

LINEAR LIBRARY

C01 0068 5136



UNIVERSITY OF CAPE TOWN

DEPARTMENT OF MATHEMATICAL STATISTICS

THE GENERALIZED PARTIAL CORRELATION MATRIX

by

J.M. Juritz

A thesis prepared under the supervision of Professor C.G. Troskie
in fulfilment of the requirements for the degree of Master of
Science in Mathematical Statistics

Copyright by the University of Cape Town
1969

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

A C K N O W L E D G E M E N T S

My warmest thanks to Professor C.G. Troskie for his guidance in the preparation of this thesis. His friendly encouragement not only made it possible, but also enjoyable. I wish also to express my appreciation and thanks to Mrs. M.I. Cousins for her skilful typing of the manuscript. Finally I must thank my husband and family for their cheerful encouragement in spite of the demands made on my time by the preparation of the thesis. It is to them that this thesis is affectionately dedicated.

C O N T E N T S

INTRODUCTION

CHAPTER I

1.1 Introduction	1
1.2 Conditional distributions	3
1.3 The Maximum Likelihood estimates of $\Sigma_{.3}$ and $\Sigma_{11.23}$	10
1.4 The Sampling distributions of $A_{.3}$ and $A_{11.23}$	12

CHAPTER II

11.1 Introduction	20
11.2 The Partial Multiple Correlation Coefficient	22
11.3 The Partial Multiple Correlation Coefficient in the Sample	33
11.4 The Partial Trace Correlation Coefficient	44
11.5 Partial Canonical Correlation Coefficients	47

CHAPTER III

111.1 The Square root of a Matrix	51
111.2 The Generalised Partial Correlation Matric in the Population	52
111.3 The Generalised Partial Correlation Matric in the Sample	54

CHAPTER IV

IV.1 Introduction	60
IV.2 An application of the Partial Multiple Correlation Coefficient	61
IV.3 An application of the Generalised Partial Correlation Matrix	68

CONCLUDING REMARKS

77

APPENDIX. The Computation of the Generalised Partial Correlation Matrix

79

BIBLIOGRAPHY

81

INTRODUCTION

This thesis is principally concerned with measures of association in conditional multivariate normal distributions. The multivariate normal distribution has the remarkable property that all the conditional distributions are again normal. It is because of this fundamental property of the normal distribution, that measures of partial association, between normal variables, such as the partial correlation coefficient, share the same distribution as their unconditional counterparts, the only difference being in those parameters that depend on the sample size.

In Chapter I we discuss the conditional distribution of a p dimensional normal variable which has been partitioned into three sets of components. In addition we also discuss distribution of the Wishart matrix when one and two of the sets are fixed.

The Wishart matrix plays an important role in the distribution theory of correlation coefficients, since this matrix carries all the sample information on the interdependences between the variables.

In Chapter II, we discuss the different measures of partial association that have been defined, the partial correlation coefficient has not been discussed because of its familiarity and excellent treatments of it are readily available in the literature. The Partial Multiple correlation coefficient is discussed in considerable detail. The partial multiple correlation coefficient, which measures the association between a single conditional variable and a linear combination of conditional variables has received scant attention in the literature, which is a great pity, because

it is potentially an extremely useful concept. In the literature the partial multiple correlation is either mentioned as a special case of more general measures of partial association, like the partial canonical correlations, or else it is used in important applications without its actual nature being recognised. Other measures that are discussed in Chapter II are the partial trace correlation coefficient and the partial canonical correlations.

In Chapter III we consider the Generalised Partial Correlation Matrix which has the remarkable property that it embodies all the measures of partial association as special cases. Its distribution in the central case will also be derived in this Chapter.

In Chapter IV we consider some practical applications of the partial multiple correlation coefficient and the Generalised Partial Correlation Matrix. The advent of digital computers has transformed multivariate analysis and has relieved the statistician of the tedium of long calculations. He is now free to spend the time saved in computation, debugging an intricate programme at least with the knowledge that if he ever has to compute the particular statistic again the answer will be ready in a few seconds. Now that elaborate statistical techniques are so readily available, it is perhaps more essential than ever before to remember the remark of A.N. Whitehead that is quoted by M.J. Moroney in Facts from Figures "There is no more common error than to assume that, because prolonged and accurate mathematical calculations have been made, the application of the result to some fact of nature is absolutely certain".

