x}'_{(m)}\right)' + \underline{\epsilon}_k \quad \bigwedge \ \ \cdots \ \ \bigwedge \\ & \beta_{m-2(m-2)}\underline{x}_{(m-2)} = \Psi_{m-2(m-1,m)}\beta_{m-2(m-1,m)}\left(\ \underline{x}'_{(m-1)} \ , \ \underline{x}'_{(m)} \ \right)' + \underline{\epsilon}_{m-2} \quad \bigwedge \\ & \beta_{m-1(m-1)}\underline{x}_{(m-1)} = \Psi_{m-1(m)}\beta_{m-1(m)}\underline{x}_{(m)} + \underline{\epsilon}_{m-1} \ , \end{aligned} \tag{4.18}$$

where $\bigwedge_{k=1}^{m-1}$ represents a conjunction of models and where some of the models

$$\beta_{k(k)}\underline{x}_{(k)} = \Psi_{k(k+1,\ldots,m)}\beta_{k(k+1,\ldots,m)}\left(\underline{x}'_{(k+1)}, \ldots, \underline{x}'_{(m)}\right)' + \underline{\epsilon}_k \qquad (k = 1, ..., m-1)$$

may be such that $\Psi_{k(k+1,\ldots,m)}$ is null (in other words, all the true canonical correlations between the set $\underline{x}_{(k)}$ and all the $m-k$ sets $\underline{x}_{(k+1)}$ through $\underline{x}_{(m)}$ may be null).

To test H_0 is then equivalent to testing sequentially $H_0^{(k)}$ for $k = 2, ..., m$ (Anderson, 1984; Andersson and Perlman, 1988). Rejecting one of the hypothesis $H_0^{(k)}$ will then imply the rejection of H_0.

When one of the hypothesis $H_0^{(k)}$ is rejected then we may want to test the set of $m-k$ hypotheses

$$H_0^{i(k)} \; : \; \Big(\Sigma_{ki} \Big| \bigwedge_{\substack{l=k+1 \\ l \neq i}}^{m} \Sigma_{kl}\Big) = 0 \qquad i = k+1, ..., m \; .$$

[$\Longleftrightarrow$ $H_0^{i(k)}$: all the canonical correlations between$\underline{x}_{(k)}$ and $\underline{x}_{(i)}$, given the canonical correlations between $\underline{x}_{(k)}$ and the sets $\underline{x}_{(k+1)}$ through $\underline{x}_{(i-1)}$ and $\underline{x}_{(i+1)}$ through $\underline{x}_{(m)}$, are null.]

We may also notice that

$$H_0^{(k)} = \bigcap_{i=k+1}^{m} H_0^{i(k)} \; .$$

When one of the vectors $\underline{x}_{(k)}$ $(k = 1, \ldots, m)$, say $\underline{x}_{(1)}$ is definitely a dependent one then our approach may be slightly different. We will then be mainly interested in testing the two (null) hypotheses

$$H_0^{(1)} \; : \; \Sigma_{1(2,\ldots,m)} = 0 \qquad \text{versus} \qquad H_1^{(1)} \; : \; \Sigma_{1(2,\ldots,m)} \neq 0$$

and

$$H_0^* \; : \; \Sigma_{ij} = 0 \quad i = 2, ..., m-1; \; j = i+1, ..., m$$

with

$$H_0^* = \bigcap_{k=2}^{m} H_0^{(k)} \; .$$

Our main interest is in $H_0^{(1)}$ and we may also note that $H_0^* \bigcap H_0^{(1)} = H_0$.

If $H_0^{(1)}$ is rejected and then we use the model

$$\mathcal{M}(\underline{x}_{(1)}|H_1^{(1)}) \;:\; \beta_{1(1)}\underline{x}_{(1)} = \Psi_{1(2\ldots m)}\beta_{1(2\ldots m)}\left(\underline{x}'_{(2)},\ldots,\underline{x}'_{(m)}\right)' + \epsilon_1$$

we may then be interested in testing H_0^* or its subhypotheses to know whether the model $\mathcal{M}(\underline{x}_{(1)}|H_1^{(1)})$ is of the type

$$\mathcal{M}(\underline{x}_{(1)}|\underline{x}_{(2)},\ldots,\underline{x}_{(m)}|H_1^{(1)})$$

where all the sets $\underline{x}_{(2)}$, ..., $\underline{x}_{(m)}$ are independent, or it is of the type

$$\mathcal{M}\left(\underline{x}_{(1)}|\beta_k\underline{x}_{(k)} = \Psi_{k(k+1,\ldots,m)}\beta_{k(k+1,\ldots,m)}\left(\underline{x}'_{(k+1)},\ldots,\underline{x}'_{(m)}\right)' + \epsilon_k\;,\; k=2,\ldots,m-1|H_1^{(1)}\right)$$

where at least one of the sets $\underline{x}_{(2)}$, ..., $\underline{x}_{(m)}$ is not independent of the others.

If $H_0^{(1)}$ is rejected then we may be interested in testing the sequence of hypotheses

$$H_0^{i(1)} \;:\; \Sigma_{1i} = 0 \qquad (i = 2,\ldots,m)$$

since

$$H_0^{(1)} = \bigcap_{i=2}^{m} H_0^{i(1)}$$

and then to drop out of the model the sets for which $H_0^{s(1)}$ is not rejected (these sets if retained in the model will only confuse it since if they are correlated with other sets in the model they will influence β_1 and $\beta_{1(2,\ldots,m)}$, correlated meaning that not all true canonical correlations between sets are zero.)

So, in this model building process, we will test $H_0^{(1)}$ first and if it is rejected then we should follow it up by testing the $m-1$ subhypotheses $H_0^{i(1)}$ $(i = 2,\ldots,m)$. The sets $\underline{x}_{(k)}$ of variables retained in the model would be the ones for which $H_0^{i(1)}$ was rejected.

A special case of the GCA model building procedure will be when all the sets of variables $\underline{x}_{(1)}$ through $\underline{x}_{(m)}$ consist of only one variable.

In this particular case the GCA model

$$\mathcal{M}(\underline{x}_{(1)}|H_1^{(1)})$$

is a Principal Components Analysis model, while each of the models corresponding to the test of each one of the subhypotheses $H_0^{(k)}$ $(k = 2, ..., m)$ is a regression model corresponding to the regression of $\underline{x}_{(k)}$ on $\underline{x}_{(k+1)}$ through $\underline{x}_{(m)}$, while the model corresponding to the test of $H_0^* = \bigcap_{k=2}^m H_o^{(k)}$ as an whole is a GCA model that in this special case is a Principal Components Analysis among the regressor variables $\underline{x}_{(2)}$ through $\underline{x}_{(m)}$. This is the theoretical justification to carry out a preliminary Principal Components Analysis on the regressor variables before fitting a Multiple Regression model, in order to better understand the structure of the correlations among the regressor variables. Sometimes a Principal Components Analysis is carried out on all the regressor variables as well as the dependent or modeling variable. This procedure is also justified since it corresponds to the fit of the model under the alternative hypothesis H_1.

4.9 Testing the hypotheses in sections 4.7 and 4.8.

The hypotheses specified in sections 4.7 and 4.8 may be tested using the test statistics derived in Theorems 4 and 5 in section 4.7. More specifically, to each hypothesis $H_0^{(k)}$ in 4.7 or 4.8 corresponds the test statistic

$$\Lambda_{k(k+1,...,m)} \overset{app}{\sim} \Lambda\left(\frac{(1-l)^2}{g},\ p_k,\ (p_k + ... + p_m)\right) \qquad (k = 1, ..., m-1)$$

derived in Theorem 4 in section 4.7., and to each hypothesis $H_0^{i(k)}$ in section 4.7 or 4.8, corresponds the test statistic

$$\Lambda_{ki|(k+1,...,i-1,i+1,...,m)} \overset{app}{\sim} \Lambda\left(\frac{(1-l)^2}{g},\ p_k,\ p_i\right)$$

$$(k = 1, ..., m-1)\ (i = k+1, ..., m)\ .$$

Given that the above test statistics are the likelihood ratio test statistics for the corresponding hypotheses, in order to obtain the p-value for the corresponding tests of hypotheses we may use the large sample distribution of the likelihood ratio test statistic. But, better approximations have been derived by Bartlett (1938), Box (1949) and Rao (1951) (see Kshirsagar (1972) for a reference). Usually Bartlett's approximation is good enough for all practical purposes.

Recently Srivastava and Yau (1989) obtained an approximation to the tail probabilities of the Wilks' lambda that is better behaved than Box's approximation for small number of error degrees of freedom. Also Schatzoff (1966), Pillai and Gupta (1969) and Lee (1972) give tables of correction factors for the p-values obtained from Bartlett's approximation.

A detailed ANOVA table for the model (4.18) in section 4.8 is given below.

Table 4.1 – ANOVA table for model (4.18) in section 4.8

($h = \frac{(1-l)^2}{g}$ — see Lemma 1 for definition of l and g)

Hypothesis being tested	Sum of Squares Matrix	degrees of freedom =d.f.(Wishart Matrix)
$H_0^{(k)}$ $_{(k=1,\dots,m-1)}$ Association $\underline{x}_{(k)}$ vs. $(\underline{x}_{(k+1)},\dots,\underline{x}_{(m)})$		
Error	$V_{kk} - V_{k(k+1,\dots,m)} V^{-1}_{(k+1,\dots,m)(k+1,\dots,m)} V_{(k+1,\dots,m)k}$	$h - (p_{k+1} + \dots + p_m)$
Hyp.	$V_{k(k+1,\dots,m)} V^{-1}_{(k+1,\dots,m)(k+1,\dots,m)} V_{(k+1,\dots,m)k}$	$p_{k+1} + \dots + p_m$
Total	V_{kk}	h
$H_0^{i(k)}$ $_{(i=k+1,\dots,m)}$ Importance of $\underline{x}_{(i)}$ in the above assoc.		
Error	$V_{kk.t} - V_{ki.t} V^{-1}_{ii.t} V_{ik.t}$	$h - (p_{k+1} + \dots + p_{i-1} + p_{i+1} + \dots + p_m) - p_i$
Hyp.	$V_{ki.t} V^{-1}_{ii.t} V_{ik.t}$	p_i
Total	$V_{kk.t}$	$h - (p_{k+1} + \dots + p_{i-1} + p_{i+1} + \dots + p_m)$

4.10 About the number of parameters in the parameter matrices in 4.8.

An interesting problem, relative to the models in section 4.8 is the determination of the number of parameters that we should retain in each of the β parameter matrices in the model (4.18).

The answer to this problem depends on the number of true non-null canonical correlations between each pair of sets in the model (4.18). This is also referred to as the dimensionality of the relationships. In order to assess this dimensionality the usual test for the number of non-null canonical correlations may be used (see for example Kshirsagar, 1972), and then the number of non-null population canonical correlations that are different from zero in the relation between $\underline{x}_{(i)}$ and the set $(\underline{x}_{(i+1)}, ..., \underline{x}_{(m)})$ may then be calculated for $i = 1, ..., m-1$. This number is then the required number of parameter columns in each matrix $\beta_i(i+1, ..., m)$ and β_i.

Another associated problem is the one of testing the adequacy of a given set of columns of $\beta_i(i+1, ..., m)$ and β_i to describe the association between $\underline{x}_{(i)}$ and $(\underline{x}_{(i+1)}, ..., \underline{x}_m)$. The given set of parameters may be suggested by an experimenter from some early experience or prior reasoning or even from some theoretical considerations. Then we may want to test if the assigned parameters really are adequate to describe the association between the two sets. This is accomplished by factorizing each test statistic $\Lambda_{i(i+1,...,m)}$ into its direction and collinearity factors (see Bartlett, 1951, Kshirsagar and Gupta, 1982).

4.11 Particular cases of the general GCA model.

Well known particular cases of the GCA model presented above (in section 4.8) arise under several situations. For example, when all the m sets of variables, $\underline{x}_{(1)}$ through $\underline{x}_{(m)}$, have only one variable then the general GCA model reduces to the well known Principal Components Analysis model. Also, when we have only two sets, $\underline{x}_{(1)}$ and $\underline{x}_{(2)}$, then the GCA model reduces to the usual Canonical Analysis model (see section 3.8 of Chapter 3). If one of these sets consists of indicator variables corresponding to the categories of a categorical variable then the GCA or Canonical Analysis model becomes a Multiple Discriminant Analysis model if the categories of such categorical variable represent different populations. Yet another interesting particular case of the GCA model occurs when all the m sets of variables $\underline{x}_{(1)}$ through $\underline{x}_{(m)}$ consist of indicator variables corresponding to the categories of m attributes. This particular case called Multiple (or Simple if $m = 2$) Correspondence Analysis is studied in detail in Chapter 6.

It should also be noted that the usual Multiple Linear Regression and Multivariate (or Univariate) Analysis of Variance and Covariance models are particular cases of the GCA model. If we have a situation where, let us suppose, $\underline{x}_{(1)}$ is formed by one or more continuous variables and the other sets are formed by the indicator variables for the levels of experimentally controlled factors, then we have a Univariate or Multivariate Analysis of Variance model. In a similar situation if there is a set of 'regressor' variables that has one or more continuous variables then we have a Univariate or Multivariate Covariance analysis model. These and other models, usually not studied in the literature, where there is one set of variables that clearly is seen as the 'dependent' set may also be studied under the GCA approach and thus the test statistics and models discussed in sections 4.7 through 4.9 may still be applied to them. These models will be studied in some more detail in Chapter 7 where a generalization of the commonly used partial F test is given.

The GCA is therefore a unifying approach to all such models, where relationships among vector variables is studied for prediction purposes.

In this Chapter we assumed that all our sets of variables had a multivariate normal distribution and then all the distributional details of the test statistics have been derived. In future Chapters we will deal with different situations. Namely in Chapter 6 we will deal with the situation when all the sets of variables consist of indicator variables for the categories of m attributes and are thus not normally but distributed but rather multinomially distributed. In Chapter 7 we will study the Block Regression model, where the situation is similar to the Multiple Linear Regression model but where we deal with sets of variables rather than with individual variables. Often in this later type of models it is assumed a multivariate normal distribution at least for the 'dependent' set, but there may also be cases where such set is formed by the indicator variables for the categories of some attribute. It will be shown in these Chapters that in each of those cases we may still use the models developed in section 4.8 and that we may still apply the test statistics developed in section 4.7, their distributions still holding, at least approximately.

CHAPTER 5

ASYMPTOTIC DISTRIBUTION OF THE LIKELIHOOD RATIO TEST STATISTIC TO TEST THE INDEPENDENCE OF m SETS OF VARIABLES.

5.1 Introduction.

As shown in 4.7, the generalized Wilks' lambda, when H_0 is true, may be split into $m-1$ independent Wilks' lambdas.

However, many times one is rather interested in a quick assessment of the independence of the m sets of variables. The testing procedure described in 4.8, is long and involves too much work. Then a quick way to test the independence of the m sets of variables is to test the significance of the generalized Wilks lambda. For that we need its distribution — at least an approximate one for large samples. This is studied in this chapter. Of course, one can use the general result about likelihood ratio test (L.R.) criteria, namely that $-2log(L.R.)$ is a chi-square with f degrees of freedom, where f is the difference in the number of degrees of parameters in the likelihood, that are estimated under H_0 and under the alternative. Many times this result is true only for very large samples. That is why in the literature one finds several improvements on this approximation, by using correction factors such as

Bartlett's (1938), Box's (1949) or Rao's (1951). Various asymptotic approximations exist in the literature for the usual Wilks' lambda. The usual tests are either χ^2 or F. In this chapter we propose yet another approximation. It is based on a normal distribution. No attempt is made to compare this approximation with others. It is intended to pursue this in the future.

5.2 Literature survey.

The exact distribution of Wilks' lambda is known. More specifically, its exact moments are known (see 4.4, Definition 2 and Theorem 4 in 4.7, see also Anderson, 1958, 1984; Muirhead, 1982) and the statistic has a finite range hence the moments determine the distribution uniquely even though an expression for its p.d.f. in a closed form is not possible. In the case $m = 2$, the exact p.d.f. of Λ has been obtained explicitly (Wilks, 1935; Schatzoff, 1966; Consul, 1966, 1967; Pillai and Gupta, 1969; Lee, 1972; Nandi, 1977). However, the exact p.d.f. of Λ is too complicated and of little practical use. For this reason, asymptotic approximations are often preferred for practical purposes. For $m = 2$, if more precise percentage points are needed, the results of Schatzoff (1966), Pillai and Gupta (1969) and Lee (1972) may be used. They have tabulated the correction factors for converting chi-square percentiles to the exact percentiles of the Wilks' lambda. Also, in a recent paper Srivastava and Yau (1989), using the saddlepoint method, obtained two explicit approximation formulae for the tail probabilities of Wilks' criterion which perform better than Box's approximation. Asymptotic null and non-null distributions of Λ are also known. They have been extensively studied for the case $m = 2$ (Rao, 1951; Sugiura, 1969; Sugiura and Fujikoshi, 1969; Lee, 1971a, 1971b, 1972; Muirhead, 1972; Kulp and Nagarsenker, 1984; Muller and Peterson, 1984) and even for the case $m > 2$ (Nagao, 1972, 1973a, 1973b).

As mentioned in the introduction, we approach this problem in a different way and obtain an asymptotic normal approximation to the distribution of Λ. Nagao (1973b) says that the distribution of Λ in the general case $m > 2$ is asymptotically normal under the alternative hypothesis but he observed that under the null hypothesis H_0 these limiting distributions were singular. We will show how, using a better expansion technique, we can avoid the singularity and get a normal approximation to the distribution of Λ. A normal approximation is of course always quite appealing.

5.3 An asymptotic distribution of $|V|$.

Assuming that $V^* \sim W_p\left(n-1,\ \frac{1}{n}\Sigma\right)$, asymptotic distributions for $|V^*|$ are usually obtained under the form

$$\sqrt{\frac{n-1}{2p}}\,\log\frac{|V^*|}{|\Sigma|} \overset{a}{\sim} N(0,1) \tag{5.1}$$

(Theorem 3.2.16 of Muirhead (1982)), or equivalently,

$$\sqrt{\frac{n-1}{2p}}\left(\frac{|V^*|}{|\Sigma|}-1\right) \overset{a}{\sim} N(0,1) \tag{5.2}$$

as in Theorem 7.5.4. of Anderson (1958, 1984).

However, they both have some limitations. The first one we may notice is that it is known that if $V^* \sim W_p\left(n-1,\ \frac{1}{n}\Sigma\right)$ then $\mathcal{E}\left(|V^*|\right) = |\Sigma|\prod_{j=1}^{p}\left[1-\frac{1}{n-1}(j-1)\right] < |\Sigma|$, with $\mathcal{E}\left(|V^*|\right) \to |\Sigma|$ as $n \to \infty$, as it is shown, for example, in the sequence of Theorem 3.2.15 of Muirhead (1982). Any 'good' asymptotic distribution of $|V^*|$ should retain this property, but that is not the case with either (5.1) or (5.2) above, because $\mathcal{E}\left(\log|V^*|\right) = \log|\Sigma|$, in (5.1) and $\mathcal{E}\left(|V^*|\right) = |\Sigma|$, in (5.2).

The result usually used in obtaining (5.1) and (5.2) is

$$\sqrt{\frac{n-1}{2}}\left(\log\chi^2_{n-i} - \log(n-1)\right) \overset{a}{\sim} N(0,1)$$

or

$$\sqrt{\frac{1}{2(n-1)}}\left(\chi^2_{n-i}-n+1\right)\ \overset{a}{\sim}\ N(0,1)$$

instead of using the Central Limit Theorem result

$$\sqrt{\frac{1}{2(n-i)}}\left(\chi^2_{(n-i)}-(n-i)\right)\ \overset{a}{\sim}\ N(0,1)\ .$$

Any asymptotic distribution obtained should as far as possible preserve known properties about the exact distribution. We shall derive our asymptotic distribution bearing this in mind.