CHAPTER I

1.1 Introduction

In this chapter we give the basic results that will be used in subsequent chapters.

Let us suppose that we have a p dimensional random variable X which has a multivariate Normal distribution with parameters μ and Σ where

$$\mu = E(X) = \begin{matrix} p \times 1 \\ \left| \begin{array}{c} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \mu_p \end{array} \right| \end{matrix}$$

$$\Sigma = E(X-\mu)(X-\mu)' = \begin{matrix} \left| \begin{array}{ccc} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \cdot & & & \\ \cdot & & & \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{array} \right| \end{matrix}$$

where σ_{ii} is the variance of the i th component of X

σ_{ij} is the covariance of the i th and j th component of X .

The density function of X will be denoted by $n(x|\mu, \Sigma)$ and the corresponding distribution function as $N(\mu, \Sigma)$. If we wish to draw special attention to the dimension of the distribution we shall indicate this by a subscript on N , i.e. $N_p(\mu, \Sigma)$ represents a p dimensional normal distribution.

We now partition X into three sets of components.

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \\ X^{(3)} \end{pmatrix} \quad \text{where} \quad \begin{array}{l} X^{(1)} \text{ has } q \text{ components} \\ X^{(2)} \text{ has } r \text{ components} \\ X^{(3)} \text{ has } s \text{ components} \end{array}$$

where $q+r+s = p$.

The vector of means, μ , and the covariance matrix, Σ , are partitioned accordingly,

$$\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \\ \mu^{(3)} \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix}$$

The dimensions of the submatrices, Σ_{ij} , can be represented schematically in the following "matrix", the dimension occupying the same position as the corresponding submatrix.

Dimensions of the Partitions of Σ

$$\begin{pmatrix} q \times q & q \times r & q \times s \\ r \times q & r \times r & r \times s \\ s \times q & s \times r & s \times s \end{pmatrix}$$

We also list the dimensions in the following table for reference purposes.

<u>SubMatrix</u>	<u>Dimension</u>
Σ_{11}	$q \times q$
Σ_{12}	$q \times r$
Σ_{13}	$q \times s$
Σ_{21}	$r \times q$
Σ_{22}	$r \times r$
Σ_{23}	$r \times s$
Σ_{31}	$s \times q$
Σ_{32}	$s \times r$
Σ_{33}	$s \times s$

In subsequent chapters our interest will be centred on the conditional

distribution of two sets of components given the third set or on the distribution of one set of components given the other two. As the numbering of the individual components, and the choice of sets of components is quite arbitrary and in practical applications depends only on the problem in hand, we may, without loss of generality, assume that it is always the third set $X^{(3)}$ that is fixed and $X^{(2)}$ and $X^{(1)}$ in the case where two sets are fixed. Since the conditional variables mentioned above are of fundamental importance in this thesis we shall derive their distribution in detail in the next section.

1.2 Conditional Distributions.

Let $\left| \begin{array}{c} X^{(1)} \\ X^{(2)} \end{array} \middle| x^{(3)} \right|$ denote the conditional random variable $\left| \begin{array}{c} X^{(1)} \\ X^{(2)} \end{array} \right|$ given that $X^{(3)} = x^{(3)}$. We note that this variable has dimension $q+r$. To find the distribution of $\left| \begin{array}{c} X^{(1)} \\ X^{(2)} \end{array} \middle| x^{(3)} \right|$ we make a nonsingular transformation of our original variable $X = \left| \begin{array}{c} X^{(1)} \\ X^{(2)} \\ X^{(3)} \end{array} \right|$ into the variable $Y = \left| \begin{array}{c} Y^{(1)} \\ Y^{(2)} \\ Y^{(3)} \end{array} \right|$

$$\begin{aligned} \text{where} \quad Y^{(1)} &= X^{(1)} + M_1 X^{(3)} \\ Y^{(2)} &= X^{(2)} + M_2 X^{(3)} \\ Y^{(3)} &= X^{(3)} \end{aligned} \tag{1.2.1}$$

and M_1 and M_2 are chosen so that the components of $Y^{(1)}$ and $Y^{(2)}$ are uncorrelated with $Y^{(3)}$.

For this condition to hold we require that

$$E \left| \begin{array}{c} Y^{(1)} - EY^{(1)} \\ Y^{(2)} - EY^{(2)} \end{array} \middle| \left| Y^{(3)} - EY^{(3)} \right| = 0 \tag{1.2.2}$$

where 0 represents a $q+r$ by $q+r$ matrix with all elements zero.

Substituting for $Y^{(1)}$ in terms of $X^{(1)}$ and M_1 in (1.2.2) the condition becomes

$$E \left| \begin{array}{l} (X^{(1)} - \mu^{(1)}) + M_1(X^{(3)} - \mu^{(3)}) \\ (X^{(2)} - \mu^{(2)}) + M_2(X^{(3)} - \mu^{(3)}) \end{array} \right| |X^{(3)} - \mu^{(3)}|^2 = 0$$

Multiplying and taking expected values we obtain the matrix equations

$$\Sigma_{13} + M_1 \Sigma_{33} = 0$$

$$\Sigma_{23} + M_2 \Sigma_{33} = 0$$

From which it follows that

$$\begin{aligned} M_1 &= -\Sigma_{13} \Sigma_{33}^{-1} \\ M_2 &= -\Sigma_{23} \Sigma_{33}^{-1} \end{aligned} \quad (1.2.3)$$

where we assume that Σ_{33} is nonsingular so that the inverse exists. Substituting (1.2.3) in (1.2.1) we obtain

$$\begin{aligned} Y^{(1)} &= X^{(1)} - \Sigma_{13} \Sigma_{33}^{-1} X^{(3)} \\ Y^{(2)} &= X^{(2)} - \Sigma_{23} \Sigma_{33}^{-1} X^{(3)} \\ Y^{(3)} &= X^{(3)} \end{aligned} \quad (1.2.4)$$

which can be written in matrix notation as

$$Y = \begin{bmatrix} Y^{(1)} \\ Y^{(2)} \\ Y^{(3)} \end{bmatrix} = \begin{bmatrix} I_q & 0 & -\Sigma_{13} \Sigma_{33}^{-1} \\ 0 & I_r & \Sigma_{23} \Sigma_{33}^{-1} \\ 0 & 0 & I_s \end{bmatrix} \begin{bmatrix} X^{(1)} \\ X^{(2)} \\ X^{(3)} \end{bmatrix} \quad (1.2.5)$$

where I_k is a $k \times k$ unit matrix.

Clearly this is a nonsingular transformation of X and therefore Y is also normally distributed. The mean and covariance of this distribution can be found using the fact that if X is distributed as $N(\mu, \Sigma)$ and C is a nonsingular transformation of X , then $Y = CX$ is distributed as $N(C\mu, C\Sigma C')$.

Taking C as the matrix in (1.2.4) we have that

$$E(Y) = E \begin{vmatrix} Y^{(1)} \\ Y^{(2)} \\ Y^{(3)} \end{vmatrix} = \begin{vmatrix} \mu^{(1)} - \Sigma_{13} \Sigma_{33}^{-1} \mu^{(3)} \\ \mu^{(2)} - \Sigma_{23} \Sigma_{33}^{-1} \mu^{(3)} \\ \mu^{(3)} \end{vmatrix} = \begin{vmatrix} \Psi^{(1)} \\ \Psi^{(2)} \\ \Psi^{(3)} \end{vmatrix} \quad (1.2.6)$$

Cov (YY') =

$$\begin{vmatrix} I_q & 0 & -\Sigma_{13} \Sigma_{33}^{-1} \\ 0 & I_r & -\Sigma_{23} \Sigma_{33}^{-1} \\ 0 & 0 & I_s \end{vmatrix} \begin{vmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{vmatrix} \begin{vmatrix} I_q & 0 & 0 \\ 0 & I_r & 0 \\ -\Sigma_{13} \Sigma_{33}^{-1} & \Sigma_{23} \Sigma_{33}^{-1} & I_s \end{vmatrix}$$