We shall use the following Taylor series expansion of order r of a function $g(T)$ around the value θ.

$$g(T)=\sum_{\alpha=0}^{r}\frac{g^{(\alpha)}(\theta)}{\alpha!}\ (T-\theta)^{\alpha}+R_{r+1}$$

where

$$R_{r+1}\ =\ \frac{g^{(r+1)}(\theta)}{(r+1)!}\ (T-\theta^*)^{r+1}\ ,$$

is the remainder term, and

$$\theta^*\ \in\ \Big(\min(T,\theta),\max(T,\theta)\Big)\ ,$$

and where $g^{(\alpha)}(\theta)$ represents the value of the α^{th} derivative of $g(.)$ at the point θ.

Theorem 1:

Let $T\sim\chi^2_k$. Then, using a Taylor series expansion of order 1

$$\text{i)}\quad \mathcal{E}(\log T)\simeq\log k$$

$$\text{ii)}\quad Var(\log T)\simeq 2k^{-1}\ .$$

Using a Taylor series expansion of order 2,

$$\text{iii)}\quad \mathcal{E}(\log T)\simeq\log k-k^{-1}$$

$$\begin{aligned}\text{iv)}\quad Var(\log T)&\simeq 2k^{-1}+\tfrac{1}{4}(k+6)(k+4)(k+2)k^{-3}-2(k+4)(k+2)k^{-2}\\&\quad+\tfrac{9}{2}(k+2)k^{-1}-\left(2+\tfrac{3}{4}\right)\\&=2k^{-1}-5k^{-2}+12k^{-3}\ .\end{aligned}$$

And using a Taylor series expansion of order 3,

$$\begin{aligned}
\text{v)}\quad \mathcal{E}(\log T) &\simeq \log k - k^{-1} + \tfrac{1}{3}(k+4)(k+2)k^{-2} - (k+2)k^{-1} + \tfrac{2}{3} && (a)\\
&= \log k - k^{-1} + \tfrac{8}{3}k^{-2} && (b)\\
\text{vi)}\quad Var(\log T) &\simeq 2k^{-1} + \tfrac{1}{9}(k+10)(k+8)(k+6)(k+4)(k+2)k^{-5} \\
&\qquad -(k+8)(k+6)(k+4)(k+2)k^{-4} \\
&\qquad +\tfrac{9}{2}(k+6)(k+4)(k+2)k^{-3} && (a)\\
&\qquad -\left(10+\tfrac{2}{9}\right)(k+4)(k+2)k^{-2} \\
&\qquad +\tfrac{27}{2}(k+2)k^{-1} \\
&\qquad -\left(6+\tfrac{23}{36}\right) \\
&= 2k^{-1} + 43k^{-2} + 4k^{-3} + \left(103+\tfrac{1}{9}\right)k^{-4} + \left(426+\tfrac{2}{3}\right)k^{-5}\ , && (b)
\end{aligned}$$

or

$$\begin{aligned}
\text{vi}')\quad Var(\log T) &\simeq 2k^{-1} + \tfrac{1}{9}(k+10)(k+8)(k+6)(k+4)(k+2)k^{-5} \\
&\qquad -(k+8)(k+6)(k+4)(k+2)k^{-4} \\
&\qquad +\tfrac{9}{2}(k+6)(k+4)(k+2)k^{-3} && (a)\\
&\qquad -\left(10+\tfrac{2}{9}\right)(k+4)(k+2)k^{-2} \\
&\qquad +\tfrac{27}{2}(k+2)k^{-1} - \left(6+\tfrac{23}{36}\right) \\
&\qquad -k^{-2} + \left(5+\tfrac{1}{3}\right)k^{-3} - \left(7+\tfrac{1}{9}\right)k^{-4} \\
&= 2k^{-1} + 42k^{-2} + \left(9+\tfrac{1}{3}\right)k^{-3} + 96k^{-4} + \left(426+\tfrac{2}{3}\right)k^{-5}\ , && (b)
\end{aligned}$$

where in obtaining the variances in ii), iv) and vi) we used $log\ k$ for $\mathcal{E}(log\ T)$, an approximation derived from i), while in vi$'$) we used v) as an approximation for $\mathcal{E}(log\ T)$.

Proof:

We will only prove v), vi) and vi$'$).

Using the Taylor series expansion of order 3 we write

$$g(T) = g(\theta) + g'(\theta)(T-\theta) + \frac{1}{2}g''(\theta)(T-\theta)^2 + \frac{1}{6}g'''(\theta)(T-\theta)^3 + R_4$$

where

$$g(T) = \log T$$

and

$$\theta = \mathcal{E}(T),$$

where, since we are dealing with a χ^2 variable,

$$\theta = k \text{ (= number of degrees of freedom) .}$$

Then

$$\begin{aligned}
\mathcal{E}(g(T)) &\simeq g(\theta) + \tfrac{1}{2}g''(\theta)Var(T) + \tfrac{1}{6}g'''(\theta)\mathcal{E}(T-\theta)^3 \\
&= \log\theta - \tfrac{1}{2}\theta^{-2}Var(T) + \tfrac{1}{3}\theta^{-3}\mathcal{E}(T-\theta)^3 \ ,
\end{aligned}$$

and

$$\begin{aligned}
Var(g(T)) &\simeq \mathcal{E}(g(T) - g(\theta))^2 \\
&= \mathcal{E}\left(g'(\theta)(T-\theta) + \tfrac{1}{2}g''(\theta)(T-\theta)^2 + \tfrac{1}{6}g'''(\theta)(T-\theta)^3 \right)^2 \\
&= \mathcal{E}\Big([g'(\theta)]^2(T-\theta)^2 + g'(\theta)g''(\theta)(T-\theta)^3 + \tfrac{1}{3}g'(\theta)g'''(\theta)(T-\theta)^4 \\
&\quad +\tfrac{1}{4}[g''(\theta)]^2(T-\theta)^4 + \tfrac{1}{6}g''(\theta)g'''(\theta)(T-\theta)^5 \\
&\quad +\tfrac{1}{36}[g'''(\theta)]^2(T-\theta)^6\Big) \\
&= \theta^{-2}Var(T) - \theta^{-3}\mathcal{E}(T-\theta)^3 + \tfrac{11}{12}\theta^{-4}(T-\theta)^4 - \tfrac{1}{3}\theta^{-5}\mathcal{E}(T-\theta)^5 \\
&\quad +\tfrac{1}{9}\theta^{-6}\mathcal{E}(T-\theta)^6 \ .
\end{aligned}$$

Now, we use

$$\begin{aligned}
\mathcal{E}(T-\theta)^3 &= \mathcal{E}(T^3) - 3\theta\,\mathcal{E}(T^2) + 2\theta^3 \\
\mathcal{E}(T-\theta)^4 &= \mathcal{E}(T^4) - 4\theta\,\mathcal{E}(T^3) + 6\theta^2\,\mathcal{E}(T^2) - 3\theta^4 \\
\mathcal{E}(T-\theta)^5 &= \mathcal{E}(T^5) - 5\theta\,\mathcal{E}(T^4) + 10\theta^2\,\mathcal{E}(T^3) - 10\theta^3\,\mathcal{E}(T^2) + 4\theta^5 \\
\mathcal{E}(T-\theta)^6 &= \mathcal{E}(T^6) - 6\theta\,\mathcal{E}(T^5) + 15\theta^2\,\mathcal{E}(T^4) - 20\theta^3\,\mathcal{E}(T^3) + 15\theta^4\,\mathcal{E}(T^2) - 5\theta^6
\end{aligned} \tag{5.3}$$

and the fact that since $T \sim \chi^2_k$

$$\begin{aligned}
\mathcal{E}(T^r) = \frac{2^r\ \Gamma\left(\frac{k}{2}+r\right)}{\Gamma\left(\frac{k}{2}\right)} &= 2^r \textstyle\prod_{i=1}^{r}\left(\frac{k}{2}+r-i\right) \\
&= \textstyle\prod_{i=1}^{r}\Big(k+2(r-i)\Big)
\end{aligned} \tag{5.4}$$

and then $\mathcal{E}(T-\theta)^2 = Var(T) = 2k$.

Finally we get v.a) and vi.a) and by simplification v.b) and vi.b).

Now in order to get vi$'$) instead of using the approximation

$$\mathcal{E}(g(T)) \simeq g(\theta)$$

we use v) as an approximation for $\mathcal{E}(g(T))$. This renders the computations slightly more complicated. Then

$$\begin{aligned} Var(g(T)) &= \mathcal{E}(g(T) - \mathcal{E}(g(T)))^2 \\ &\simeq \mathcal{E}\Big(g(\theta) + g'(\theta)(T-\theta) + \tfrac{1}{2}g''(\theta)(T-\theta)^2 + \tfrac{1}{6}g'''(\theta)(T-\theta)^3 + R_4 \\ &\qquad -g(\theta) + k - \tfrac{1}{3}(k+4)(k+2)k^{-2} + (k+2)k^{-1} - \tfrac{2}{3}\Big)^2 \end{aligned}$$

(using now $g(.) = \log(.)$)

$$\begin{aligned} &= \mathcal{E}\Big(k^{-1}(T-\theta) - \tfrac{1}{2}k^{-2}(T-\theta)^2 + \tfrac{1}{3}k^{-3}(T-\theta)^3 + k^{-1} \\ &\qquad -\tfrac{1}{3}(k+4)(k+2)k^{-1} - \tfrac{2}{3}\Big)^2 \\ &= k^{-2}\,\mathcal{E}(T-\theta)^2 - k^{-3}\,\mathcal{E}(T-\theta)^2 + k^{-2} + (k+2)^2k^{-2} - \tfrac{4}{3}(k+2)k^{-1} \\ &\quad -\tfrac{2}{3}(k+4)(k+2)^2k^{-3} + \tfrac{4}{9}(k+4)(k+2)k^{-2} + \tfrac{1}{9}(k+4)^2(k+2)^2k^{-4} \\ &\quad + \Big(-\tfrac{13}{9}k^{-3} + \tfrac{2}{3}k^{-4} - \tfrac{2}{9}(k+4)(k+2)k^{-5} + \tfrac{2}{3}(k+2)k^{-4}\Big)\,\mathcal{E}(T-\theta)^3 \\ &\quad + \Big(\tfrac{2}{3} + \tfrac{1}{4}\Big)\,k^{-4}\,\mathcal{E}(T-\theta)^4 - \tfrac{1}{3}k^{-5}\,\mathcal{E}(T-\theta)^5 + \tfrac{1}{9}k^{-6}\,\mathcal{E}(T-\theta)^6\ , \end{aligned}$$

now keeping in mind that $\theta = k$ and using the above relations (5.3) and (5.4) we get vi$'$.a) and further simplifications held vi$'$.b). ∎

Theorem 2: If $V^* \sim W_p\left(n-1,\ \frac{1}{n}\Sigma\right)$, then (using a Taylor series expansion of order 3 and a remainder of order 4)

$$\log|V^*| \overset{a}{\sim} N\left(\log|\Sigma| + a + \sum_{j=1}^{2} b_j d_j\ \ ,\ \ \sum_{j=1}^{5} c_j d_j\right) \tag{5.5}$$

with

$$a = \textstyle\sum_{i=1}^{p} \log\left(\frac{n^*-i+1}{n}\right) \qquad , \qquad d_j = \textstyle\sum_{i=1}^{p}(n^*-i+1)^{-j}$$

$b_1 = -1, \quad b_2 = \frac{8}{3},$

$c_1 = 2, \quad c_2 = 43, \quad c_3 = 4, \quad c_4 = 103 + \frac{1}{9}, \quad c_5 = 426 + \frac{2}{3},$

and where $n^* = n - 1$.

Proof:

It is known that if $V^* \sim W_p\left(n-1, \ \frac{1}{n}\Sigma\right)$ then

$$n^p \frac{|V^*|}{|\Sigma|} = \prod_{i=1}^{p} \chi^2_{(n-i)} \tag{5.6}$$

where $\chi^2_{(n-i)}$ $(i = 1, \ldots, p)$ stands for p independently distributed χ^2 variables, degrees of freedom $(n-i)$ $(i = 1, \ldots, p)$.

We may write (5.6) as

$$p \log n + \log |V^*| - \log |\Sigma| = \sum_{i=1}^{p} \log \chi^2_{(n-i)} \ .$$

Let $\mathcal{M}_i$ and $\mathcal{V}_i$ represent respectively the approximate values of $\mathcal{E}(\log \chi^2_{k_i})$ and $Var(\log \chi^2_{k_i})$ given by v) and vi) from Theorem 1 above, for $i = 1, \ldots, p$. Let

$$T_i = \log \chi^2_{(n-i)} - \mathcal{M}_i \qquad (i = 1, \ldots, p) \ .$$

Then we know that, for $i = 1, \ldots, p$

$$\mathcal{E}(T_i) \simeq 0$$

$$Var(T_i) \simeq \mathcal{V}_i$$

and that all the T_i are independent and have the same type of distribution. Then we may apply Lindeberg's (or Lindeberg's-Feller) (Feller, 1968) variant of the Central Limit Theorem to obtain

$$\frac{\sum_{i=1}^{p} T_i}{\sqrt{\sum_{i=1}^{p} \mathcal{V}_i}} = \frac{\sum_{i=1}^{p}\left(\log \chi^2_{n-i} - \mathcal{M}_i\right)}{\sqrt{\sum_{i=1}^{p} \mathcal{V}_i}} \quad \overset{a}{\sim} \quad N(0,1)$$

or, by a standard argument,

$$\sum_{i=1}^{p} \log \chi^2_{(n-i)} \stackrel{a}{\sim} N\left(\sum_{i=1}^{p} \mathcal{M}_i \ , \ \sum_{i=1}^{p} \mathcal{V}_i\right)$$

or

$$p \log n + \log |V^*| - \log |\Sigma| \stackrel{a}{\sim} N\left(\sum_{i=1}^{p} \mathcal{M}_i \ , \ \sum_{i=1}^{p} \mathcal{V}_i\right) \tag{5.7}$$

or

$$\log |V^*| \stackrel{a}{\sim} N\left(\log |\Sigma| + \sum_{i=1}^{p} (\mathcal{M}_i - \log n) \ , \ \sum_{i=1}^{p} \mathcal{V}_i\right) . \tag{5.8}$$

Now all we have to do is to substitute $\mathcal{M}_i$ and $\mathcal{V}_i$ in (5.8) by the formulas in v.b) and vi.b) of Theorem 1 where in turn we replace k by $n-i$. ∎

Note that (5.7) has a form similar to (5.1), the result in Theorem 3.2.15 in Muirhead (1982) and (5.2), Theorem 7.5.4 in Anderson (1958, 1984) but now it has the desirable property that $\mathcal{E}(|V^*|) < |\Sigma|$ and at the same time $\mathcal{E}(|V^*|) \to |\Sigma|$ as $n \to \infty$, while the variance is not a function of $|\Sigma|$, contrary to (5.2).

In a similar way, we obtain the following results.

Corollary 1:

Using a Taylor series expansion of order 1,

$$\log |V^*| \stackrel{a}{\sim} N\left(\log |\Sigma| + a \ , \ \sum_{i=1}^{p} 2(n^* - i + 1)^{-1}\right)$$

with a defined in Theorem 2 above.

Using a Taylor series expansion of order 2,

$$\log |V^*| \stackrel{a}{\sim} N\left(\log |\Sigma| + a - \sum_{i=1}^{p} (n^* - i + 1)^{-1} \ , \ \sum_{i=1}^{p} \sum_{j=1}^{3} c'_j (n^* + i - 1)^{-j}\right)$$

with a defined as in Theorem 2 and

$$c'_1 = 11, \quad c'_2 = 7, \quad c'_3 = 12 \ . \ ∎$$

In our case, V does not have an exact Wishart distribution but rather the approximate one (as shown in Lemma 1 of Chapter 4)

$$V \overset{app}{\sim} W_p\left(\frac{(1-l)^2}{g},\ \frac{g}{1-l}\Sigma\right)\ , \tag{5.9}$$

thus, in order to obtain the asymptotic distributions of $log|V|$ we need to replace n by $\frac{g}{1-l}$ in the expression for a, and n^* by $\frac{(1-l)^2}{g}$ in the above distributions in Theorem 2 and in Corollary 1.

5.4 An asymptotic distribution of the likelihood ratio test statistic.

Theorem 3: (An asymptotic null distribution of Λ)

If $V^* \sim W_p\left(n-1,\ \frac{1}{n}\Sigma\right)$, with Σ partitioned as in (4.2), then under the null hypothesis $H_0 : \Sigma_{ij} = 0,\ \forall i \neq j;\ i,j \in \{1,\ldots,m\}$, and based on a Taylor series expansion of order 3, the distribution of $\log\Lambda^*$, where (for a partition of V^* corresponding to the partition of Σ) $\Lambda^* = \frac{|V^*|}{\prod_{k=1}^{m}|V^*_{kk}|}$, is asymptotically normal with

$$\log\Lambda \overset{a}{\sim} N\left(a^* + b^*,\ c^*\right)$$

where

$$a^* = a - \sum_{k=1}^{m} a_k \quad , \quad b^* = \sum_{j=1}^{2} b_j e_j \quad , \quad c^* = \sum_{j=1}^{5} c_j e_j$$

with

$$a_k = \sum_{i=1}^{p_k} log\frac{n^*-i+1}{n} \quad \text{and} \quad e_j = \sum_{i=1}^{p}(n^*-i+1)^{-j} - \sum_{k=1}^{m}\sum_{i=1}^{p_k}(n^*-i+1)^{-j}\ . \tag{5.10}$$

with $n^* = n-1$ and where a, b_j and c_j are defined as in Theorem 2.

Proof:

Under the above null hypothesis $H_0 : \Sigma_{ij} = 0,\ \forall i \neq j;\ i,j \in \{1,\ldots,m\}$, and given that $V^* \sim W_p\left(n-1,\ \frac{1}{n}\Sigma\right)$, any diagonal block of V^*, V^*_{kk} $(k = 1,\ldots,m)$

is $W_{p_k}\left(n-1,\ \frac{1}{n}\Sigma_{kk}\right)$ and is independent of any $V^*_{k'k'}$ for $k \neq k'$. Then from the previous Theorem 2 we have, for $k = 1, ..., m$,

$$\log|V^*_{kk}| \overset{a}{\sim} N\left(\log|\Sigma_{kk}| + a_k + \sum_{i=1}^{p_k}\sum_{j=1}^{2} b_j(n-i)^{-j}\ ,\ \ \sum_{i=1}^{p_k}\sum_{j=1}^{5} c_j(n-i)^{-j}\right)$$

with

$$a_k = \sum_{i=1}^{p_k} \log\frac{n-i}{n}\ ,$$

and b_j $(j = 1, 2)$ and c_j $(j = 1, \ldots, 5)$ defined as in Theorem 2.

Then since $\log|V^*_{kk}|$ is independent from $\log|V^*_{k'k'}|$ for any $k \neq k'$, we will have

$$\sum_{k=1}^{m}\log|V^*_{kk}| \overset{a}{\sim} N\left(\log|\Sigma_D| + \sum_{k=1}^{m} a_k + \sum_{k=1}^{m}\sum_{i=1}^{p_k}\sum_{j=1}^{2} b_j(n-i)^{-j}\ ,\ \sum_{k=1}^{m}\sum_{i=1}^{p_k}\sum_{j=1}^{5} c_j(n-i)^{-j}\right) \tag{5.11}$$

where $\Sigma_D = bdiag(\Sigma_{11}, ..., \Sigma_{mm})$. Then, from (5.5) and (5.11) we may see how the asymptotic expected value of

$$\log\Lambda^* = \log|V| - \sum_{k=1}^{m}\log|V^*_{kk}|$$

is equal to

$$\log|\Sigma| - \log|\Sigma_D| + a^* + b^* = \log|\Sigma\Sigma_D^{-1}| + a^* + b^*\ .$$

But, under $H_0 : \Sigma = \Sigma_D$, we have $\log|\Sigma\Sigma_D^{-1}| = 0$, and then the expected value of $\log\Lambda$ however different from zero tends to zero as n tends to infinity.