$$= \begin{vmatrix} \Sigma_{11} - \Sigma_{13} \Sigma_{33}^{-1} \Sigma_{31} & \Sigma_{12} - \Sigma_{13} \Sigma_{33}^{-1} \Sigma_{32} & 0 \\ \Sigma_{21} - \Sigma_{23} \Sigma_{33}^{-1} \Sigma_{32} & \Sigma_{22} - \Sigma_{23} \Sigma_{33}^{-1} \Sigma_{32} & 0 \\ 0 & 0 & \Sigma_{33} \end{vmatrix}$$

$$= \begin{vmatrix} \Sigma_{11 \cdot 3} & \Sigma_{12 \cdot 3} & 0 \\ \Sigma_{21 \cdot 3} & \Sigma_{22 \cdot 3} & 0 \\ 0 & 0 & \Sigma_{33} \end{vmatrix}$$

where $\Sigma_{ij \cdot 3} = \Sigma_{ij} - \Sigma_{i3} \Sigma_{33}^{-1} \Sigma_{3j}$ $i, j = 1, 2$.

Since $\begin{vmatrix} Y^{(1)} \\ Y^{(2)} \end{vmatrix}$ are independent of $Y^{(3)}$ we can write the joint density

function of these variables as

$$f|y^{(1)}, y^{(2)}, y^{(3)}| = n \begin{vmatrix} y^{(1)} \\ y^{(2)} \\ y^{(3)} \end{vmatrix} \begin{vmatrix} \Psi^{(1)} \\ \Psi^{(2)} \end{vmatrix} \Sigma_{\cdot 3} \begin{vmatrix} y^{(3)} \\ \Psi^{(3)} \end{vmatrix} \Sigma_{33}$$

$$\text{where } \Sigma_{\cdot 3} = \begin{vmatrix} \Sigma_{11 \cdot 3} & \Sigma_{12 \cdot 3} \\ \Sigma_{21 \cdot 3} & \Sigma_{22 \cdot 3} \end{vmatrix} \quad (1.2.7)$$

The joint density function of $\begin{vmatrix} X^{(1)} \\ X^{(2)} \end{vmatrix}$ and $X^{(3)}$ can be found by substituting

for the $Y^{(i)}$ in terms of the $X^{(i)}$ using equations (1.2.4) and multiplying by the Jacobian which is one in this case

$$f|x^{(1)}x^{(2)}x^{(3)}| = n \left| \begin{array}{c} x^{(1)} - \Sigma_{13}\Sigma_{33}^{-1}x^{(3)} \\ x^{(2)} - \Sigma_{23}\Sigma_{33}^{-1}x^{(3)} \end{array} \right| \left| \begin{array}{c} \psi^{(2)}, \Sigma_{\cdot 3} \\ \psi^{(3)} \end{array} \right| n \left| x^{(3)} \right| \left| \begin{array}{c} \psi^{(3)}, \Sigma_{33} \end{array} \right| \quad (1.2.8)$$

where $\psi^{(i)} = E(Y^{(i)})$ as defined in (1.2.6).

The conditional density function of $\left| \begin{array}{c} X^{(1)} \\ X^{(2)} \end{array} \middle| x^{(3)} \right|$ is the quotient of

(1.2.8) and the marginal density function of $X^{(3)}$ at the point $x^{(3)}$.

This marginal density function of $X^{(3)}$ is the last factor of (1.2.8).

Thus we have, writing out the density function in full

$$f \left| \begin{array}{c} x^{(1)} \\ x^{(2)} \end{array} \middle| x^{(3)} \right| \quad (1.2.9) \\ = \frac{|\Sigma_{\cdot 3}|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}(q+r)}} \exp -\frac{1}{2} \left| \begin{array}{c} x^{(1)} - \mu^{(1)} - \Sigma_{13}\Sigma_{33}^{-1}(x^{(3)} - \mu^{(3)}) \\ x^{(2)} - \mu^{(2)} - \Sigma_{23}\Sigma_{33}^{-1}(x^{(3)} - \mu^{(3)}) \end{array} \right| \left| \begin{array}{c} \Sigma_{\cdot 3}^{-1} \\ \Sigma_{33}^{-1} \end{array} \right| \left| \begin{array}{c} x^{(1)} - \mu^{(1)} - \Sigma_{13}\Sigma_{33}^{-1}(x^{(3)} - \mu^{(3)}) \\ x^{(2)} - \mu^{(2)} - \Sigma_{23}\Sigma_{33}^{-1}(x^{(3)} - \mu^{(3)}) \end{array} \right|$$