Then, given the independence under H_0 of $\log|V^*_{kk}|$ and $\log|V^*_{k'k'}|$ for $k \neq k'$, we have

$$\begin{aligned}
Var(\log\Lambda^*) &= Var\left(\log|V^*| - \textstyle\sum_{k=1}^{m}\log|V^*_{kk}|\right) = \\
&= Var\left(\log|V^*|\right) + \textstyle\sum_{k=1}^{m} Var\left(\log|V^*_{kk}|\right) - \\
&\qquad\qquad -2Cov\left(\log|V^*|,\ \textstyle\sum_{k=1}^{m}\log|V^*_{kk}|\right) \\
&= Var\left(\log|V^*|\right) + \textstyle\sum_{k=1}^{m} Var\left(\log|V^*_{kk}|\right) - \\
&\qquad\qquad -2\textstyle\sum_{k=1}^{m} Cov\left(\log|V^*|,\ \log|V^*_{kk}|\right)
\end{aligned}$$

where

$$Cov\left(\log|V^*|,\ \log|V^*_{kk}|\right) = Var\left(\log|V^*_{kk}|\right)$$

since if we consider a partition of V^* corresponding to the partition of Σ in (4.2), we may write

$$V^* = \left[\begin{array}{c|c} V^*_{11} & \tilde{V}^*_{12} \\ \hline \tilde{V}^*_{21} & \tilde{V}^*_{22} \end{array}\right] \tag{5.12}$$

with $\tilde{V}^*_{12} = [V^*_{12}|V^*_{13}|\ldots|V^*_{1k}|\ldots|V^*_{1m}] = \tilde{V}^{*\prime}_{21}$ and

$$\tilde{V}^*_{22} = \begin{bmatrix} V^*_{22} & V^*_{23} & \ldots & V^*_{2k} & \ldots & V^*_{2m} \\ V^*_{32} & V^*_{33} & \ldots & V^*_{3k} & \ldots & V^*_{3m} \\ \vdots & \vdots & & \vdots & & \vdots \\ V^*_{k2} & V^*_{k3} & \ldots & V^*_{kk} & \ldots & V^*_{km} \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ V^*_{m2} & V^*_{m3} & \ldots & V^*_{mk} & \ldots & V^*_{mm} \end{bmatrix}$$

and with $|V^*| = |V^*_{11}|\ |\tilde{V}^*_{22.1}|$, where $\tilde{V}^*_{22.1} = \tilde{V}^*_{22} - \tilde{V}^*_{21}V^{*-1}_{11}\tilde{V}^*_{12}$, and where V^*_{11} is independent of $\tilde{V}^*_{22.1}$. But then

$$\log|V^*| = \log|V^*_{11}| + \log|\tilde{V}^*_{22.1}| \tag{5.13}$$

and

$$Var\left(\log|V^*|\right) = Var\left(\log|V^*_{11}|\right) + Var\left(\log|\tilde{V}^*_{22.1}|\right)$$

since $\log|V^*_{11}|$ and $\log|\tilde{V}^*_{22.1}|$ are independent given the independence of V^*_{11} and $\tilde{V}^*_{22.1}$. But also from (5.13) we have

$$\log|V^*| - \log|V^*_{11}| = \log|\tilde{V}^*_{22.1}|\ .$$

This would give

$$Var\left(\log|\tilde{V}^*_{22.1}|\right) = Var\left(\log|V^*|\right) + Var\left(\log|V^*_{11}|\right) - 2\,Cov\left(\log|V^*|,\ \log|V^*_{11}|\right)$$

so that

$$Cov\left(\log|V^*|,\ \log|V^*_{11}|\right) = Var\left(\log|V^*_{11}|\right) \ .$$

Similarly

$$Cov\left(\log|V^*|,\ \log|\tilde{V}^*_{22}|\right) = Var\left(\log|\tilde{V}^*_{22}|\right) \ .$$

Then we may partition $\tilde{V}^*_{22}$ in a manner similar to (5.12), and by induction we can show that

$$Cov\left(\log|V^*|,\ \log|V^*_{kk}|\right) = Var\left(\log|V^*_{kk}|\right) \qquad (k = 1, ..., m) \ .$$

Finally

$$Var\left(\log\Lambda^*\right) = Var\left(\log|V^*|\right) - \sum_{k=1}^{m} Var\left(\log|V^*_{kk}|\right) \ ,$$

and using Theorem 2, we obtain

$$Var\left(\log\Lambda^*\right) = \sum_{j=0}^{5} c_j d_j \ ,$$

with c_j defined as in Theorem 2 and e_j defined as in (5.10). ■

We may obtain several asymptotic distributions of $\log\Lambda$, based upon Taylor series expansions of different orders.

Corollary 2: Using a Taylor series expansion of order 1,

$$\log\Lambda^* \ \overset{a}{\sim} \ N\left(a^* \ , \ \ 2e_1\right)$$

where a^* and e_1 are defined as in Theorem 3.

Corresponding to a Taylor series expansion of order 2,

$$\log\Lambda^* \ \overset{a}{\sim} \ N\left(a^* - e_1 \ , \ \ c^{*\prime}\right)$$

with a^* and e_1 defined as in Theorem 3 and

$$c^{*\prime} = \sum_{j=0}^{3} c_j' e_j$$

where e_j $(j = 1, ..., 5)$ are also defined as in Theorem 3. ■

Since V has the approximate Wishart distribution (5.9) (see also (4.6), Lemma 1 in section 4.4 of Chapter 4), we need to replace n^* with $\frac{(1-l)^2}{g}$ to obtain the approximate asymptotic distribution of $log\Lambda = log\frac{|V|}{\prod_{k=1}^{m}|V_{kk}|}$.

CHAPTER 6

THE MULTIPLE CORRESPONDENCE ANALYSIS

6.1 Introduction.

In this Chapter we will be studying the application of the GCA model to the study of the relationship among m categorical variables or attributes. Each categorical variable or attribute will be represented by the set of indicator variables for its categories (modalities). Therefore, each of the m sets of variables $\underline{x}_{(k)}$ $(k = 1, ..., m)$ will be a set of indicator variables for the categories (modalities) of a categorical variable.

For this situation the GCA model is equivalent to the Multiple Correspondence Analysis model. Multiple Correspondence Analysis is usually regarded as a method of assigning scores to the categories of the variables under analysis (Gower, 1990). We shall consider applying the models and test statistics developed in Chapter 4, for the GCA to the Multiple Correspondence Analysis (MCA) situation. This way we will give MCA the power to analyze the relationships among categorical variables that GCA has to study the relationships among groups of continuous variables. We shall then be able to see that Multiple Correspondence Analysis is also a modeling and hypotheses testing technique.

In section 6.4 we shall analyze the equivalence between the GCA algorithm applied to the m sets of indicator variables and two other analyses commonly performed on totally disjunctive matrices or matrices derived from them.

We will see how models in section 4.8 of Chapter 4, corresponding to the study of the relationship between one set of variables and all or some of the others, may be used to study the relationship between one categorical variable (attribute) and others, corresponding to the analysis of a band or part of a band of the Burt table $B = X'X$ (where X stands for the original non-centered data matrix).

Let us assume that we have a set of n observations for m attributes and let us assume that the k^{th} attribute, A_k, has p_k mutually exclusive and exhaustive categories, with $p = \sum_{k=1}^{m} p_k$.

For each attribute A_k $(k = 1, ..., m)$ we will define $\underline{x}_{(k)}$ as the set of p_k indicator variables corresponding to its p_k categories. If no individuals are observed or no observation falls into a particular category, we will remove this category from our consideration. Each submatrix X_k (defined in section 3.1 of Chapter 3) will be formed by columns by the indicator variables of the p_k categories of the k^{th} attribute ($x_{ki}^{j} = 1$ if observation i falls into the j^{th} $(j = 1, ..., p_k)$ category of the k^{th} attribute, $x_{ki}^{j} = 0$ otherwise).

First we shall consider the matrices X_k and thus also X without centering. For these noncentered matrices, it is then easy to see that the elements of each row of each submatrix X_k (since the categories of each A_k are considered disjoint and exhaustive) will add up to 1, and therefore each row of X will add up to m. The sum of the elements of each column of X_k will be equal to the absolute frequency of the corresponding category. Thus, the sum of all the elements in each X_k $(k = 1, ..., m)$ is n and the sum of all the elements in X is mn.

We shall continue using the same notation and considerations in Chapter 3. The Multiple Correspondence Analysis or Generalized Canonical Analysis on categorical variables gives identical results whether the data is centered or not. This will be further discussed ahead.

6.2 The observations space and the structure of the metric Q.

As in Chapter 3, we may obtain $m+1$ representations for each observation i: a 'global' representation and m 'partial' ones. In this analysis we will pay less attention to the 'partial' representations of the observations, since they really lie on the vertices of a simplex, as they are the representations of the observations through one of the categories of one of the (categorical) variables.

Since, in this analysis, it usually does not make much sense to assign different masses (or weights) to the observations, we shall assume,

$$m_i = \frac{1}{n} \quad \forall i \in \{1, \ldots, n\} \ .$$

If, however, the weights are unequal, our formulae will need only some minor changes and adjustments being then possible to use the general formulation presented in Chapter 3.

For a non-centered matrix X, each of our matrices $V_{kk} = X_k' D X_k$, (diagonal blocks of V) will be diagonal and so also the metric Q. Also, since

$$D = diag(m_i) = diag(1/n)$$

then,

$$V_{kk} = X_k' D X_k = \frac{1}{n} X_k' X_k = diag(n_{kj}/n) = \frac{1}{n} diag(n_{kj}) \tag{6.1}$$

where

$$n_{kj} = \sum_{i=1}^{n} x_{ki}^j = (\underline{x}_k^j)' \underline{x}_k^j \qquad (k = 1, ..., m) \quad (j = 1, ..., p_k) \tag{6.2}$$

is the absolute frequency of the j^{th} category (modality) of the k^{th} categorical variable, and where x_{ki}^{j} (which is either zero or one) is the value of the i^{th} observation for the j^{th} $(j = 1, \ldots, p_k)$ category of the k^{th} categorical variable. In (6.2) the vector $\underline{x}_k^j \in \mathbb{E}^n$ represents the j^{th} $(j = 1, \ldots, p_k)$ column of X_k, i.e.

$$\underline{x}_k^j = \begin{bmatrix} \\ x_{ki}^j \\ \\ \end{bmatrix} \qquad k = 1, ..., m; \quad i = 1, ..., n; \quad j = 1, ..., p_k \ .$$

The matrices V_{kk} will thus be diagonal matrices of the absolute frequencies of the categories (modalities) of the k^{th} $(k = 1, \ldots, m)$ categorical variable divided by

$$n = \sum_{j=1}^{p_k} n_{kj} \ .$$

The metric Q is, therefore,

$$Q = diag(n/n_{kj}) = n \ diag(1/n_{kj}) \qquad k = 1, \ldots, m; \ \ j = 1, \ldots, p_k \ . \tag{6.3}$$

The matrix Q is a diagonal matrix that gives gives to each category (modality) a weight equal to the inverse of its relative frequency. Therefore, each category (modality) gets an identical contribution to the total inertia of the cloud of observation-points (relate to Appendix 2.A), and thus, each categorical variable a contribution that is proportional to the number of categories in it.

We may also note that with the definition of this metric in $\mathbb{E}^p$, we are using in each subspace $\mathbb{E}^{p_k}$ its Mahalanobis metric which in this analysis are called chi-square metrics. These metrics differ from the Euclidean canonical metric only by the fact that each $(x_{ki}^j - x_{ki'}^j)^2$ is weighted by the inverse of the frequency of the j^{th} modality (of the k^{th} categorical variable) in the computation of the squared distance between two observations i and i'.

6.3 Norm and scalar product in $\mathbb{E}^p$ associated with the metric Q.

If Q_k denotes the restriction of Q to each subspace $\mathbb{E}^{p_k}$ of $\mathbb{E}^p$, as in Chapter 3, then, from (6.1) and (6.3),

$$Q_k = V_{kk}^{-1} = n\ diag(1/n_{kj}) \qquad j = 1, \ldots, p_k \quad \text{for a given } k \in \{1, ..., m\}\ .$$

We will then have

$$\| \underline{x}_{ki} \|^2_{Q_k} = n/n_{kj} \qquad \text{for a given } j \in \{1, \ldots, p_k\}\ ,$$

where j represents the category (modality) of the k^{th} categorical variable to which the observation i belongs.

Observe that,

$$\| \underline{x}_i \|^2_Q = \sum_{k=1}^{m} \| \underline{x}_{ki} \|^2_{Q_k} = \sum_{k=1}^{m} (n/n_{kj}) = n \sum_{k=1}^{m} (1/n_{kj})\ ,$$

where n_{kj} is the frequency of the category of the k^{th} attribute where the i^{th} observation falls (those categories are generically denoted by j here, however j is not the same for each k). This shows that observations that fall in 'rare' categories have larger norms, and are thus projected to the border of the cloud of observation-points. This is an interesting and useful property of this method if we are using a purely geometric-algebraic approach, because it enables us to identify observations corresponding to rare categories.

The scalar product of two observation-points (or observation-vectors) i and i' will be

$$\begin{aligned} \underline{x}_i' Q \underline{x}_{i'} &= \textstyle\sum_{k=1}^{m} \underline{x}_{ki}' Q_k \underline{x}_{ki'} = n \sum_{k=1}^{m} \sum_{j=1}^{p_k} (\underline{x}_{ki}' \underline{x}_{ki'})/n_{kj} \\ &= n \textstyle\sum_{k=1}^{m} (\delta/n_{kj}) \qquad \text{(for a given } j) \end{aligned} \tag{6.4}$$

where

$$\delta = \sum_{j=1}^{p_k} (\underline{x}'_{ki}\underline{x}_{ki'})$$

is 1 or 0 (zero) according as the two observations have or have not a common category (modality) of the k^{th} categorical variable. In (6.4), n_{kj} represents the frequency of the category j ($j \in \{1, \ldots, p_k\}$) of the k^{th} variable in which either of observations i or i' fall in. The scalar product of two observation-vectors in $I\!E^p$ will be, n times the sum of the reciprocals of the frequencies of the categories in which both fall at the same time. Then it is easy to see that two observation-vectors will be orthogonal if they do not fall simultaneously in any category of any categorical variable.

6.4 Factorial axes and subspace of $I\!E^p$ closest to the cloud of observation-points. Equivalence among several analyses.

In this section we will show the equivalence between the GCA applied to categorical variables and other analyses commonly carried out on completely disjunctive matrices or matrices derived therefrom.

6.4.1 The Multiple Correspondence Analysis as Generalized Canonical Analysis of the matrix X.

Once we see that the Multiple Correspondence Analysis is a special case of the Generalized Canonical Analysis, the factorial axes will be (as in Chapter 3, section 3.6) the directions $\Delta\underline{u}_\alpha$ in $I\!E^p$ that lie along the vectors $\underline{u}_\alpha$ ($\alpha = 1, \ldots q$) (eigenvectors of VQ associated with the eigenvalues λ_α, where λ_α are the inertias of the cloud of observation-points $N(I)$ in relation to the respective factorial axis of same order α).

The matrix $V = X'DX$, is a matrix of blocks

$$V_{kk'} = X'_k D X_{k'} \; .$$

Since we assumed

$$m_i \;=\; \frac{1}{n} \;\; \forall i \in \{1, \ldots m\} \;,$$

D becomes $diag(\frac{1}{n}) = \frac{1}{n} I_n$. Therefore, $V_{kk'}$ ($k \neq k'$; $k, k' \in \{1, \ldots, m\}$) are contingency tables (except that the entries are multiplied by the factor $1/n$) for the k^{th} and k'^{th} categorical variables. The entries are thus proportions, rather than actual frequencies. For $k = k'$, V_{kk}, is a diagonal matrix of the relative frequencies n_{kj}/n of the modalities of the k^{th} categorical variable.

Note that for any k and $k' \in \{1, \ldots, m\}$

$$V_{kk'} = \frac{1}{n} X_k' X_{k'} \;.$$

VQ will then be the matrix

$$VQ = \frac{1}{n} X' X diag\left(\frac{n}{n_{kj}}\right) = X' X diag\left(\frac{1}{n_{kj}}\right) = X' X D_C$$

where

$$D_C = diag\left(\frac{1}{n_{kj}}\right) = bdiag\left(X_k' X_k\right)^{-1} = n \; bdiag\left(V_{kk}^{-1}\right)$$

is the diagonal matrix of the reciprocals of the sums of the columns of X, i.e. the diagonal matrix of the reciprocals of the frequencies of all the $p = \sum_{k=1}^{m} p_k$ categories (modalities) of the m categorical variables.

Note that

$$VQ\underline{u}_\alpha = X' X D_C \underline{u}_\alpha = \lambda_\alpha \underline{u}_\alpha \;,$$

where the vectors $\underline{u}_\alpha$ will be vectors of D_C–norm equal to $1/\sqrt{n}$. Alternatively we may write

$$X' X D_C \underline{\omega}_\alpha = \lambda_\alpha \underline{\omega}_\alpha \;, \tag{6.5}$$

where now the vectors $\underline{\omega}_\alpha = \sqrt{n} \underline{u}_\alpha$ are D_C-unitary. The expression (6.5) corresponds really to a factorial analysis using the metric I_n in $\mathbb{E}^n$ and the metric D_C in $\mathbb{E}^p$. Clearly, the factorial axes $\Delta \underline{u}_\alpha$ and $\Delta \underline{\omega}_\alpha$ ($\alpha = 1, \ldots, q$) are collinear.

6.4.2 Equivalence of GCA with the Correspondence Analysis of the matrix X.

It is usually called Correspondence Analysis or Factorial Correspondence Analysis of a matrix to the Factorial Analysis that in the space spanned by the columns of such matrix uses the diagonal metric of the reciprocals of the sums of such columns and that has as dual analysis the one in the space spanned by the rows of the same matrix, in which the metric used is the one that has as matrix the diagonal matrix of the reciprocals of the sums of the rows of such matrix.

The metrics used are, in the observations space, $\mathbb{E}^p$,

$$D_C = diag(\text{reciprocal of the sum of columns of } X) = diag(1/n_{kj})$$

and in the variables space, $\mathbb{E}^n$, the metric

$$D_L = diag(\text{reciprocal of the sum of rows of } X) = diag(1/m) \ .$$

The problem will then be to determine the eigenvalues and D_C-unitary eigenvectors of

$$X'D_LXD_C = \frac{1}{m}X'XD_C \ .$$

Using (6.5),

$$\frac{1}{m}X'XD_C\underline{\omega}_\alpha = \frac{1}{m}\lambda_\alpha\underline{\omega}_\alpha \ , \tag{6.6}$$

or,

$$X'XD_C\underline{\omega}_\alpha = \lambda_\alpha\underline{\omega}_\alpha \ . \tag{6.7}$$

When we use (6.6) instead of (6.7) not only are we using D_L as a metric in $\mathbb{E}^n$ instead of I_n but also we will get λ_α/m instead of λ_α as the inertias of the observations cloud of points in relation to the factorial axis of order α.