Thus $\left| \begin{array}{c} X^{(1)} \\ X^{(2)} \end{array} \middle| x^{(3)} \right|$ is a $q+r$ dimensional normal variable with mean

$$E \left| \begin{array}{c} X^{(1)} \\ X^{(2)} \end{array} \middle| x^{(3)} \right| = \left| \begin{array}{c} \mu^{(1)} + \Sigma_{13}\Sigma_{33}^{-1}(x^{(3)} - \mu^{(3)}) \\ \mu^{(2)} + \Sigma_{23}\Sigma_{33}^{-1}(x^{(3)} - \mu^{(3)}) \end{array} \right| = \left| \begin{array}{c} \omega^{(1)}(x^{(3)}) \\ \omega^{(2)}(x^{(3)}) \end{array} \right| \quad (1.2.10)$$

and covariance matrix

$$\Sigma_{\cdot 3} = \left| \begin{array}{cc} \Sigma_{11\cdot 3} & \Sigma_{12\cdot 3} \\ \Sigma_{21\cdot 3} & \Sigma_{22\cdot 3} \end{array} \right| \quad (1.2.11)$$

The conditional distribution of $(X^{(1)}, X^{(2)}, X^{(3)})$ can be derived using the distribution of $\left| \begin{array}{c} X^{(1)} \\ X^{(2)} \end{array} \middle| x^{(3)} \right|$.

As above we choose a matrix M such that the two vectors

$$\begin{aligned} Y^{(1)} &= (X^{(1)} | X^{(3)}) + M(X^{(2)} | X^{(3)}) \\ Y^{(2)} &= (X^{(2)} | X^{(3)}) \end{aligned} \quad (1.2.12)$$

are uncorrelated.

Thus we must have

$$E(Y^{(1)} - E(Y^{(1)} | Y^{(2)} - E(Y^{(2)})))' = 0 \quad (1.2.13)$$

Substituting for $(X^{(1)} | X^{(3)})$ and $(X^{(2)} | X^{(3)})$ we find that condition

(1.2.13) reduces to

$$\Sigma_{12 \cdot 3} + M \Sigma_{22 \cdot 3} = 0$$

From which it follows that

$$M = -\Sigma_{12 \cdot 3} \Sigma_{22 \cdot 3}^{-1} \quad (1.2.14)$$

The transformation in matrix notation now becomes

$$\begin{vmatrix} Y^{(1)} \\ Y^{(2)} \end{vmatrix} = \begin{vmatrix} 1 - \Sigma_{12 \cdot 3} \Sigma_{22 \cdot 3}^{-1} & \\ 0 & 1 \end{vmatrix} \begin{vmatrix} X^{(1)} | X^{(3)} \\ X^{(2)} \end{vmatrix} \quad (1.2.15)$$

From which it follows that $\begin{vmatrix} Y^{(1)} \\ Y^{(2)} \end{vmatrix}$ is again normally distributed with mean

$$E \begin{vmatrix} Y^{(1)} \\ Y^{(2)} \end{vmatrix} = \begin{vmatrix} \omega^{(1)}(x^{(3)}) - \Sigma_{12 \cdot 3} \Sigma_{22 \cdot 3}^{-1} \omega^{(2)}(x^{(3)}) \\ \omega^{(2)}(x^{(3)}) \end{vmatrix} \quad (1.2.16)$$

where $\omega^{(1)}(x^{(3)})$ is defined in (1.2.10). The covariance matrix of

$$\begin{vmatrix} Y^{(1)} \\ Y^{(2)} \end{vmatrix} \text{ is}$$

$$\begin{vmatrix} \Sigma_{11 \cdot 3} - \Sigma_{12 \cdot 3} \Sigma_{22 \cdot 3}^{-1} \Sigma_{21 \cdot 3} & 0 \\ 0 & \Sigma_{22 \cdot 3} \end{vmatrix} = \begin{vmatrix} \Sigma_{11 \cdot 23} & 0 \\ 0 & \Sigma_{22 \cdot 3} \end{vmatrix} \quad (1.2.17)$$

Clearly $Y^{(1)}$ and $Y^{(2)}$ are independent and their joint density function will be

$$n(y^{(1)} | E(y^{(1)}), \Sigma_{11 \cdot 23}) n(y^{(2)} | \omega^{(2)}(x^{(3)}), \Sigma_{22 \cdot 3}) \quad (1.2.18)$$

The joint density of $\begin{vmatrix} X^{(1)} \\ X^{(2)} \end{vmatrix} | x^{(3)}$ can be found by using equations

(1.2.15) and (1.2.18) and multiplying by the Jacobian which is 1.