The identity of formulas (6.7) and (6.5) proves that Generalized Canonical Analysis and Correspondence Analysis of the complete matrix X are completely equivalent. Both analyses give the same factorial axes.

6.4.3 Equivalence of GCA with the Correspondence Analysis of the Burt table.

Associated with a completely disjunctive data matrix X, we define the matrix

$$B = X'X = n\ V\ ,$$

a symmetric matrix of blocks $X_k'X_{k'}$. This is called the Burt matrix or Burt table (Burt, 1950).

The Burt table may be taken as our initial data matrix. In practice, it often happens that the only data matrix available, is the Burt table. This is usually the case when dealing with studies involving large numbers of observations.

The Burt table is a symmetric table. It is possible to start with it as the initial data table and analyze it as a contingency table (Burt, 1950; Cazès et allii, 1976; Pagès et allii, 1979). It is easy to see that in a Correspondence Analysis of the Burt table, the two clouds of points, direct and dual (corresponding to the row and column spaces of the Burt table) coincide, and our problem thus reduces to the analysis of a single cloud of p points, representing $p = \sum_{p_k}^{m}$ categories (modalities) of the m categorical variables, in a $(p-m)$-dimensional space ($p-m$ and not p since the j^{th} $(j = 1, ..., p_k)$ row in each block $B_{kk'}$ of B adds up to $n_{k'j}$, where $B_{k'k'} = diag(n_{k'j})$). Correspondingly, also the metric to use in both spaces and both analysis, i.e. direct and dual, will be the same.

The metric matrix, corresponding to the definition of Correspondence Analysis, will be the diagonal matrix with elements as the reciprocals of the sums of the columns (or sums of rows) of B. These column sums are m times the column sums of X, and so the metric used will be

$$D_B = diag(1/mn_{kj}) = \frac{1}{m} diag(1/n_{kj}) = \frac{1}{m} D_C\ .$$

The factorial axes will then be the directions in $\mathbb{E}^p$ along the D_B-unitary eigenvectors of

$$B'D_BBD_B = \frac{1}{m^2}X'XD_CX'XD_C \ .$$

These will be the vectors $\underline{\eta}_\alpha \in \mathbb{E}^p$, such that

$$\frac{1}{m^2}X'XD_CX'XD_C\underline{\eta}_\alpha = \mu_\alpha\underline{\eta}_\alpha \ .$$

From the definition of the vectors $\underline{\omega}_\alpha$ given by (6.7), we have $\underline{\eta}_\alpha = \sqrt{m}\,\underline{\omega}_\alpha$, and, using again (6.7),

$$\frac{1}{m^2}X'XD_CX'XD_C\,\underline{\omega}_\alpha = \frac{1}{m^2}X'XD_C\lambda_\alpha\,\underline{\omega}_\alpha = \frac{1}{m^2}\lambda_\alpha^2\,\underline{\omega}_\alpha \ . \tag{6.8}$$

Therefore,

$$\mu_\alpha = \lambda_\alpha^2/m^2 \ .$$

Multiplying both sides of (6.8) by m^2 we get,

$$BD_CBD_C\,\underline{\omega}_\alpha = \lambda_\alpha^2\,\underline{\omega}_\alpha \ .$$

This corresponds to carry out a Correspondence Analysis of B using the metric D_C instead of $D_B = \frac{1}{m}D_C$, in both spaces.

The Correspondence Analysis of the Burt table is equivalent to the Factorial Analysis of a cloud of points

$$N_B(J) = \{(\underline{b}_{kj}, 1/(mn_{kj})),\ j = 1, \ldots, p_k;\ k = 1, \ldots, m\}$$

where n_{kj} represents the frequency of the j^{th} category (modality) of the k^{th} categorical variable, showing that to each row or column of B, denoted by $\underline{b}_{kj}$ we are assigning a weight which is the reciprocal of m times the frequency of the corresponding category, if we use the metric $D_B = \frac{1}{m}D_C$, or the analysis of an equivalent cloud of points with weights $1/n_{kj}$ if we use the metric D_C. These two Factorial or Correspondence Analyses will give exactly the same factorial axes but while the use of the metric D_B will give μ_α as the inertia of the cloud of indicator variable points in relation to each direction $\Delta\underline{\eta}_\alpha \equiv \Delta\underline{\omega}_\alpha$, the analysis using D_C will give λ_α^2 as the inertias.

From (6.8) and (6.5), we also conclude that the Factorial Correspondence Analysis of B, using the metric D_C, and the Generalized Canonical Analysis of the total matrix X give the same factorial axes. However, while the inertia of the clouds of points (both the cloud of observation-points and the cloud of indicator-variable points) considered in the Generalized Canonical Analysis of X is λ_α (relative to each factorial axis of order α $(\alpha = 1, \ldots, q)$), in the Correspondence Analysis of the Burt table, the cloud $N_B(J)$ has λ_α^2 as its inertia relative to the same directions.

The major difference between the analysis of the Burt table and the analysis of the matrix X is that while the last one allows us to obtain both the representation of the observations and of the indicator variables on the new basis formed by the first q factorial axes, the former one (i.e. the analysis of the Burt table) in principle only allows the representation of the indicator variables. But this is not necessarily a disadvantage of the first analysis because quite often we are really not much interested in the representation of the individual observations which are usually very large in number, but rather only in the representation of the categories (or of the indicator variables for the categories) of the m categorical variables. In the case of numerous observations the Burt table also becomes more easily manageable than the matrix X.

Methods are available that allow for an analysis of the Burt table by bands corresponding to the crossing of the categories (or rather of the indicator variables for the categories) of one categorical variable with the indicator variables for the modalities of all or some of the other categorical variables (Benzecri, 1982). We will show later how the models developed in 4.8 and the test statistics developed in section 4.7 could be also used to this purpose.

6.5 The three equivalent ways to implement the GCA algorithm.

We have already shown that the GCA applied to the total matrix X is equivalent to the other methods of analysis of data matrices for qualitative variables that are commonly used. We shall, therefore, focus only on the GCA hereafter.

When all our m sets of variables are formed by indicator variables for the categories (modalities) of m categorical variables there are three variations of carrying out the GCA on our data. These are:

(i) to use all the data, i.e., all the $p = \sum_{k=1}^{m} p_k$ indicator variables represented in X, without centering;

(ii) to use all but one of the p_k indicator variables in each set $\underline{x}_k$, using only a total of $p - m$ of the p indicator variables in X, without centering;

(iii) to use the same approach as in ii) but with centering.

All the three approaches give the same results, with the only difference that the approach in i) gives the largest eigenvalue $\lambda_1 = m$ and $m - 1$ smallest eigenvalues equal to zero. This is so because each row of the data matrix adds to m, and as such all these m eigenvalues are really spurious ones. The eigenvalue λ_1 has an associated eigenvector $\underline{u}_1$, proportional to $Q^{-1}E_{p1} = [n_{kj}/n]$ (with proportionality constant $1/\sqrt{m}$). Therefore, λ_1 has an associated factor $\underline{a}^1 = Q\underline{u}_1$ (see definition in section 2.7 of Chapter 2) (eigenvector of QV or QB associated with λ_1) proportional to E_{p1}. This factor gives equal scores to all categories and is thus of no use. Therefore, such eigenvector and eigenvalue as well as all the other $m - 1$ null eigenvalues and associated eigenvectors will be discarded.

But since then all the other eigenvectors $\underline{u}_\alpha$ of VQ associated with the nonnull eigenvalues are Q-orthogonal to $Q^{-1}E_{p1}$ they are centered in the sense that for $\alpha > 1$, $\underline{u}'_\alpha QQ^{-1}E_{p1} = \underline{u}'_\alpha E_{p1} = 0$. This means that then all the factors $\underline{a}^\alpha$ for $\alpha > 1$ are

Q^{-1} orthogonal to E_{p1} and the generalized canonical variables $\underline{c}^{\alpha}$ are centered, in the sense that they will be D-orthogonal to $\underline{c}^{1} = XQ\underline{u}_{1} = \sqrt{\frac{m}{n}}E_{p1}$.

Once we see this, it is algebraically involved but conceptually easy to see the equivalence among the results of the three approaches (i), (ii) and (iii) and so the details will be omitted.

We should however note that there are situations where approach (ii) may be not an adequate one. These situations are quite rare but they may arise mainly when the indicator variables instead of representing the categories of attributes observed they represent rather the levels of experimentally controlled variables or factors in the some experimental design type of situation and in situations when the levels of one factor are nested in the levels of another or somehow conditioned by the levels of another factor as for example in situations of the type of repeated measures analyses. Then the removal of one indicator variable from each set without centering may alter the structure of the relationships among two or more of such sets. A situation where this occurs will be the one of an example treated in section 6.8.

As we will see ahead, we may use in MCA the test statistics proposed in section 4.7 of Chapter 4. We should then note that when using approach (i) we should then remove from our consideration all the spurious and structural eigenvalues from our consideration before computing such test statistics.

We should also note that when we use approaches (ii) or (iii) we will not obtain the coordinates in the vectors $\underline{u}_{\alpha}$ for the categories corresponding to the modalities removed. These coordinates are easy to be obtained since we know that, from the above referred relation, $\underline{u}_{\alpha}' E_{p1}$ (for $\alpha > 1$), that the coordinates for the complete vectors $\underline{u}_{\alpha}$ (for $\alpha > 1$) should add up to zero. If we do not consider this coordinate then our scores in the vectors $\underline{a}^{\alpha}$ will be differences between the actual scores for the categories and what would be the score for the removed category, this is the score for this removed category is taken as being zero.

6.6 The matrix B as an MLE.

It should be noted that centering or not centering the indicator variables leads to different interpretations for the blocks of $B = X'DX$ (the Burt table or Burt matrix). Strictly speaking, it is $B = X'X$ that is called Burt table and this only if the columns of the matrix X are not centered for the masses m_i. In this case the diagonal blocks of B are themselves diagonal matrices of the frequencies of the p_k classes (categories) of each of the m categorical variables. The off-diagonal blocks are then the contingency tables corresponding to the cross-classification of the n observations generated by the categories (classes) of two categorical variables. For example B_{kl} will be the contingency table corresponding to the cross-classification generated by the categories (classes) of the k^{th} and l^{th} categorical variables. When $B = V = X'DX$ rather than $B = X'X$ (as we shall consider hereon) the diagonal blocks of B are diagonal matrices of the relative frequencies of the p_k classes of each of the m categorical variables and the off-diagonal blocks are two-way tables of sample proportions.

But, if the matrix X has been centered for the masses m_i, then the diagonal blocks of $B = X'DX$ are not diagonal any more.

If we assume a multinomial distribution for the indicator variables in each set $\underline{x}_{(k)}$ $(k = 1, ..., m)$, with

$$f(\underline{x}_{(k)}) \;=\; \frac{1!}{\prod_{j=1}^{p_k} x_{(k)j}!} \prod_{k=1}^{p_k} {\pi_j}^{x_{(k)j}} \;\equiv\; \prod_{j=1}^{p_k} {\pi_j}^{x_{(k)j}} \quad (j = 1, ..., p_k)$$

where π_j is the probability that observation i falls in the j^{th} category of the k^{th} attribute, with $j = 1, ..., p_k$ and with $\underline{x}_{(k)} = [x_{(k)j}]$, where $x_{(k)j}$ is either zero or one, then B is now a sample variance-covariance matrix with elements

$$b_{ij} \;=\; n^*_{ij} - n\hat{\pi}_i\hat{\pi}_j \qquad i, j = 1, ..., p\ , \tag{6.9}$$

where n is the total sample size, $\hat{\pi}_i = n_i/n$ and $\hat{\pi}_j = n_j/n$, the MLE's of π_i and π_j are the relative frequencies of categories i and j and n^*_{ij} is the number of observations that fall at the same time in categories i and j $(i, j = 1, ..., p)$. Then the diagonal blocks of B, the matrices $B_{kk} = X'_k D X_k$ will be the sample variance-covariance matrices of m multinomial distributions with parameters $\pi_{k1}, ..., \pi_{kp_k}$ $(k = 1, ..., m)$. We may note that for the diagonal elements of these m diagonal blocks of B, $n^*_{ii} = n_{ki}$ $(i = 1, ..., p_k)$ and $\hat{\pi}_{kj} = \hat{\pi}_{ki}$ and then

$$b_{k_i k_i} \;=\; n_{ki} - n\pi^2_{ki} \;=\; n\pi_{ki} - n\pi^2_{ki} \;=\; n\pi_{ki}(1 - \pi_{ki}) \quad (i = 1, ..., p_k;\; k = 1, ..., m)\ .$$

These are the MLE's of the variances of variables in a multinomial distribution with parameters $\pi_{1k}, ..., \pi_{1p_k}$ $(k = 1, ..., m)$. The off-diagonal elements of B_{kk} are

$$b_{k_i k_j} \;=\; n\hat{\pi}_i \hat{\pi}_j$$

as $n^*_{ij} = 0$. These are the MLE's of the covariance between two multinomially distributed variables $x_{(k)i}$ and $x_{(k)j}$. The elements of the off-diagonal blocks of B, having the general form in (6.9), will then be (by the invariance property of the MLE's) the MLE's of the covariance between two random variables in correlated multinomial distributions. This way B may be seen as the MLE of the variance-covariance matrix of p variables in m correlated multinomial distributions with parameters $\pi_{k1}, \pi_{k2}, ..., \pi_{kp_k}$ $(k = 1, ..., m)$.

But, then, under our general null hypothesis (4.3) of independence (now hypothesis of independence of the m multinomial distributions) and by the invariance property of the MLE's, the MLE of the variance covariance matrix of the m multinomials is

$$Q^{-1} \;=\; bdiag(B_{kk}) \qquad (k = 1, ..., m)\ .$$

We may then argue that

$$\Lambda = |VQ| = |BQ| = \frac{|B|}{\prod_{k=1}^{m} |B_{kk}|}$$

is a measure of the 'closeness' of B to $bdiag(B_{kk})$ and as both of these matrices are formed by MLE's we may still look upon the test statistic $\Lambda = |BQ|$ as a Likelihood Ratio test statistic.

Another peculiarity about the matrix B is that when it originates from the non-centered data matrix, then the sum of each row or column of each block in B will be n_j $(j = 1, ..., p)$, and there are m of such blocks in each row and column of B. This is why when we consider all the non-centered indicator variables we obtain the largest eigenvalue equal to m with the associated eigenvector proportional to $Q^{-1}E_{p1}$ (Note that $Q = diag(n/n_j)$). But when we use a matrix B obtained from a centered X matrix, then , B having elements $b_{ij} = N^*_{ij} - n\hat{\pi}_i\hat{\pi}_j$, each row and column of each block of B adds up to zero. In this analysis we would therefore have a first eigenvalue of zero associated with an eigenvector proportional to $Q^{-1}E_{p1}$ and that is why we follow the approach in (iii) similar to the one in (ii) of removing one row and the corresponding column from each block of B, corresponding to use only $p_k - 1$ indicator variables for each qualitative variable.

6.7 The use in MCA of the test statistics developed in Chapter 4, section 4.7.

The test statistics in Chapter 4, section 4.7, have been developed for normal variables, i.e., under the assumption that each set of variables $\underline{x}_{(k)}$ $(k = 1, ..., m)$ had a normal p_k-variate distribution.

We noted in the previous sections (6.4 and 6.6) that the MCA could be seen as a particular case of GCA and also, that in MCA, $\Lambda = |VQ| = |BQ|$ can still bee seen as a likelihood ratio test statistic. These reasons, however enough to justify the use in MCA of the test statistics developed in 4.7, do not allow us to use for them the distributions derived in 4.7, since our variables now are not normal but multinomial.

The argument that in MCA, $\Lambda = |VQ| = |BQ|$ may still be seen as a likelihood ratio test statistic is only useful in order to obtain for our generalized Wilks' lambda statistic a large sample distribution, in the sense that for a large sample size, i.e. for large n,

$$-2\ log\Lambda \overset{a}{\sim} \chi^2_f \qquad \text{with} \qquad f = (p-m)^2 - \sum_{k=1}^{m}(p_k - 1)^2\ .$$

Both types of test statistics developed in 4.7 of Chapter 4, $\Lambda_{i(i+1,\ldots,m)}$ and $\Lambda_{ik|(i+1,\ldots,k-1,k+1,\ldots,m)}$ may be seen as referring to the canonical analysis of two sets of variables. This is clear for $\Lambda_{i(i+1,\ldots,m)}$, from formula (4.12) and from the ANOVA table in section 4.9 of Chapter 4. $\Lambda_{i(i+1,\ldots,m)}$ may easily be seen as the Wilks' lambda corresponding to the canonical analysis of set $\underline{x}_{(i)}$ with the set $\left(\underline{x}_{(i+1)}, \ldots, \underline{x}_{(m)}\right)$. However for $\Lambda_{ik|(i+1,\ldots,k-1,k+1,\ldots,m)}$ this interpretation may at first sight be not so clear. However, from formula (4.13) and the ANOVA table in section 4.9 of Chapter 4, it is still possible to write it as $|A|/|A+B|$, where B is the hypothesis matrix and A the error matrix (see sections 4.8 and 4.9 of Chapter 4). We will hereon refer to the hypothesis matrix in either (4.12) or (4.13) as B and to the error matrix as A.

First consider the analysis of only two sets of indicator variables, say $\underline{x}_{(i)}$ and $\underline{x}_{(i+1)}$, respectively formed by p_i and p_{i+1} indicator variables, corresponding to the two attributes or categorical variables A_i and A_{i+1}. Then the usual X^2 test statistic used to test the null hypothesis of independence between the two attributes, or rather between the two sets of indicator variables, is known to have (under such null hypothesis) the chi-square distribution $\chi^2_{(p_i-1)(p_{i+1}-1)}$. Whichever among the three approaches (i), (ii) or (iii) in section 6.5 is used, it has been shown that X^2 is related to the Pillai's trace (Kendall and Stuart, 1977) through

$$X^2_{i(i+1)} \ = \ \mathcal{V} \ = \ n\ tr\left(B(B+A)^{-1}\right)$$

where A is the error matrix and B is the hypothesis matrix for the test of the hypothesis of independence between $\underline{x}_{(i)}$ and $\underline{x}_{(i+1)}$, and where it is assumed that if we took approach (i), any spurious or trivial roots, arising from the fact that all the indicator variables are used have been removed from $\mathcal{V}$.

But then, under the null hypothesis of independence of $\underline{x}_{(i)}$ and $\underline{x}_{(i+1)}$, the Wilks' lambda $|A|/|A+B|$ is known to be asymptotically equivalent to X^2 or the Pillai's trace criterion $\mathcal{V}$, with

$$log\Lambda_{i(i+1)} \;\simeq\; \mathcal{V} \;.$$

The approximation follows from the expansion of $log(1+r)$ (see Kshirsagar, 1972). So that we may consider that

$$-n\; log\Lambda_{i(i+1)} \;\stackrel{a}{\sim}\; \chi^2_{(p_i-1)(p_{i+1}-1)} \;.$$

But then, under the null hypothesis (4.3) of independence among all the m sets of variables, $-n\; log\Lambda_{i(i+1,...,m)}$ is approximately the sum of $m-i$ independent chi-square distributed quantities, each with $(p_i-1)(p_k-1)$ $(k=i+1,...,m)$ degrees of freedom, i.e.

$$-n\; log\Lambda_{i(i+1,...,m)} \;\simeq\; n \sum_{k=i+1}^{m} X^2_{i(k)}$$

where $X^2_{i(k)}$ $(k=i+1,...,m)$ represents $m-i$ independent chi-square distributed variables, each with $(p_i-1)(p_k-1)$ degrees of freedom. Each $\chi^2_{i(k)}$ is the test statistic for the test of independence between $\underline{x}_{(i)}$ and $\underline{x}_{(k)}$, or $\sum_{k=i+1}^{m} X^2_{i(k)}$ $(i = 1,...,m-1)$ is the test statistic for the test of independence corresponding to the hypotheses $H_0^{(i)}$ in 4.8, i.e. the test statistic to test the independence between $\underline{x}_{(i)}$ and $\left(\underline{x}_{(i+1)},...,\underline{x}_{(m)}\right)$, i.e. the chi-square test statistic for the sequence of $m-i$ contingency tables corresponding to the crossing of the modalities for the i^{th} qualitative variable with the modalities of the $m-i$ qualitative variables A_{i+1} through A_m.

So we may then say that

$$-\,n\; log\Lambda_{i(i+1,...,m)} \;\stackrel{a}{\sim}\; \chi^2_{(p_i-1)(p_{i+1}+...+p_m-(m-i))} \;, \tag{6.10}$$

corresponding to the use of $p_k - 1$ out of the p_k indicator variables in each of the sets $\underline{x}_{(i)}$ through $\underline{x}_{(m)}$.

Concerning the test statistic $\Lambda_{ik|(i+1,\ldots,k-1,k+1,\ldots,m)}$ we may remember that in section 4.7 of Chapter 4 it was defined as

$$\Lambda_{ik|(i+1,\ldots,k-1,k+1,\ldots,m)} = \frac{\Lambda_{i(i+1,\ldots,m)}}{\Lambda_{i(i+1,\ldots,k-1,k+1,\ldots,m)}} ,$$

or

$$-n\ log\Lambda_{ik|(i+1,\ldots,k-1,k+1,\ldots,m)} = -n\ log\Lambda_{i(i+1,\ldots,m)} + n\ log\Lambda_{i(i+1,\ldots,k-1,k+1,\ldots,m)}$$

where, from (6.10) above,

$$-n\ log\Lambda_{i(i+1,\ldots,m)} \overset{a}{\sim} \chi^2_{(p_i-1)(p_{i+1}+\ldots+p_m-(m-i))}$$

and from similar arguments

$$-n\ log\Lambda_{i(i+1,\ldots,k-1,k+1,\ldots,m)} \overset{a}{\sim} \chi^2_{(p_i-1)(p_{i+1}+\ldots+p_{k-1}+p_{k+1}+\ldots+p_m-(m-i)+1)} .$$

However, even under the general null hypothesis of independence among all the sets of variables $\Lambda_{i(i+1,\ldots,m)}$ and $\Lambda_{i(i+1,\ldots,k-1,k+1,\ldots,m)}$ are not independent. Nevertheless, if we use a further approximation, we may then take the approximate distribution of $\Lambda_{ik|(i+1,\ldots,k-1,k+1,\ldots,m)}$ as

$$-\ n\ log\Lambda_{ik|(i+1,\ldots,k-1,k+1,\ldots,m)} \overset{a}{\sim} \chi^2_{(p_1-1)(p_k-1)} . \tag{6.11}$$

It may be possible to improve upon these approximations by replacing the multiplier n by a Bartlett type correction factor. However this was not considered in this dissertation.

Another reason that we may use in order to still further justify the use of the test statistics $\Lambda_{i(i+1,\ldots,m)}$ and $\Lambda_{ik|(i+1,\ldots,k-1,k+1,\ldots,m)}$ is that as seen above they really are Wilks' lambda statistics in the canonical analysis of two sets of variables, $\underline{x}_{(i)}$ and $\left(\underline{x}_{(i+1)},\ldots,\underline{x}_{(m)}\right)$. Then it is true that even if we assume a large sample size it would still be hard to assume limiting normal distributions for the sets of indicator variables $\underline{x}_{(k)}$ $(k = i, ..., m)$, but as Bartlett (1951) says: "Even although any precise tests of significance are based on the assumption of normality of at least one set of variables, an assumption which cannot strictly be true in this example, the structural analysis is independent of this assumption, and the tests, if correctly formulated, will still be asymptotically correct, and therefore still informative.", a statement that he reiterated twelve years later when he said that "provided tests of significance are not taken too precisely, it is believed that they are broadly correct" (Bartlett, 1963). We agree with this perspective of the problem by Bartlett. Further, when we remarked above that the fact that the MCA may be seen as a particular case of the GCA may be used as a justification to use in MCA the test statistics used in GCA, our thought was in accordance with Bartlett's (1951) when he said that "...the structural analysis is independent of this assumption...", referring to the assumption of normality of at least one set of variables when carrying the canonical analysis of two sets of variables.

Really when using the matrix X centered for the masses m_i, and when using $p_k - 1$ indicator variables for each qualitative variable then for a large sample size it seems that the distribution of $B = X'DX$ really approaches a Wishart distribution but further study in this area would surely be adequate.

These reasons may also be used as a justification to use the above two test statistics and still assume for them approximate Wilks' lambda distributions similar to the ones derived in section 4.7 of Chapter 4, with each p_k replaced by $p_k - 1$ for $k = 1, ..., m$.

We should also note here that as already called the attention to in section 6.5, if approach (i) there described is used, i.e. if all the indicator variables are used, then we should then remove from consideration all the spurious and structural eigenvalues from our consideration before computing the above test statistics.

6.8 Analysis of Taylor's blood serological data set and the need for other models.

Generalized Canonical Analysis models, when dealing with categorical variables, i.e., Multiple Correspondence Analysis and related models are often used with the main purpose of assigning 'optimal' scores to the categories of all or some of the categorical variables involved in the analysis.

Such scores are the generalized canonical factors of order α, $\underline{a}^{\alpha} = Q\underline{u}_{\alpha} \in I\!E^{p*}$, introduced under the general approach of Factorial Analysis in section 2.7 of Chapter 2. They are optimal in the sense that the vectors $\underline{u}_{\alpha}$ and the factorial axes $\Delta\underline{u}_{\alpha}$ have the optimal properties defined in section 2.6 of Chapter 2, for the general Factorial Analysis and also in the sense that the vectors $\underline{c}^{\alpha} = XQ\underline{u}_{\alpha}$ and $\underline{c}_k^{\alpha} = X_k Q_k \underline{u}_{k\alpha}$ have the optimal properties shown in sections 3.7 and 3.8 of Chapter 3. Frequently, one retains the scores corresponding to $\underline{a}^1$ (or $\underline{a}^2$ if $\underline{a}^1$ is a trivial factor as it happens in MCA in some situations, as seen earlier). One should really retain as many sets of scores as the number of 'significant' eigenvalues λ_{α} of VQ. But there are no tests of significance available for individual eigenvalues λ_{α} when $m > 2$. For $m = 2$, however, references may be made to Hsu (1941), Constantine (1963), Kshirsagar (1972), Glynn and Muirhead (1978) and Glynn (1980).

There are mainly two reasons for the unique focus on the use of MCA to obtain scores. For Multiple Correspondence Analysis there are generally no test statistics available and as such the method is used as a descriptive one, with the purpose of obtaining scores. The second reason is that however it is becoming more and more accepted that, when one uses the MCA, there surely is an underlying model being fitted to our data (Gower, 1990). An explicit model, with corresponding hypotheses and associated test statistics clearly stated, has never been presented.

As a matter of fact the objective of obtaining scores alone is inadequate and should be supplemented by the model to be used and the hypotheses to be tested.

We are going to show that the findings of Gower (1990) are not surprising and that Fisher's (1938) method is more closely related to the Canonical Analysis than to MCA. The only surprising thing is that Gower (1990) was able to obtain the same scores as Fisher (1938) by using a technique that, however being more sophisticated, may be really not the most adequate to analyze these data. The two results agree because of the peculiar structure of the data and this fact is overlooked by Gower (1990). We shall also show that Fisher's (1938) approach is really the most appropriate one of the two.

An interesting question is how many sets of scores to retain. From our earlier exposition of the GCA model and associated test statistics sections 4.7 and 4.8 of Chapter 4, one way to choose the adequate number of sets of scores one should retain in the model could be to choose that number to be the maximum number of columns used in any of the $\beta_{i(i+1,\ldots,m)}$ parameter matrices in the MCA model (in model (4.18)). However, we never established a correspondence between the $\beta_{i(i+1,\ldots,m)}$ matrices in the submodels in (4.18) and the 'overall' parameter matrix QU from the overall GCA (or MCA) model, for the case $m > 2$ (the relation was established in section 3.11 of Chapter 3, for the case $m = 2$). In other words, we did not establish a correspondence between the canonical correlations in each matrix $\Psi_{i(i+1,\ldots,m)}$ and the generalized canonical correlations $\lambda_1, \ldots, \lambda_p$. This criterion of choosing the number of sets of scores needs further discussion and research.

We may break down each vector $\underline{a}^\alpha = Q\underline{u}_\alpha$ in the subvectors corresponding to the vectors $\underline{u}_{k\alpha}$ $(k = 1, ..., m)$ defined in section 3.7 of Chapter 3. We would then have

$$\begin{bmatrix} \underline{a}_1^\alpha \\ \vdots \\ \underline{a}_m^\alpha \end{bmatrix} = Q \begin{bmatrix} \underline{u}_{1\alpha} \\ \vdots \\ \underline{u}_{m\alpha} \end{bmatrix}$$

where $\underline{a}_k^\alpha = Q\underline{u}_{k\alpha}$ $(k = 1, ..., m)$ is the α^{th} set of scores for the k^{th} attribute (categorical variable).

To illustrate the MCA technique, model building process and testing procedure we will use Taylor's serological data set, a well known data set that has been extensively referred in the literature (Fisher, 1938; Bartlett, 1951; Kshirsagar, 1972; Gower, 1990). The reasons that lead us to choose this data set could not be better expressed than by Bartlett (1951). He says: "As emphasized by Fisher, these non-numerical data would hardly justify an elaborate analysis, nevertheless, they constitute a very convenient and interesting example for purposes of illustrating possible tests" (and also modeling approaches).

Another reason why this data set is used is a recent paper by Gower (1990) in which he compares the scores obtained for the response variable using the MCA to the scores obtained by Fisher (1938).

In Taylor's data set we have $m = 3$ discrete characteristics or attributes under study: responses, sera and subjects or individuals. The first with 5 levels: no reaction, a trace, weak, positive and strongly positive; and the last two with 12 levels each, corresponding respectively to the 12 sera and the 12 individuals.

For the sake of making the exposition clearer, we will use the following notation:

$\underline{x}_{(1)}$ – the set of indicator variables for 'response'

$\underline{x}_{(2)}$ – the set of indicator variables for 'sera'

$\underline{x}_{(3)}$ – the set of indicator variables for 'individuals' .

For this example we do not have the X_k matrices available but that is of no importance since we do not need the representation of the individual observations and the non-availability of the original data matrices does not at all prevent us from carrying any part of our analysis neither the test of hypotheses nor even from writing the model. The data is presented in the following Table, reproduction of Table 61.9 from Fisher (1938).

Table 6.1 – Table 61.9 from Fisher (1938) – Taylor's serological data

	Serum											
Cell	1	2	3	4	5	6	7	8	9	10	11	12
1	w	w	w	(+)	w	(+)	?	w	w	(+)	w	w
2	?	w	?	w	w	w	?	w	w	w	w	?
3	w	w	w	w	w	w	w	w	w	w	w	w
4	w	w	w	w	w	w	–	w	w	w	w	?
5	w	(+)	w	(+)	w	w	?	(+)	w	(+)	w	w
6	w	w	(+)	(+)	w	w	?	w	w	(+)	w	w
7	(+)	(+)	(+)	(+)	+	+	w	(+)	w	(+)	(+)	w
8	w	+	(+)	(+)	w	(+)	w	(+)	w	(+)	(+)	w
9	w	(+)	(+)	(+)	w	(+)	w	(+)	w	(+)	w	w
10	?	?	w	w	w	w	?	w	w	w	w	?
11	w	w	(+)	w	w	w	?	w	w	w	w	w
12	w	(+)	+	(+)	(+)	(+)	w	(+)	w	(+)	+	w

The symbols in the above table have been transformed into numbers, –, ?, w, (+) and + to 1, 2, 3, 4, and 5 respectively. Then by counting we obtain the Burt matrix described in Gower (1990)

$$B = \begin{bmatrix} W & R' & C' \\ R & 12I_{12} & E_{12,12} \\ C & E_{12,12} & 12I_{12} \end{bmatrix} \tag{6.12}$$

where

$$R' = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 4 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 4 & 1 & 0 \\ 8 & 8 & 12 & 10 & 7 & 8 & 3 & 5 & 6 & 8 & 10 & 4 \\ 3 & 0 & 0 & 0 & 4 & 3 & 7 & 6 & 6 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & 2 \end{bmatrix}, \tag{6.13}$$

$$C' = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 1 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & 0 & 3 \\ 9 & 6 & 5 & 5 & 10 & 7 & 5 & 7 & 12 & 5 & 9 & 9 \\ 1 & 4 & 5 & 7 & 1 & 4 & 0 & 5 & 0 & 7 & 2 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \tag{6.14}$$

and, as described by Gower (1990)

$$W = diag(1,\ 13,\ 89,\ 36,\ 5)\ . \tag{6.15}$$

Once we choose the MCA model to analyze these data it would mean that our objective is to test the independence among all the three sets of indicator variables: for response, for sera and for individuals. However, the aim of the analysis is really to test if 'response' is related with 'sera', conditionally on 'individuals'. Even if innapropriate, we shall first illustrate the use of the MCA model, for this data.

For the general testing procedure outlined in section 4.8 of Chapter 4, we choose the following hypotheses for testing:

$$H_0^{(1)} \; : \; \Sigma_{1(23)} = 0$$

and

$$H_0^{(2)} \; : \; \Sigma_{2(3)} = 0$$

where

$$\hat{\Sigma}_{1(23)} \;=\; [R' \; C']$$

and

$$\hat{\Sigma}_{2(3)} \;=\; E_{12,12} \; .$$

The choice of $H_0^{(1)}$ and $H_0^{(2)}$ is governed by the fact that the set $\underline{x}_{(1)}$ can be regarded as a response set.

Also if $H_0^{(1)}$ is rejected, we then would like to test

$H_0^{2(1)}$: the partial canonical correlations between sets $\underline{x}_{(1)}$ and $\underline{x}_{(2)}$, given set $\underline{x}_{(3)}$, are all null ,

and

$H_0^{3(1)}$: the partial canonical correlations between sets $\underline{x}_{(1)}$ and $\underline{x}_{(3)}$, given set $\underline{x}_{(2)}$, are all null .

For this data set we had $n = 144$, so that we take

$$D \;=\; diag(1/n) \;=\; diag(1/144)$$

and thus, from (6.12),

$$V = \frac{1}{144}B = \begin{bmatrix} \frac{1}{144}W & \frac{1}{144}R' & \frac{1}{144}C' \\ \frac{1}{144}R & \frac{1}{12}I_{12} & \frac{1}{144}E_{12,12} \\ \frac{1}{144}C & \frac{1}{144}E_{12,12} & \frac{1}{12}I_{12} \end{bmatrix} . \tag{6.16}$$

From sections 4.7 and 4.8 of Chapter 4, the test statistics to test $H_0^{(1)}$ and $H_0^{(2)}$ above are respectively

$$\begin{aligned} \Lambda_{1(23)} &= \frac{|V_{11} - V_{1(23)}V_{(23)(23)}^{-1}V_{(23)1}|}{|V_{11}|} \\ &= \left| W - [R' \ C'] \begin{bmatrix} \frac{1}{12}I_{12} & E_{12,12} \\ E_{12,12} & \frac{1}{12}I_{12} \end{bmatrix}^{-1} \begin{bmatrix} R \\ C \end{bmatrix} \right| \Big/ |W| \end{aligned}$$

and

$$\begin{aligned} \Lambda_{2(3)} &= \frac{\left|12I_{12} - E_{12,12}\frac{1}{12}I_{12}E_{12,12}\right|}{|12I_{12}|} \\ &= \frac{|12I_{12} - E_{12,12}|}{12^{12}} . \end{aligned}$$

When trying to compute $\Lambda_{1(23)}$, the matrix

$$\begin{bmatrix} \frac{1}{12}I_{12} & \frac{1}{144}E_{12,12} \\ \frac{1}{144}E_{12,12} & \frac{1}{12}I_{12} \end{bmatrix}$$

is singular (its rank being 23). Also $\Lambda_{2(3)}$ as calculated above is zero and this would appear to lead us to reject immediately the hypothesis $H_0^{(2)}$ of independence between $\underline{x}_{(2)}$ and $\underline{x}_{(3)}$.

For the matrix

$$\begin{bmatrix} V_{22} & V_{23} \\ V_{32} & V_{33} \end{bmatrix} ,$$

the sum of every row and column in each of the four blocks is equal to $\frac{1}{6}$, and the eigenvalues in the canonical analysis of $\underline{x}_{(2)}$ with $\underline{x}_{(3)}$, i.e. the eigenvalues of

$$\begin{bmatrix} V_{22} & V_{23} \\ V_{32} & V_{33} \end{bmatrix} \begin{bmatrix} V_{22}^{-1} & 0 \\ 0 & V_{33}^{-1} \end{bmatrix}$$

are $\lambda_1 = 2$, $\lambda_2 = ... = \lambda_{23} = 1$, $\lambda_{24} = 0$.

However, some aspects in this analysis deserve a more detailed look. We may begin with the eigenvalues for the overall MCA analysis, i.e. the eigenvalues of VQ where V is as described above and

$$Q \; = \; bdiag(144W^{-1}, \; 12I_{12}, \; 12I_{12}) \; . \tag{6.17}$$

These eigenvalues are, when rounded to 5 decimal places: $\lambda_1 = 3.00000$, $\lambda_2 = 1.84187$, $\lambda_3 = 1.51993$, $\lambda_4 = 1.34230$, $\lambda_5 = 1.32633$, $\lambda_6 = ... = \lambda_{23} = 1.00000$, $\lambda_{24} = .67406$, $\lambda_{25} = .65916$, $\lambda_{26} = .48019$, $\lambda_{27} = .15818$, $\lambda_{28} = \lambda_{29} = .0000$.

Our approach here corresponds to the use of all the indicator variables. Therefore, we ignore the first and the last two eigenvalues.

A careful look at the other eigenvalues shows that they would be the eigenvalues in a canonical analysis between two sets of variables (the comment on the eigenvalues λ_α for such an analysis, in section 3.11 of Chapter 3, are relevant here).

But if we look at the structure of the matrix

$$V_2 \; = \; \begin{bmatrix} V_{22} & V_{23} \\ V_{32} & V_{33} \end{bmatrix} \; = \; \frac{1}{144} \begin{bmatrix} 12I_{12} & E_{12,12} \\ E_{12,12} & 12I_{12} \end{bmatrix}$$

we easily observe that the above eigenvalues should be the same as the eigenvalues in the canonical analysis of $\underline{x}_{(1)}$ with $\left(\underline{x}_{(2)}, \underline{x}_{(3)}\right)$ together. We still would have the problem of the noninvertibility of V_2. We could (as mentioned in section 3.10 of Chapter 3) a generalized inverse. Using a little algebra it is easy to see that the appropriate generalized inverse of V_2 is

$$V_2^- = \begin{bmatrix} 12I_{12} - \frac{3}{4}E_{12,12} & \frac{1}{4}E_{12,12} \\ \frac{1}{4}E_{12,12} & 12I_{12} - \frac{3}{4}E_{12,12} \end{bmatrix}.$$

The MCA and the Canonical Analysis of set $\underline{x}_{(1)}$ with $\left(\underline{x}_{(2)}, \underline{x}_{(3)}\right)$ produce the same eigenvalues and eigenvectors. This is strictly true only for λ_2 through λ_5 and λ_{24} through λ_{27}, the relevant eigenvalues, and associated eigenvectors. In the Canonical Analysis of set $\underline{x}_{(1)}$ with $\left(\underline{x}_{(2)}, \underline{x}_{(3)}\right)$ we get eighteen instead of the nineteen unit eigenvalues and only one out of the two null eigenvalues from MCA, this is because we used a generalized inverse in Q and then $tr(VQ) = 28$ and not 29. Thus all the eigenvalues of the Canonical Analysis will be exactly the same as the ones from the MCA except for the first that will be 2 in the Canonical Analysis and for one of the unit eigenvalues that will not exist.