This expression will have the form

$$\begin{aligned}
 & f(x^{(1)}, x^{(2)} | x^{(3)}) \tag{1.2.19} \\
 &= \frac{1}{(2\pi)^{\frac{1}{2}q}} \frac{1}{\sqrt{|\Sigma_{11 \cdot 23}|}} \exp -\frac{1}{2} \left| (x^{(1)} - v^{(1)})' \Sigma_{11 \cdot 23}^{-1} (x^{(1)} - v^{(1)}) \right| \\
 & \frac{1}{(2\pi)^{\frac{1}{2}r}} \frac{1}{\sqrt{|\Sigma_{22 \cdot 3}|}} \exp -\frac{1}{2} \left| (x^{(2)} - w^{(2)}(x^{(3)}))' \Sigma_{22 \cdot 3}^{-1} (x^{(2)} - w^{(2)}(x^{(3)})) \right|
 \end{aligned}$$

where $v^{(1)}$ will be defined later.

As previously the conditional distribution of $X^{(1)}$ given $x^{(2)}$ and $x^{(3)}$ will be the quotient of (1.2.19) and the conditional distribution of $(X^{(2)} | x^{(3)})$ at the point $X^{(2)} = x^{(2)}$ (and $X^{(3)} = x^{(3)}$). This quotient is clearly the first factor of (1.2.19) and thus we have that

$$\begin{aligned}
 & (X^{(1)} | x^{(2)}, x^{(3)}) \text{ is also normally distributed with mean} \\
 & v^{(1)} = E(X^{(1)} | x^{(3)}) + \Sigma_{12 \cdot 3} \Sigma_{22 \cdot 3}^{-1} (x^{(2)} - E(X^{(2)} | x^{(3)})) \\
 & = \mu^{(1)} + \Sigma_{13} \Sigma_{33}^{-1} (x^{(3)} - \mu^{(3)}) + (\Sigma_{12} - \Sigma_{13} \Sigma_{33}^{-1} \Sigma_{32}) (\Sigma_{22} - \Sigma_{23} \Sigma_{33}^{-1} \Sigma_{32})^{-1} \\
 & \quad (x^{(2)} - \mu^{(2)} - \Sigma_{23} \Sigma_{33}^{-1} (x^{(3)} - \mu^{(3)})) \tag{1.2.20}
 \end{aligned}$$

and covariance matrix

$$\Sigma_{11 \cdot 23} = \Sigma_{11 \cdot 3} - \Sigma_{12 \cdot 3} \Sigma_{22 \cdot 3}^{-1} \Sigma_{21 \cdot 3} \tag{1.2.21}$$

Alternatively, we could find the distribution of $\begin{vmatrix} X^{(1)} \\ X^{(2)} \\ X^{(3)} \end{vmatrix}$ using the

joint density function of $X = \begin{vmatrix} X^{(1)} \\ X^{(2)} \\ X^{(3)} \end{vmatrix}$ expressed in a suitable form, and

dividing this density function by the joint density function of $\begin{vmatrix} X^{(2)} \\ X^{(3)} \end{vmatrix}$

at the point $\begin{vmatrix} x^{(2)} \\ x^{(3)} \end{vmatrix}$. We would then find that again $\begin{vmatrix} X^{(1)} & | & x^{(2)} \\ & & x^{(3)} \end{vmatrix}$ is a