The equivalence between the results may be easily shown algebraically. For the MCA model consider the matrices V and Q defined by (6.16), (6.13), (6.14), (6.15) and (6.17), then

$$VQ = \begin{bmatrix} I_5 & \frac{1}{12}R' & \frac{1}{12}C' \\ RW^{-1} & I_{12} & \frac{1}{12}E_{12,12} \\ CW^{-1} & \frac{1}{12}E_{12,12} & I_{12} \end{bmatrix}.$$

Let us denote by

$$\underline{u}_\alpha = \begin{bmatrix} \underline{u}_{1\alpha} \\ \underline{u}_{2\alpha} \\ \underline{u}_{3\alpha} \end{bmatrix}$$

(where $\underline{u}_{1\alpha}$ is 5×1, $\underline{u}_{2\alpha}$ is 12×1 and $\underline{u}_{3\alpha}$ is 12×1) the eigenvector of VQ associated with λ_α $(\alpha = 1, ..., 27)$. Then we have, for $\alpha = 2, ..., 27$

$$VQ\underline{u}_\alpha = \lambda_\alpha \underline{u}_\alpha \iff \begin{cases} \underline{u}_{1\alpha} + \frac{1}{12}R'\underline{u}_{2\alpha} + \frac{1}{12}C'\underline{u}_{3\alpha} = \lambda_\alpha \underline{u}_{1\alpha} \\ RW^{-1}\underline{u}_{1\alpha} + \underline{u}_{2\alpha} + \frac{1}{12}E_{12,12}\underline{u}_{3\alpha} = \lambda_\alpha \underline{u}_{2\alpha} \\ CW^{-1}\underline{u}_{1\alpha} + \frac{1}{12}E_{12,12}\underline{u}_{2\alpha} + \underline{u}_{3\alpha} = \lambda_\alpha \underline{u}_{3\alpha} \end{cases} \tag{6.18}$$

$$\iff \begin{cases} \frac{1}{12}R'\underline{u}_{2\alpha} + \frac{1}{12}C'\underline{u}_{3\alpha} = (\lambda_\alpha - 1)\underline{u}_{1\alpha} \\ RW^{-1}\underline{u}_{1\alpha} = (\lambda_\alpha - 1)\underline{u}_{2\alpha} \\ CW^{-1}\underline{u}_{1\alpha} = (\lambda_\alpha - 1)\underline{u}_{3\alpha} \quad . \end{cases}$$

This follows from $E_{12,12}\underline{u}_{2\alpha} = E_{12,12}\underline{u}_{3\alpha} = 0$ for $\alpha = 2, ..., 27$, as $\underline{u}_{21} = \underline{u}_{31} = \frac{1}{12}\frac{1}{\sqrt{3}}E_{12,1}$ (since $\underline{u}'_{21}Q_2\underline{u}_{21} = \underline{u}'_{31}Q_3\underline{u}_{31} = \frac{1}{3}$), so that $\underline{u}'_{2\alpha}Q_2\underline{u}_{21} = 12\underline{u}'_{2\alpha}I_{12}\underline{u}_{21} = \frac{1}{12}\frac{1}{\sqrt{3}}\underline{u}'_{2\alpha}E_{12,1} = 0$ (for $\alpha = 2, ..., 27$), since $Q_2 = 12I_{12}$, and similarly for $\underline{u}_{3\alpha}$ ($\alpha = 2, ..., 27$). Then for $\alpha = 1$, we would have $\underline{u}_{1\alpha} = \frac{1}{144}\frac{1}{\sqrt{3}}WE_{5,1} = \frac{1}{144}\frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 13 & 89 & 36 & 5 \end{bmatrix}' = \frac{1}{144}\frac{1}{\sqrt{3}}\underline{w}$ and $\underline{u}_{2\alpha} = \underline{u}_{3\alpha} = \frac{1}{12}\frac{1}{\sqrt{3}}E_{12,1}$. Hence for $\alpha = 1$

$$R'\underline{u}_{2\alpha} = \frac{1}{12}\frac{1}{\sqrt{3}}R'E_{12,1} = \frac{1}{12}\frac{1}{\sqrt{3}}\underline{w} \ ,$$

$$C'\underline{u}_{3\alpha} = \frac{1}{12}\frac{1}{\sqrt{3}}C'E_{12,1} = \frac{1}{12}\frac{1}{\sqrt{3}}\underline{w} \ ,$$

$$RW^{-1}\underline{u}_{1\alpha} = \frac{1}{144}\frac{1}{\sqrt{3}}RW^{-1}WE_{5,1} = \frac{1}{144}\frac{1}{\sqrt{3}}RE_{5,1} = \frac{1}{12}\frac{1}{\sqrt{3}}E_{12,1}$$

and

$$CW^{-1}\underline{u}_{1\alpha} = \frac{1}{144}\frac{1}{\sqrt{3}}CW^{-1}WE_{5,1} = \frac{1}{144}\frac{1}{\sqrt{3}}CE_{5,1} = \frac{1}{12}\frac{1}{\sqrt{3}}E_{12,1} \ .$$

Then for $\alpha = 1$ from (6.18)

$$VQ\underline{u}_\alpha = \lambda_\alpha \underline{u}_\alpha \iff \begin{cases} 3\frac{1}{144}\frac{1}{\sqrt{3}}\underline{w} = \lambda_\alpha \frac{1}{144}\frac{1}{\sqrt{3}}\underline{w} \\ 3\frac{1}{12}\frac{1}{\sqrt{3}}E_{12,1} = \lambda_\alpha \frac{1}{12}\frac{1}{\sqrt{3}}E_{12,1} \\ 3\frac{1}{12}\frac{1}{\sqrt{3}}E_{12,1} = \lambda_\alpha \frac{1}{12}\frac{1}{\sqrt{3}}E_{12,1} \end{cases}$$

$$\iff \lambda_\alpha = 3 \ .$$

For the Canonical Analysis model we would use

$$Q^* = \begin{bmatrix} 144W^{-1} & 0 & 0 \\ 0 & 12I_{12} - \frac{3}{4}E_{12,12} & \frac{1}{4}E_{12,12} \\ 0 & \frac{1}{4}E_{12,12} & 12I_{12} - \frac{3}{4}E_{12,12} \end{bmatrix}$$

and then

$$VQ^* = \begin{bmatrix} I_5 & \frac{1}{12}R' - \frac{1}{144}\frac{1}{4}(3R' - C')E_{12,12} & \frac{1}{12}C' - \frac{1}{144}\frac{1}{4}(3C' - R')E_{12,12}) \\ RW^{-1} & I_{12} - \frac{1}{24}E_{12,12} & \frac{1}{24}E_{12,12} \\ CW^{-1} & \frac{1}{24}E_{12,12} & I_{12} - \frac{1}{24}E_{12,12} \end{bmatrix}$$

and thus for $\alpha = 2, ..., 27$

$$\begin{aligned} VQ^*\underline{u}_\alpha &= \begin{bmatrix} \underline{u}_{1\alpha} + \frac{1}{12}(R'\underline{u}_{2\alpha} + C'\underline{u}_{3\alpha}) - \frac{1}{144}\frac{1}{4}(3R' - C')(E_{12,12}\underline{u}_{2\alpha} + E_{12,12}\underline{u}_{3\alpha}) \\ RW^{-1}\underline{u}_{1\alpha} + \underline{u}_{2\alpha} - \frac{1}{24}E_{12,12}\underline{u}_{2\alpha} + \frac{1}{24}E_{12,12}\underline{u}_{3\alpha} \\ CW^{-1}\underline{u}_{1\alpha} + \frac{1}{24}E_{12,12}\underline{u}_{2\alpha} + \underline{u}_{3\alpha} - \frac{1}{24}E_{12,12}\underline{u}_{3\alpha} \end{bmatrix} \\ &= \begin{bmatrix} \underline{u}_{1\alpha} + \frac{1}{12}R'\underline{u}_{2\alpha} + \frac{1}{12}C'\underline{u}_{3\alpha} \\ RW^{-1}\underline{u}_{1\alpha} + \underline{u}_{2\alpha} \\ CW^{-1}\underline{u}_{1\alpha} + \underline{u}_{3\alpha} \end{bmatrix}, \end{aligned}$$

while for $\alpha = 1$ we would have

$$R'E_{12,12}\underline{u}_{2\alpha} = \frac{1}{12\sqrt{3}}R'E_{12,12}E_{12,1} = \frac{1}{\sqrt{3}}R'E_{12,1} = \frac{1}{\sqrt{3}}\underline{w}$$

and

$$C'E_{12,12}\underline{u}_{2\alpha} = R'E_{12,12}\underline{u}_{3\alpha} = C'E_{12,12}\underline{u}_{3\alpha} = R'E_{12,12}\underline{u}_{2\alpha} = 0,$$

and thus, for $\alpha = 1$,

$$VQ^*\underline{u}_\alpha = \begin{bmatrix} 2\frac{1}{144\sqrt{3}}\underline{w} - \frac{1}{144\sqrt{3}}\frac{3}{4}\underline{w} + \frac{1}{144\sqrt{3}}\frac{1}{4}\underline{w} + \frac{1}{144\sqrt{3}}\underline{w} - \frac{1}{144\sqrt{3}}\frac{3}{4}\underline{w} + \frac{1}{144\sqrt{3}}\frac{1}{4}\underline{w} \\ 2\frac{1}{12\sqrt{3}}E_{12,1} - \frac{1}{24\sqrt{3}}E_{12,1} + \frac{1}{24\sqrt{3}}E_{12,1} \\ 2\frac{1}{12\sqrt{3}}E_{12,1} + \frac{1}{24\sqrt{3}}E_{12,1} - \frac{1}{24\sqrt{3}}E_{12,1} \end{bmatrix}$$

$$= \begin{bmatrix} 2\frac{1}{144\sqrt{3}}\underline{w} \\ 2\frac{12\sqrt{3}}{E}{}_{12,1} \\ 2\frac{1}{12\sqrt{3}}E_{12,1} \end{bmatrix}$$

what shows that for $\alpha = 2, ..., 27$

$$VQ^*\underline{u}_\alpha = VQ\underline{u}_\alpha$$

while for $\alpha = 1$

$$VQ\underline{u}_\alpha = 3\underline{u}_\alpha$$

but

$$VQ^*\underline{u}_\alpha = 2\underline{u}_\alpha \ .$$

Thus, in this case, it so happens that the MCA model is just the Canonical Analysis model between the sets $\underline{x}_{(1)}$ and $\left(\underline{x}_{(2)}, \underline{x}_{(3)}\right)$. This fact is a consequence of the structure of the data in this particular example or rather a consequence of the structure of the indicator variables, probably not noticed by Gower (1990), and indicates that Fisher's (1938) approach is an adequate one to analyze these data. Bartlett (1951) also essentially used the same approach though his emphasis was on test of goodness of fit of hypothetical scores.

For this particular data set, approach (ii) described in section 6.5 would however not yield satisfactory results given that the removal of one indicator variable from each set without centering would modify the original structure of the matrix

$$V_2 = \begin{bmatrix} V_{22} & V_{23} \\ V_{32} & V_{33} \end{bmatrix} = \begin{bmatrix} \frac{1}{12}I_{12} & \frac{1}{144}E_{12,12} \\ \frac{1}{144}E_{12,12} & \frac{1}{12}I_{12} \end{bmatrix} .$$

However the use of approach (iii) in section 6.5 would have been a better one since centering the original indicator variables and removing one of them from each set would not alter the structure in V or Q, for this example we do not have the original X data matrix available, however it is possible with a little work to obtain it from Table 6.1. This extra work would not be compensated by the gain in simplicity in the analysis, which would be reflected in the non-obtention of the first eigenvalue equal to three and the last two null eigenvalues. This is why we preferred approach (i) when dealing with this example.

Besides, when using approach (iii) instead of approach (i) we also do not obtain the coordinates in the vectors $\underline{u}_\alpha$ and $\underline{a}^\alpha$ for the categories corresponding to the indicator variables removed. These coordinates may however be easily obtained for the vectors $\underline{a}^\alpha$, from the fact that each vector $\underline{a}_2^\alpha$ and $\underline{a}_3^\alpha$ ($\alpha = 2, ..., 26$) with that coordinate or score added should be orthogonal to $E_{12,1}$ while each vector $\underline{a}_1^\alpha$ ($\alpha = 2, ..., 27$) with the missing coordinate added should be orthogonal to $\underline{w} = WE_{5,1}$.

With the use of this approach we would also overcome the difficulty associated with the obtention of an irrelevant eigenvalue of 2 and a last one of zero in the Canonical Analysis of set $\underline{x}_{(2)}$ with $\underline{x}_{(3)}$. Now all the eigenvalues are equal to 1 since those two spurious eigenvalues have been discarded.

As mentioned in section 6.5, if approach (i) is taken then we will not have to consider the spurious eigenvalues when computing our test statistics developed in section 4.7 of Chapter 4.

After these observations, we continue with our analysis and test $H_0^{(1)}$ and $H_0^{(2)}$ using the test statistics (6.10) and (6.11).

We find

$$\Lambda_{1(23)} = .16820$$

and

$$\Lambda_{2(3)} \ = \ 1 \ .$$

We now use the chi-square approximation in (6.10), comparing

$$-n \ log\Lambda_{1(23)} \ = \ -144 \ log(.16820) \ = \ 256.69$$

with the percentiles of the chi-square distribution with 88 degrees of freedom. We obtain a p-value, $p < .001$. Then we compare

$$-n \ log\Lambda_{2(3)} \ = \ -144 \ log(1) \ = \ 0$$

with the percentiles of the chi-square distribution with 121 degrees of freedom and obtain a p-value $p = 1$.

These results lead us to reject $H_0^{(1)}$ and not reject $H_0^{(2)}$ (i.e. to reject the independence of set $\underline{x}_{(1)}$ (response) with the other two sets (sera and individuals), and not to reject the independence of $\underline{x}_{(2)}$ (sera) and $\underline{x}_{(3)}$ (individuals)).

Now if we want to investigate the dimensionality of the relationship between sets $\underline{x}_{(1)}$ and $\left(\underline{x}_{(2)}, \underline{x}_{(3)}\right)$ we may try to assess the significance of successive products

$$\prod_{i=l+1}^{p_1-1} \mu_i \qquad , \qquad l = 1, ..., min(p_1 - 1, p_2 + p_3 - 2)$$

where μ_i is the square of one minus the i^{th} sample canonical correlation between sets $\underline{x}_{(1)}$ and $\left(\underline{x}_{(2)}, \underline{x}_{(3)}\right)$, this is, for the λ eigenvalues reported in the beginning of this section, $\mu_i = \lambda_{i+1}$ for $i = 1, ..., 4$.

Each of such products would then be compared with the percentile of a chi-square distribution. Using a Bartlett-Lawley (Lawley, 1959) correction factor would correspond to comparing

$$-\left(n - l - \frac{1}{2}(p_i - 1 + p_{i+1} + ... + p_m - m + i + 1)\right) log \prod_{i=l+1}^{p_1-1} \mu_i$$

with the percentiles of the distribution

$$\chi^2_f \qquad \text{with} \qquad f = (p_i - 1 - l)(p_{i+1} + ... + p_m - m + i - l) \ .$$

In our case, for $l = 1$,

$$\prod_{i=2}^{p_1-1} \mu_i \;=\; .57733\;, \quad \text{and thus} \quad -\left(144 - 1 - \frac{1}{2}(4 - 1 + 22 + 1)\right)log(.57733) \;=\; 71.41$$

should be compared with the percentiles of χ^2_{63} ($\chi^2_{63(.10)} = 75.50$), giving a p-value, $p > .10$. We, therefore, do not reject the null hypothesis that all but the first canonical correlation are equal to zero, or equivalently we do not reject the hypothesis that the relation between $\underline{x}_{(1)}$ and $\left(\underline{x}_{(2)}, \underline{x}_{(3)}\right)$ may be explained by only one dimension. This will lead us to rely only on the first set of scores $\underline{a}^2$ ($\underline{a}^1$ for the analysis with all indicator variables was discarded), which agrees with Gower's (1990) approach. However, it is not clear if he reached it by intuition or if he somehow tested if only one set of scores was enough to describe the association.

Given the independence between sets $\underline{x}_{(2)}$ and $\underline{x}_{(3)}$, our model (as specified by the model (4.18)) is thus (using parameter estimates rounded to 3 decimal places)

$$Q_1 \underline{u}_{12}\ \underline{x}_{(1)} \;=\; \sqrt{.842}\left(Q_2 \underline{u}_{22}\ \underline{x}_{(2)} + Q_3 \underline{u}_{32}\ \underline{x}_{(3)}\right) \tag{6.19}$$

where

$$Q_1 = 144W^{-1} \quad , \quad Q_2 = 12I_{12} \quad , \quad Q_3 = 12I_{12}$$

and

$$\begin{aligned}
\underline{u}'_{12} &= \left[\ .014 \quad .128 \quad .131 \quad -.236 \quad -.037\ \right] \\
\underline{u}'_{22} &= \left[\ .002 \quad .061 \quad .021 \quad .046 \quad -.007 \quad .002 \quad -.067 \quad -.047 \quad -.036 \quad .061 \quad .021 \quad -.057\ \right] \\
\underline{u}'_{32} &= \left[\ .031 \quad -.018 \quad -.027 \quad -.046 \quad .001 \quad -.028 \quad .096 \quad -.027 \quad .021 \quad -.046 \quad -.009 \quad .051\ \right].
\end{aligned}$$

Model (6.19) then becomes

$$\begin{aligned}
&\left[2.017 \quad 1.421 \quad .211 \quad -.942 \quad -1.072\right]\ \underline{x}_{(1)} \;= \\
&\quad = .91753\Big(\left[\ .028 \quad .730 \quad .251 \quad .549 \quad -.086 \quad .028 \quad -.802 \quad -.561 \quad -.434 \quad .730 \quad .256 \quad -.688\right]\ \underline{x}_{(2)} \;+ \\
&\qquad + \left[\ .376 \quad -.213 \quad -.327 \quad -.549 \quad .010 \quad -.333 \quad 1.148 \quad -.320 \quad .251 \quad -.549 \quad -.105 \quad .610\right]\ \underline{x}_{(3)} \quad \Big)
\end{aligned}$$

In fact, before writing model (6.19) or (6.20) we should first examine whether both sets of variables, $\underline{x}_{(2)}$ and $\underline{x}_{(3)}$ (sera and individuals) contribute significantly to the model. This question is answered by assessing the significance of the two test statistics

$$\Lambda_{12|3} = \frac{\Lambda_{1(23)}}{\Lambda_{1(3)}} \tag{6.20}$$

and

$$\Lambda_{13|2} = \frac{\Lambda_{1(23)}}{\Lambda_{1(2)}} \tag{6.21}$$

(see sections 4.7 through 4.10 of Chapter 4).

The statistics $\Lambda_{12|3}$ and $\Lambda_{13|2}$ could also be obtained directly from their determinantal equations but since we already have $\Lambda_{1(23)}$ and both Λ_{13} and Λ_{12} are very easy to obtain we will rather use the expressions in (6.21) and (6.22). The eigenvalues λ_α for the GCA model corresponding to the Canonical Analysis of sets $\underline{x}_{(1)}$ with $\underline{x}_{(2)}$ and $\underline{x}_{(1)}$ with $\underline{x}_{(3)}$, respectively, and with the use of all the indicator variables are:

– for the Canonical Analysis of set $\underline{x}_{(1)}$ with $\underline{x}_{(2)}$

$(\lambda_1 = 2.00000),\ \lambda_2 = 1.62800,\ \lambda_3 = 1.33159,\ \lambda_4 = 1.25707,\ \lambda_5 = 1.18095,$
$\lambda_6 = ... = \lambda_{12} = 1.00000,\ \lambda_{13} = .81905,\ \lambda_{14} = .74294,\ \lambda_{15} = .66842,$
$\lambda_{16} = .37200,\ (\lambda_{17} = .00000)$

– for the Canonical Analysis of set $\underline{x}_{(1)}$ with $\underline{x}_{(3)}$

$(\lambda_1 = 2.00000),\ \lambda_2 = 1.60913,\ \lambda_3 = 1.40876,\ \lambda_4 = 1.21228,\ \lambda_5 = 1.12670,$
$\lambda_6 = ... = \lambda_{12} = 1.00000,\ \lambda_{13} = .87330,\ \lambda_{14} = .78772,\ \lambda_{15} = .59124,$
$\lambda_{16} = .39087,\ (\lambda_{17} = .00000)$.

Thus,

$$\Lambda_{1(2)} = .48694$$

and

$$\Lambda_{1(3)} = .49848\ ,$$

so that

$$\Lambda_{12|3} = \frac{\Lambda_{1(23)}}{\Lambda_{1(3)}} = .33742$$

and

$$\Lambda_{13|2} = \frac{\Lambda_{1(23)}}{\Lambda_{1(2)}} = .34542 \ .$$

We should than compare

$$-n \ log\Lambda_{12|3} = -144 \ log.33742 = 156.45$$

and

$$-n \ log\Lambda_{13|2} = -144 \ log.34542 = 153.07$$

with the percentiles of the distribution χ^2_{44} $\left(\chi^2_{44(.001)} = 74.82\right)$. Both give a p-value, $p < .001$.

So our conclusion is that both sets sera and individuals contribute significantly to the model.

Fisher (1938) and Bartlett (1951) put together $\underline{x}_{(2)}$ and $\underline{x}_{(3)}$ and ignore their separate existence. Gower (1990), on the other hand, considers the three individual sets and analyzes them together using an MCA, but loosing the clear treatment of set $\underline{x}_{(1)}$ as the dependent set. Our approach of the problem, however beginning with the use of the MCA model, lead us to a final model that rather agrees with Fisher's (1938) approach. This example brings to our attention the need for a general Canonical Analysis type model of one set with two or more independent sets and test criteria for measuring the conditional importance of these in the model.

Really the final model obtained also leads us to think about another aspect of the analysis of this data set.

The analysis is more like a repeated measures analysis of variance, where it really happens that the dependent set is not formed by one (or more) continuous variables but by the levels of a categorical response variable.

What we really want to assess is if the relationship between 'response' (i.e. set $\underline{x}_{(1)}$) and 'sera' (set $\underline{x}_{(2)}$) is significant, when accounting in the model for 'individuals (set $\underline{x}_{(3)}$).

This means that we should really deal with the analysis of this data set by fitting the model (6.19) that really corresponds to test the null hypothesis of no association between $\underline{x}_{(1)}$ and $\left(\underline{x}_{(2)},\ \underline{x}_{(3)}\right)$ and then to test $H_0^{2(1)}$: 'all the canonical correlations between sets $\underline{x}_{(1)}$ and $\underline{x}_{(2)}$, given that set $\underline{x}_{(3)}$ is in the model, are null'. The test statistic corresponding to this hypothesis is $\Lambda_{12|3}$ and as we saw above would lead us to reject $H_0^{2(1)}$.

The above discussion points rather in the direction of the need for models in which we model one set of variables as a linear function of other sets, but in which we want to be able to test for the significance in the model of individual sets of variables, conditional on the presence in the model of other sets. These models that may be seen as a generalization of the usual Multiple Regression Analysis model and will in general be designated by Block Regression models will be discussed in further detail in the next Chapter. The essentials elements have already been developed in sections 4.7 through 4.10 of Chapter 4, where we discussed the more general GCA model and associated hypotheses and test statistics.

CHAPTER 7

THE BLOCK REGRESSION MODEL

7.1 Introduction.

Often there is a need for models that express one set of variables as a linear function of the others. Such models were already introduced in section 4.8 of Chapter 4, as the submodels which are parts of the global GCA model (4.18), corresponding to the tests of the hypotheses $H_0^{(k)}$ $(k = 1, ..., m)$.

Suppose, without loss of generality and only for the sake of simplicity of exposition, that we want to model the set $\underline{x}_{(1)}$ as a linear function of the $m-1$ sets $\underline{x}_{(2)}$ through $\underline{x}_{(m)}$. Suppose also that we want to test if a given set of variables $\underline{x}_{(k)}$ $(k = 2, ..., m)$ is significant in the model or not.

This situation is well known in MANOVA and MANCOVA models where set $\underline{x}_{(1)}$ is formed by say p_1 continuous variables assumed to have a joint normal distribution and where the other $m-1$ sets are either all formed by indicator variables for the levels of experimentally controlled factors and interactions or among them some may be formed by continuous variables, then assumed to have a joint normal distribution.

7.2 The model under the assumption of normality.

First, let us assume that all the m sets of variables, $\underline{x}_{(1)}$ through $\underline{x}_{(m)}$ have a joint normal distribution specified by (4.1) and (4.2), i.e. that all the $p = \sum_{k=1}^{m} p_k$ variables that form the m sets of variables have a joint p-variate normal distribution.

Suppose that set $\underline{x}_{(1)}$ is our dependent or predictand set and we want to model it as a linear function of the other $m-1$ sets. We would then use the model

$$\begin{aligned} \beta_{1(1)}\underline{x}_{(1)} &= \Lambda_{1(2,...,m)}\beta_{1(2,...,m)}\left(\underline{x}'_{(2)}, ..., \underline{x}'_{(m)}\right)' \ + \ \underline{\epsilon} \\ &= \Lambda_{1(2,...,m)}\left(\beta_{1(2)}\underline{x}_{(2)} + \beta_{1(3)}\underline{x}_{(3)} + ... + \beta_{1(k)}\underline{x}_{(k)} + ... + \beta_{1(m)}\underline{x}_{(m)}\right) \ + \ \underline{\epsilon}\ , \end{aligned} \tag{7.1}$$

where the order of the parameter matrices $\beta_{1(k)}$ $(k = 1, ..., m)$ is $p^* \times p_k$, with

$$p^* \ \leq \ min(p_1,\ p_2 + ... + p_m)$$

and where the parameter matrix $\beta_{1(2,...,m)}$ is

$$\underset{p^* \times (p_2 + ... + p_m)}{\beta_{1(2,...,m)}} \quad = \quad \left[\beta_{1(2)} \mid \beta_{1(3)} \mid ... \mid \beta_{1(k)} \mid ... \mid \beta_{1(m)}\right]\ .$$

Then the test statistic for the overall fit of the model is (from sections 4.7 through 4.9, in Chapter 4) the test statistic corresponding to the hypothesis $H_0^{(1)}$ (see hypotheses $H_0^{(k)}$, for $k = 1$ in sections 4.7 and 4.8 of Chapter 4), namely

$$\Lambda_{1(2,...,m)} \ = \ \frac{|V_{11} - V_{1(2,...,m)}V^{-1}_{(2,...,m)(2,...,m)}V_{(2,...,m)1}|}{|V_{11}|}\ , \tag{7.2}$$

where,

$$V \ = \ \left[\begin{array}{c|c} V_{11} & V_{1(2,...,m)} \\ \hline V_{(2,...,m)1} & V_{(2,...,m)(2,...,m)} \end{array}\right] \ = \ X'DX$$

is the same matrix as in Chapters 3 and 4 and throughout this dissertation.

The test statistic $\Lambda_{1(2,...,m)}$ is (from Theorem 5 in section 4.7 of Chapter 4) distributed as

$$\Lambda_{1(2,...,m)} \ \overset{app}{\sim} \ \Lambda\left(\frac{(1-l)^2}{g},\ p_1,\ p_2 + ... + p_m\right) \tag{7.3}$$

where l and g have been defined in Lemma 1 of Chapter 4.

Then, to test for the importance (or significance) of set $\underline{x}_{(k)}$ $(k = 2, ..., m)$ in model (7.1) is the same as to test $H_0^{k(1)}$, the null hypothesis of independence between $\underline{x}_{(1)}$ and $\underline{x}_{(k)}$, given that the canonical correlations between $\underline{x}_{(1)}$ and all the other sets ($\underline{x}_{(2)}$ through $\underline{x}_{(k-1)}$ and $\underline{x}_{(k+1)}$ through $\underline{x}_{(m)}$) are known (see Theorem 5 in section 4.7 of Chapter 4 and also sections 4.8 and 4.9 of the same Chapter), i.e. $H_0^{k(1)}$ may also be written as

$H_0^{k(1)}$: the partial canonical correlations between $\underline{x}_{(1)}$ and $\underline{x}_{(k)}$ are all equal to zero, given the sets $\underline{x}_{(2)}$ through $\underline{x}_{(k-1)}$ and $\underline{x}_{(k+1)}$ through $\underline{x}_{(m)}$ are in the model.

The associated test statistic is (see Theorem 5 in section 4.7 of Chapter 4 and sections 4.8 and 4.9 of the same Chapter)

$$\Lambda_{1k|(2,...,k-1,k+1,...,m)} = \frac{\Lambda_{1(2,...,m)}}{\Lambda_{1(2,...,k-1,k+1,...,m)}} = \frac{|V_{11.t} - V_{1k.t}V_{kk.t}^{-1}V_{k1.t}|}{|V_{11.t}|} \tag{7.4}$$

where t is the sequenced set

$$t = \{2, ..., k-1, k+1, ..., m\} .$$

Under $H_0^{k(1)}$ $\Lambda_{1k|(2,...,k-1,k+1,...,m)}$ has the distribution

$$\Lambda_{1k|(2,...,k-1,k+1,...,m)} \overset{app}{\sim} \Lambda\left(\frac{(1-l)^2}{g} - (p_2 + ... + p_{k-1} + p_{k+1} + ... + p_m),\ p_1,\ p_k\right) . \tag{7.5}$$

Model (7.1) is really a Canonical Analysis type model. However there a subtle difference from a usual Canonical Analysis model. The usual Canonical Analysis model between sets $\underline{x}_{(1)}$ and the macroset $\left(\underline{x}'_{(2)}, ..., \underline{x}'_{(m)}\right)'$ would use in the observations space a metric

$$Q^* = \left[\begin{array}{c|c} V_{11}^{-1} & 0 \\ \hline 0 & V_{(2,...,m)(2,...,m)}^{-1} \end{array}\right] \tag{7.6}$$

while we are using the metric Q as defined in sections 3.3 of Chapter 3 and 4.5 of Chapter 4, i.e.

$$Q = \begin{bmatrix} V_{11}^{-1} & 0 & \dots & 0 & \dots & 0 \\ 0 & V_{22}^{-1} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \dots & V_{kk}^{-1} & \dots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & V_{mm}^{-1} \end{bmatrix} \tag{7.7}$$

i.e. $Q = bdiag(Q_k) = bdiag(X_k' D X_k)^{-1}$. In the particular case when $\underline{x}_{(1)}$ consists of only one variable, model (7.1) is a Multiple Regression type model. The difference to the usual Multiple Regression model being, that the metric in the observations space is still of the type (7.7) and not of the the type (7.6), with V_{11}^{-1} being 1×1, as it happens in the usual Multiple Regression model. Anyway, similarly to what happens in the Multiple Regression model, in this particular case the test statistic (7.2), for the overall fit of the model, is an F test since (see for example Kshirsagar, 1972)

$$\frac{\frac{(1-l)^2}{g} - (p_2 + \ldots + p_m)}{p_2 + \ldots + p_m} \; \frac{1 - \Lambda_{1(2,\ldots,m)}}{\Lambda_{1(2,\ldots,m)}} \overset{app}{\sim} F_{(p_2+\ldots+p_m),\; \frac{(1-l)^2}{g} - (p_2+\ldots+p_m)} \; . \tag{7.8}$$

In this particular case we may then see the test statistic (7.4) as a generalization of the partial or conditional F test commonly used in Regression, since, for $p_1 = 1$,

$$\frac{c - p_k}{p_k} \; \frac{1 - \Lambda_{1k|(2,\ldots,k-1,k+1,\ldots,m)}}{\Lambda_{1k|(2,\ldots,k-1,k+1,\ldots,m)}} \overset{app}{\sim} F_{p_k,\; c-p_k} \; , \tag{7.9}$$

where

$$c = \frac{(1-l)^2}{g} - (p_2 + \ldots + p_{k-1} + p_{k+1} + \ldots + p_m) \; .$$

We should note that the above F distributions in (7.8) and (7.9), as well as the distributions in (7.3) and (7.5) are exact when we use masses $m_i = \frac{1}{n}$ for all our n observations, being the case that then we may replace $\frac{(1-l)^2}{g}$ by $n - 1$ in those distributions (see Lemma 1 in Chapter 4 and the note after that Lemma).

7.3 The model when the assumption of normality does not hold.

In the most of situations when for some or all of the m sets of variables involved in model (7.1) it is not possible to assume a joint multivariate normal distribution, the use of the above test statistics can be justified only when the sample sizes are large enough and tests of significance are not taken too precisely.

However, we shall analyze in further detail some situations where further approximations of the distributions of the above test statistics may still be used with a fair degree of approximation.

Let us first consider the case where the set $\underline{x}_{(1)}$ has a p_1-variate normal distribution (with $p_1 > 1$) and all the other sets are formed by indicator variables for the levels of factors controlled by the experimenter (and possibly also for the levels of interactions of some of such factors). In this situation the sets of indicator variables should be regarded as fixed. This is the situation in MANOVA (or Univariate Analysis of Variance, if $p_1 = 1$). Our model is then still related to the Canonical Analysis model in the same way that model (7.1) was when under the normal assumptions (see for example Kshirsagar (1972)). Then we may still apply the above test statistics in (7.2) and (7.4) to test the above corresponding hypotheses $H_0^{(1)}$ and $H_0^{k(1)}$ $(k = 2, ..., m)$. The distributions of these test statistics are still given by (7.3) and (7.4) respectively, where if we take p_k as the number of levels of the factor or interaction A_k, we should substitute each p_k by $p_k - 1$ for $k = 2, ..., m$. This means that in this case the two test statistics have the distributions

$$\Lambda_{1(2,...,m)} \overset{app}{\sim} \Lambda\left(\frac{(1-l)^2}{g},\ p_1,\ p_2 + ... + p_m - m + 1\right)$$

and

$$\Lambda_{1k|(2,...,k-1,k+1,...,m)}$$

$$\overset{app}{\sim} \Lambda\left(\tfrac{(1-l)^2}{g} - (p_2 + ... + p_{k-1} + p_{k+1} + ... + p_m - m + 2),\ p_1,\ p_k - 1\right) .$$

We should note that if set $\underline{x}_{(1)}$ consists of only one variable, i.e. if we are in the Univariate Analysis of Variance situation, and if we then apply the F distributions in (7.8) and (7.9) with p_k replaced by $p_k - 1$ for $k = 2, ..., m$ we will be using just the usual F test statistics in Analysis of Variance.

A similar situation is the one when $\underline{x}_{(1)}$ is still assumed to have a p_1-variate normal distribution but we assume for all the other $m-1$ sets, $\underline{x}_{(2)}$ through $\underline{x}_{(m)}$ a multinomial distribution with

$$f\left(\underline{x}_{(k)}\right) = \frac{1}{\prod_{j=1}^{p_k} x_{(k)j}} \prod_{j=1}^{p_k} \pi_j^{x_{(k)j}} = \prod_{j=1}^{p_k} \pi_j^{x_{(k)j}} \qquad (7.10)$$

$$(j = 1, ..., p_k \quad ; \quad k = 2, ..., m)$$

where $x_{(k)j}$ is either zero or one and π_j is the probability that the i^{th} observation falls in the j^{th} category of the k^{th} $(k = 2, ..., m)$ attribute.

Then we may still apply the above test statistics (7.2) and (7.4) and assume for them the same distributions as in the situation where these $m-1$ regressor sets were regarded as fixed. Since the multinomial distribution is singular, it will be appropriate to discard one variable from each set and use $p_k - 1$ instead of p_k as the effective number of variables in the sets $\underline{x}_{(2)}$ through $\underline{x}_{(m)}$, this way obtaining the same distributions as in (7.3) and (7.5) with p_k replaced by $p_k - 1$ for $k = 2, ..., m$.

Yet another situation of interest is when all the m sets of variables, $\underline{x}_{(1)}$ through $\underline{x}_{(m)}$, have a multinomial distribution of the type (7.10). In this case model (7.1) corresponds to the analysis of a band of the Burt table (see Chapter 6). Once again now the problem is the applicability of the test statistics (7.2) and (7.4) as well as the problem of their distributions. This problem was discussed in Chapter 6 when we dealt with the MCA model and the simultaneous analysis of the correlation structure among m categorical variables or m sets of indicator variables. From the discussion in that Chapter, mainly in section 6.7, and from the above note on the singularity of the multinomial distribution we may conclude that we may still use the same test statistics, now having with distributions similar to the ones in (7.3) and (7.5) with p_k replaced by $p_k - 1$ for $k = 1, ..., m$. This is, in this situation

$$\Lambda_{1(2,...,m)} \overset{app}{\sim} \Lambda\left(\frac{(1-l)^2}{g},\ p_1 - 1,\ p_2 + ... + p_m - m + 1\right)$$

and

$$\Lambda_{1k|(2,...,k-1,k+1,...,m)}$$
$$\overset{app}{\sim} \Lambda\left(\tfrac{(1-l)^2}{g} - (p_2 + ... + p_{k-1} + p_{k+1} + ... + p_m - m + 2),\ p_1 - 1,\ p_k - 1\right) ,$$

or we may use the chi-square approximations specified in section 6.7 of Chapter 6, more precisely,

$$-n\ log\Lambda_{1(2,...,m)} \overset{app}{\sim} \chi^2_{(p_1-1)(p_2+...+p_m-m+1)}$$

and

$$-n\ log\Lambda_{1k|(2,...,k-1,k+1,...,m)} \overset{app}{\sim} \chi^2_{(p_1-1)(p_k-1)} .$$

Thus, model (7.1) with the above test statistics and corresponding approximate distributions, when applied to sets of indicator variables for the categories of categorical variables may present an alternative to models developed by Benzecri (1982) to analyze bands of the Burt table.

Situations may occur when some of the sets $\underline{x}_{(2)}$ through $\underline{x}_{(m)}$ consist of continuous variables while the others consist of indicator variables, and $\underline{x}_{(1)}$ may be of either one of these two types. Then, if we assume a multivariate normal distribution for the sets of continuous variables and a multinomial distribution for the sets of indicator variables, we may say that by a kind of 'interpolation principle' we may still apply the test statistics (7.3) and (7.5) and use for them approximate distributions as the ones in (7.4) and (7.6) with p_k replaced by $p_k - 1$ for each of the sets that are formed by indicator variables.

7.4 Some final remarks about these models.

Models of this type are very interesting and have important applications in several areas where the modeling of one set of variables as a linear function of $m-1$ other sets may be useful for prediction.