q dimensional normal variate with mean

$$E \begin{vmatrix} X^{(1)} & | & x^{(2)} \\ & & x^{(3)} \end{vmatrix} = \mu^{(1)} + \Sigma_{12} \Sigma_{13} \begin{vmatrix} \Sigma_{22} \Sigma_{23} \\ \Sigma_{32} \Sigma_{33} \end{vmatrix}^{-1} \begin{vmatrix} x^{(2)} - \mu^{(2)} \\ x^{(3)} - \mu^{(3)} \end{vmatrix} \quad (1.2.22)$$

and covariance matrix

$$\Sigma_{11 \cdot 23} = \Sigma_{11} - \begin{vmatrix} \Sigma_{12} \Sigma_{13} \\ \Sigma_{22} \Sigma_{23} \\ \Sigma_{32} \Sigma_{33} \end{vmatrix}^{-1} \begin{vmatrix} \Sigma_{21} \\ \Sigma_{31} \end{vmatrix} \quad (1.2.23)$$

Clearly (1.2.20) and (1.2.22) are equivalent as are (1.2.21) and (1.2.23). We remark that the second method of finding the distribution

$\begin{vmatrix} X^{(1)} & | & x^{(2)} \\ & & x^{(3)} \end{vmatrix}$ is more direct, but we chose the first method because we

felt that it gives more insight into what is to follow. We also note that the conditional mean is a linear combination of the variables held fixed, but the conditional covariance matrices $\Sigma_{\cdot 3}$ and $\Sigma_{11 \cdot 23}$ do not involve the fixed values at all.

We shall state the results of this section in a theorem.

Theorem 1.2.1

If $X = \begin{vmatrix} X^{(1)} \\ X^{(2)} \\ X^{(3)} \end{vmatrix}$ is distributed as $N_p(\mu, \Sigma)$ then $\begin{vmatrix} X^{(1)} & | & x^{(3)} \\ X^{(2)} & & \end{vmatrix}$ is

distributed as $N_{q+r} \left(\begin{vmatrix} \omega^{(1)} \\ \omega^{(2)} \end{vmatrix}, \Sigma_{\cdot 3} \right)$ and $\begin{vmatrix} X^{(1)} & | & x^{(2)} \\ & & x^{(3)} \end{vmatrix}$ is distributed as

$N_q(v^{(1)}, \Sigma_{11 \cdot 23})$ where $\begin{vmatrix} \omega^{(1)} \\ \omega^{(2)} \end{vmatrix}$ is defined in (1.2.19) and $v^{(1)}$ in

(1.2.20) or (1.2.22).

1.3. The Maximum Likelihood Estimates of $\Sigma_{.3}$ and $\Sigma_{11.23}$

In general the parameters μ and Σ of the distribution of X are unknown and must be estimated from a random sample. Suppose a random sample, $X_{(1)}, X_{(2)}, \dots, X_{(N)}$ where $N > p$ is drawn from the population of X .

Let $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_{(i)}$ be the vector of sample means and

$$A = \sum_{i=1}^N (X_{(\alpha)} - \bar{X})(X_{(\alpha)} - \bar{X})'$$

be the Wishart matrix of sums of squares and cross-products.

Then it is shown in Anderson 1958 Chapter 3, that

$$\hat{\mu} = \bar{X} \quad \text{and} \quad \hat{\Sigma} = \frac{1}{N} A$$

estimates of μ and Σ .

We are interested in obtaining the maximum likelihood estimates of the parameters of the conditional distributions of $\begin{vmatrix} X^{(1)} & | & x^{(3)} \\ X^{(2)} & & \end{vmatrix}$ and

$$\begin{vmatrix} X^{(1)} & | & x^{(2)} \\ & & x^{(3)} \end{vmatrix}.$$

These estimates can be computed from the original sample $X_{(1)}, X_{(2)}, \dots, X_{(N)}$.

Since there is a one to one correspondence between the parameters of the conditional distributions and the parameters of X , maximum likelihood estimates of the parameters of both conditional distributions can be found by using appropriate functions of $\hat{\mu}$ and $\hat{\Sigma}$. (Anderson 1958 page 48).

Partitioning \bar{X} and A according to the partitioning of μ and Σ we have

$$\bar{X} = \begin{vmatrix} \bar{X}^{(1)} \\ \bar{X}^{(2)} \\ \bar{X}^{(3)} \end{vmatrix} \quad \text{and} \quad A = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix}$$

and $\bar{X}^{(i)}$ and $\frac{1}{N} A_{ij}$ are maximum likelihood estimates of $\mu^{(i)}$ and Σ_{ij} .