Models of the type (7.1) are particular cases of the GCA model in the sense that they are a part of the overall model (4.18) specified in section 4.8 of Chapter 4, and represent a generalization of the usual Canonical Analysis model where now we have not only the possibility to test the overall fit of the model, using $\Lambda_{1(2,\ldots,m)}$, we may also test for the importance of an individual set of variables in the model, using partial Wilks' lambda statistics such as $\Lambda_{1k|(2,\ldots,k-1,k+1,\ldots,m)}$.

CHAPTER 8

DIRECTIONS FOR FUTURE RESEARCH

The GCA technique, model, associated hypotheses, and respective test statistics and their distributions have been the main topics of this dissertation. GCA was introduced, as its name indicates, as a generalization of the usual Canonical Analysis to more than two sets of variables. But, we should note that GCA is also a generalization of the Principal Components Analysis (more precisely, Principal Components Analysis (PCA) on the correlation matrix), since in the particular case when each of the m sets of variables consists of only one variable, the GCA technique and model reduces to the PCA technique and model. In this case, we should also note that each of the submodels of the model (4.18) are multiple regression models and that the test statistics $\Lambda_{k(k+1,\ldots,m)}$ and $\Lambda_{ki|(k+1,\ldots,i-1,i+1,\ldots,m)}$ are related to the F distribution through

$$\frac{\frac{(1-l)^2}{g}-m+k}{m-k}\;\frac{1-\Lambda_{k(k+1,\ldots,m)}}{\Lambda_{k(k+1,\ldots,m)}}\quad\overset{app}{\sim}\quad F_{(m-k),\frac{(1-l)^2}{g}-m+k} \tag{8.1}$$

and

$$\left(\frac{(1-l)^2}{g}-m+k\right)\;\frac{1-\Lambda_{ki|(k+1,\ldots,i-1,i+1,\ldots,m)}}{\Lambda_{ki|(k+1,\ldots,i-1,i+1,\ldots,m)}}\quad\overset{app}{\sim}\quad F_{1,\frac{(1-l)^2}{g}-m+k}\,, \tag{8.2}$$

where l and g are defined in Lemma 1 of Chapter 4 and $m=p$, the total number of variables.

The above distributions, as in the case of the models discussed in Chapter 7, are exact if we take $m_i = \frac{1}{n}$ for all $i = 1, ..., n$. In this case $\frac{(1-l)^2}{g}$ reduces to $n - 1$. Also similar to what happened for the models discussed in Chapter 7, when p_1 was equal to 1, $\Lambda_{ki|(k+1,...,i-1,i+1,...,m)}$ becomes a partial F (or t) test statistic.

With the test statistics (8.1) and (8.2) we may then test in PCA, the hypotheses corresponding to $H_0^{(k)}$ $(k = 1, ..., m-1)$ and $H_0^{i(k)}$ $(k = 1, ..., m-1; i = k+1, ..., m)$ in GCA (see sections 4.7 through 4.9 in Chapter 4).

Another remark concerning the GCA model is that when applied to the study of sets of indicator variables, it may be used to study any nonlinear relationships among m continuous variables. The range of these m variables should first be subdivided in a convenient way such that it leads to an appropriate categorization of these variables. This process of categorization would depend, among other factors, on the distributions assumed for such variables. The resulting distribution of the frequencies will still be multinomial but the probabilities will be functions of parameters of the underlying continuous distribution. For this reason the appropriateness of the approximate distributions derived in Chapter 6 for the test statistics may be questionable. However these approximate distributions have been based on approximations of such test statistics by chi-square distributed test statistics that are applicable even under a non-parametric setting. Further research in this area is surely appropriate.

Another interesting area of future research would be to relate the models in Chapter 7, when applied to categorical variables, with nonlinear models, like log-linear models, commonly used to study the relationships among categorical variables. This relationship is almost obvious for the models discussed in Chapter 7, when applied to sets of indicator variables, if only the first set of scores is retained, but does not seem so straightforward when we retain, in the model, more than the first set of scores.

Further research on the distribution of the, Burt matrix, $B = X'DX$ (used in

Chapter 6) when we are using a centered X data matrix, would be much useful and appropriate. When using a centered data matrix X, then, the matrix B is formed by the MLE's of the variances and covariances of m correlated multinomial distributions. We feel that, for a large sample size, such a distribution may converge to a Wishart or a related distribution. The distribution of such type of matrices is unknown and if possible to derive, it will be useful in obtaining better approximations to the distributions of the test statistics used.

Another interesting area of research would be the further detailed study of the distribution of the Wilks' lambda type test statistics used when we have a mixture of sets of continuous and indicator variables in our analysis, and we may assume a multivariate normal distribution for the sets of continuous variables and a multinomial distribution for the sets of indicator variables. This problem was only touched upon briefly, in Chapter 7, where we used an 'interpolation principle' to justify the use of the test statistics developed with an approximate Wilks' lambda distribution in such type of situations. Still, the study of the distribution of the test statistics could be further extended to the case of other distributions, such as elliptically contoured distributions, where the strict assumption of normality is not imposed.

Also, since at present there exist some techniques to transform to near normality the distributions of some non-normal variables (Young, 1991), a related topic of interest would be to study how the application of such transformations may affect the distributions of the test statistics and their ability to detect true correlations among the original variables.

Comparison of the performance of the asymptotic distributions of $\Lambda = |VQ|$ obtained in Chapter 5 with other asymptotic distributions in the literature and the problem of tests of significance of individual eigenvalues λ_α, of VQ, as well as the related problem of assessing the adequate number of eigenvectors of VQ necessary to obtain a good representation of all the n observations, or equivalently, the number of sets of scores or factors in our model are also other points of interest.

The studies conducted in order to assess the performance of the asymptotic distributions derived in Chapter 5 for $\Lambda = |VQ|$ were insufficient to deserve any detailed mention in this dissertation but still they seemed encouraging even in the case of $m = 2$, when comparing with Bartlett's approximation.

Tests of significance for individual eigenvalues λ_α, for the case $m > 2$ could be based, for example, on the asymptotic distributions developed in Chapter 5 or any other asymptotic distribution of $\Lambda = |VQ|$, much in the same way as we usually test the for the number of non-null population canonical correlations. But this method is not so easy. For example, in the case $m = 2$ $\lambda_1 + \lambda_{p_1+p_2} = 2$ and $\lambda_1\lambda_{p_1+p_2}$ $= 1 - \hat{\rho}_1^2$ (see section 3.11 of Chapter 3), where ρ_1 represents the largest population canonical correlation. Therefore, for testing the canonical correlation ρ_1, we need not just a test of λ_1 but tests of functions of λ's. For $m > 2$, we know that the eigenvalues λ_α are functions of the canonical correlations between set $\underline{x}_{(i)}$ and the macroset $\left(\underline{x}_{(i+1)}, ..., \underline{x}_{(m)}\right)$ for $i = 1, ..., m-1$, but the possible relations among the eigenvalues λ_α are not known, except for the obvious trivial one, namely $\sum_{\alpha=1}^{p} \lambda_\alpha = p$. This makes the problem hard.

These problems, of the significance of individual eigenvalues λ_α, (called in this dissertation as generalized canonical correlations) and their relation with the canonical correlations between the $m-1$ pairs of sets $\underline{x}_{(i)}$ and $\left(\underline{x}_{(i+1)}, ..., \underline{x}_{(m)}\right)$ for $i = 1, ...m-1$, are intimately related with the assessment of the number of columns in

$$U = \left[\underline{u}_1 \mid ... \mid \underline{u}_\alpha \mid ... \mid \underline{u}_q\right] \qquad q \leq p \ ,$$

(or equivalently, the number of columns in QU) that we need in order to get a good representation of the set of n observations, or, correspondingly a good fit of the overall GCA model. Another important related problem is to relate the parameter matrix QU with the regression parameters $\beta_{i(i+1,...,m)}$ in the models in (4.18).

BIBLIOGRAPHY

BIBLIOGRAPHY

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons, New York.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis.* 2^{nd} ed. John Wiley & Sons, New York.

Andersson, S. A. and Perlman, M. D. (1988). *Lattice models for conditional independence in a multivariate normal distribution.* Technical report n. 9. Inst. Math. Stat., Univ. Copenhagen.

Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Proc. Camb. Phil. Soc.*, **34**, 33 - 40.

Bartlett, M. S. (1951). The goodness of fit of a single hypothetical discriminant function in the case of several groups. *Ann. Eugen.*, **16**, 199 - 214.

Bartlett, M. S. (1963). *Multivariate Statistics – A Chapter in Theoretical and Mathematical Biology.* Blaisdell, New York.

Benzecri, J. P. (1982). Sur la généralisation du Tableau de Burt et son analyse par bandes. *Cahiers Anal. Donn.*, **7**, 33 - 43.

Boullion, T. L. and Odell, P. L. (1971). *Generalized Inverse Matrices.* Wiley-Interscience, New York.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, **36**, 317 - 346.

Burt, C. (1950). The factorial analysis of qualitative data. *Brit. J. Stat. Pshycol.*, **3**, 166 - 185.

Cailliez, F. and Pagès, J. P. (1976). *Introduction à l'Analyse des Données.* SMASH, Paris.

Carroll, J. D. (1968). Generalization of Canonical Analysis to three or more sets of variables. *Proc.* 76^{th} *Convent. Amer. Psychol. Assoc.*, 227 - 228.

Cazès, P., Bonnefous, S., Baumerder, A., Pagès, J. P. (1976). Déscription cohérente des variables qualitatives prises globalement et de leurs modalités. *Stat. Anal. Donn.*, **2**, 48 - 62.

Constantine, A. G. (1963). Some noncentral distribution problems in multivariate analysis. *Ann. Math. Statist.*, **34**, 1270 - 1285.

Consul, P. C. (1966). On the exact distributions of the likelihood ratio criteria for testing linear hypotheses about regression coefficients. *Ann. Math. Stat.*, **37**, 1319 - 1330.

Consul, P. C. (1967). On the exact distributions of likelihood ratio criteria for testing independence of sets of variates under the null hypothesis. *Ann. Math. Stat.*, **38**, 1160 - 1169.

Dauxois, J. and Pousse, A. (1976). Essai de synthèse sur quelques problèmes mathèmatiques de l'analyse des données. *Stat. Anal. Donn.*, **1**, 47 - 67.

Escofier, B. and Pagès, J. P. (1982). Comparaison de groupes de variables définis sur le même ensemble d'individus. *Rapport IRISA n.166*, Rennes.

Escofier, B. and Pagès, J. P. (1984). L'Analyse Factorielle Multiple. *Cahiers BURO*, n.42, 1 - 68.

Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics*, **29**, 751 - 760.

Escoufier, Y. (1980). L'Analyse conjointe de plusieurs matrices de données. in Jolivet, E., Legay, J. M. and Tomassone, R. (eds.) *Biometrie et Temps*, n.3, 59 - 76.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications.* Vol. I. 3rd ed. John Wiley & Sons, New York.

Fisher, R. A. (1938). *Statistical Methods for Research Workers.* 7th ed. Oliver and Boyd, Edinburgh.

Geer, J. P. Van de (1984). Linear relations among k sets of variables. *Psychometrika*, **49**, 79 - 94

Glaçon, F. (1981). *Analyse Conjointe de Plusieurs Matrices de Données. Comparaison de Différentes Méthodes.* Thesis 3ème cycle. Univ. Scient. et Méd. Grenoble, Grenoble.

Glynn, W. J. (1980). Asymptotic representations of the densities of canonical correlations and latent roots in MANOVA when the population parameters have arbitrary multiplicity. *Ann. Statist.*, **8**, 958 - 976.

Glynn, W. J. and Muirhead, R. J. (1978). Inference in canonical correlation analysis. *J. Mult. Anal.*, **8**, 468 - 478.

Gower, J. C. (1975). Generalized Procustes analysis. *Psychometrika*, **40**, 33 - 51.

Gower, J. C. (1990). Fisher's Optimal Scores and Multiple Correspondence Analysis. *Biometrics*, **46**, 947 - 961.

Gupta, R. D. and Richards, D. (1979). Exact distributions of Wilks' Λ under the null and non-null (linear) hypotheses. *Statistica*, **39**, 333 - 342.

L'Hermier des Plantes, H. (1976). *Structuration des Tableaux à Trois Indices de la Statistique.* Thesis, U.S.T.L., Montpellier II.

Hotelling, H. (1936). Relation between two sets of variables. *Biometrika.*, **28**, 129 - 149.

Hsu, P. L. (1941). On the limiting distribution of the canonical correlations. *Biometrika*, **32**, 38 - 45.

Huygens, C. (1986). *The Pendulum Clock or Geometrical Demonstrations Concerning the Motion of Pendula as Applied to Clocks.* The Iowa State University Press, Ames. Translation by Richard J. Blackwell.

James, A. T. (1954). Normal multivariate analysis and the orthonormal group. *Ann. Math. Stat.*, **25**, 40 - 75.

James, A. T. (1955). The non-central Wishart distribution. *Proc. Roy. Soc.*, Ser. A, **229**, 364 - 366.

James, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.*, **35**, 475 - 501.

Kendall, M. G. and Stuart, A. (1977). *The Advanced Theory of Statistics.* 4th ed. C. Griffin, London.

Kettenring, J. R. (1971). Canonical Analysis of several sets of variables. *Biometrika*, **58**, 433 - 451.

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

Kshirsagar, A. M. (1983). Lambda criterion, Wilks's. in Kotz-Johnson (ed.) *Encyclopedia of Statistical Sciences*, Vol. 4. John Wiley & Sons, New York.

Kshirsagar, A. M. (1989). Effect of a missing observation on Hotelling's generalized T_0^2. *Letters in Prob. and Statist.*, **8**, 163 - 166.

Kshirsagar, A. M. and Gupta, R. P. (1982). *Present status of the research in the area of direction and collinearity factors of Wilks'* Λ. Research report n. 12. Dep. of Mathematics, Dalhousie Univ.

Kulp, R. W. and Nagarsenker, B. N. (1984). An asymptotic expansion of the nonnull distribution of Wilks criterion for testing the multivariate linear hypothesis. *Ann. Statist.*, **12**, 1576 - 1583.

Lafosse, R. (1985). *Analyses Procustéennes de Deux Tableaux.* Thesis 3ème cycle. Univ. Paul Sabatier, Toulouse.

Lavit, C. and Roux, C. (1982). Analyse conjointe de plusieurs tableaux de données par la méthode STATIS. *Séminaire ISUP.* Paris.

Lawley, D. N. (1959). Tests of significance in canonical analysis. *Biometrika*, **46**, 59 - 66.

Lee, Y. S. (1971a). Distribution of the canonical correlations and asymptotic expansions for distributions of certain independence test statistics. *Ann. Math. Stat.*, **42**, 526 - 537.

Lee, Y. S. (1971b). Asymptotic formulae for the distribution of a multivariate test statistic: power comparisons of certain multivariate tests. *Biometrika*, **58**, 647 - 651.

Lee, Y. S. (1972). Some results on the distribution of Wilks's likelihood-ratio criterion. *Biometrika*, **95**, 649 - 664.

Muirhead, R. J. (1972). On the test of independence between two sets of variates. *Ann. Math. Stat.*, **43**, 1491 - 1497.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, New York.

Muller, K. E. and Peterson, B. L. (1984). Practical methods for computing power in testing the multivariate general linear hypothesis. *Comput. Statist. Data Anal.*, **2**, 143 - 158.

Nagao, H. (1972). Non-null distributions of the likelihood ratio criteria for independence and equality of mean vectors and covariance matrices. *Ann. Inst. Stat. Math.*, **24**, 67 - 79.

Nagao, H. (1973a). On some test criteria for covariance matrix. *Ann. Statist.*, **1**, 700 - 709.

Nagao, H. (1973b). Nonnull distributions of two test criteria for independence under local alternatives. *J. Mult. Anal.*, **3**, 435 - 444.

Nandi, S. B. (1977). The exact null distribution of Wilks' criterion. *Sankhya*, Ser. B, **39**, 307 - 315.

Neudecker, H. (1986). The dispersion matrix of $vecX'AX$, $A' = A$, when $X'X$ is a Wishart matrix. *Pacif. Stat. Congress*, I. S. Francis, B. F. J. Manly and F. C. Lam (eds.), Elsevier Sci. Publ.

Pagès, J. P., Cailliez, F., Escoufier, Y. (1979). Analyse factorielle: un peu d'histoire et de géométrie. *Rev. Stat. Appl.*, **27**, 5 - 28.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, **2**, 559 - 572.

Pillai, K. C. S. and Gupta, A. K. (1969). On the exact distribution of Wilks's criterion. *Biometrika*, **56**, 109 - 118.

Rao, C. R. (1951). An asymptotic expansion of the distribution of Wilks' criterion. *Bull. Inst. Intern. Stat.*, **33**, 177 - 180.

Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and Its Applications.* John Wiley & Sons, New York.

Salle, P. de La (1979). Analyse conjointe de plusieurs tableaux. *Stat. Anal. Donn.*, **1**, 13 - 30.

Schatzoff, M. (1966). Exact distribution of Wilks's likelihood ratio criterion. *Biometrika*, **53**, 347 - 358.

Srivastava, M. S. and Yau, W. K. (1989). Saddlepoint method for obtaining tail probability of Wilks' likelihood ratio test. *J. Mult. Anal.*, **31**, 117 - 126.

Sugiura, N. (1969). Asymptotic expansions of the distributions of the likelihood ratio criteria for covariance matrix. *Ann. Math. Stat.*, **40**, 2051 - 2063.

Sugiura, N. and Fujikoshi, Y. (1969). Asymptotic expansions of the non-null distributions of the likelihood ratio criteria for multivariate linear hypothesis and independence. *Ann. Math. Stat.*, **40**, 942 - 952.

Wilks, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrika*, **3**, 309 - 326.

Young, M. R. (1991). *Estimating Optimal Transformations for Multivariate Data.* Ph.D. Thesis, Univ. of Michigan, Dep. of Biostatistics, Ann Arbor, Michigan.

Corrections

Page	line	where it reads	should read
78	2^{nd} line after expression (4.7)	V^*_{kk} (both occurrences)	$V^*_{kk.k}$
109	1^{st} expression in section 4.9	$(p_k + \ldots + p_m)$	$(p_{k+1} + \ldots + p_m)$
109	2^{nd} expression in section 4.9	$\frac{(1-l)^2}{g}$	$\frac{(1-l)^2}{g} - (p_{k+1} + \ldots + p_{i-1} + p_{i+1} + \ldots + p_m)$
118	last line of vi)	$43k^{-2}$	$3k^{-2}$
118	last line of vi')	there are some misprints	
119	last line of expression preceeding (5.3)	$\frac{11}{12}\theta^{-4}(T-\theta)^4$	$\frac{11}{12}\theta^{-4}\mathcal{E}(T-\theta)^4$
120	2^{nd} line of expression for $Var(g(T))$	$+R_4 - g(\theta) + k$	$-g(\theta) + k^{-1}$
120	3^{rd} line of expression for $Var(g(T))$	$-\frac{1}{3}(k+4)(k+2)k^{-1}$	$-\frac{1}{3}(k+4)(k+2)k^{-2} + (k+2)k^{-1}$
121	2^{nd} line	$c_2 = 43$	$c_2 = 3$
125	line after expression (5.13)	and	and, given the independence relations under H_0,
126	last expression in the proof	$Var(log\,\Lambda^*) = \sum_{j=0}^{5} c_j d_j$	$Var(log\,\Lambda^*) = \sum_{j=1}^{5} c_j e_j$
127	first expression	$c^{*\prime} = \sum_{j=0}^{3} c_j^{\prime} e_j$	$c^{*\prime} = \sum_{j=1}^{3} c_j^{\prime} e_j$ (where the $c_j^{\prime}$'s are defined as in Corollary 1)