From the partitioned matrix A we can compute

$$\begin{aligned} A_{13} &= \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} - \begin{vmatrix} A_{13} \\ A_{23} \end{vmatrix} A_{33}^{-1} \begin{vmatrix} A_{31} & A_{32} \end{vmatrix} \\ &= \begin{vmatrix} A_{11 \cdot 3} & A_{12 \cdot 3} \\ A_{21 \cdot 3} & A_{22 \cdot 3} \end{vmatrix} \end{aligned} \quad (1.3.1)$$

$A_{\cdot 3}$ is then the Wishart matrix of the conditional distribution of $\begin{vmatrix} X^{(1)} \\ X^{(2)} \end{vmatrix} | X^{(3)}$.

Thus the maximum likelihood estimator of $\Sigma_{\cdot 3}$, the covariance matrix of

$\begin{vmatrix} X^{(1)} \\ X^{(2)} \end{vmatrix} | X^{(3)}$ is

$$\hat{\Sigma}_{\cdot 3} = \frac{1}{N} A_{\cdot 3} \quad (1.3.2)$$

Having obtained $A_{\cdot 3}$ we can use it to compute a maximum likelihood estimate of $\Sigma_{11 \cdot 23}$, the covariance matrix of $(X^{(1)} | X^{(2)}, X^{(3)})$. We compute

$$A_{11 \cdot 23} = A_{11 \cdot 3} - A_{12 \cdot 3} A_{22 \cdot 3}^{-1} A_{21 \cdot 3} \quad (1.3.3)$$

and

$$\hat{\Sigma}_{11 \cdot 23} = \frac{1}{N} A_{11 \cdot 23} \quad (1.3.4)$$

Alternatively $\hat{\Sigma}_{11 \cdot 23}$ could be computed directly from A using the matrix

$A_{11 \cdot 23}$ in the form

$$A_{11 \cdot 23} = A_{11} - \begin{vmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix}^{-1} \begin{vmatrix} A_{21} \\ A_{31} \end{vmatrix} \quad (1.3.5)$$

We note that the maximum likelihood estimators of $\Sigma_{\cdot 3}$ and $\Sigma_{11 \cdot 23}$ are biased.

It will be shown in Section 4 that $A_{\cdot 3}$ can be expressed as

$\sum_{\alpha=1}^{N-1-s} U_{(\alpha)} U_{(\alpha)}'$ where $U_{(\alpha)}$ are independent $N_{q+r}(0, \Sigma_{\cdot 3})$ variables. Hence

$$\begin{aligned} E(\hat{\Sigma}_{.3}) &= \frac{1}{N} E(A_{.3}) = \frac{1}{N} \sum_{\alpha=1}^{N-1-s} E(U_{(\alpha)} U_{(\alpha)}') \\ &= \frac{N-1-s}{N} \Sigma_{.3} \end{aligned}$$

Thus we can remove the bias by considering

$$S_{.3} = \frac{1}{N-s-1} A_{.3}$$

Similarly by Theorem 1.4.2 Section 4

$$E(\hat{\Sigma}_{11 \cdot 23}) = \frac{N-s-r-1}{N} \Sigma_{11 \cdot 23}$$

In this case we remove the bias by considering

$$S_{11 \cdot 23} = \frac{1}{N-s-r-1} A_{11 \cdot 23}$$

1.4 The sampling distributions of $A_{.3}$ and $A_{11 \cdot 23}$

The distribution of $A_{.3}$ and $A_{11 \cdot 23}$ will be of particular interest in subsequent chapters and their distributions will be given in this section.

It is well known that if $X_{(1)}, X_{(2)}, \dots, X_{(N)}$ is a random sample from

$N_p(\mu, \Sigma)$ with $N > p$ then

$$A = \sum_{\alpha=1}^N (X_{(\alpha)} - \bar{X})(X_{(\alpha)} - \bar{X})' \quad \text{can be written as}$$

$$A = \sum_{\alpha=1}^{N-1} Z_{(\alpha)} Z_{(\alpha)}' \quad \text{where } Z_{(\alpha)} \text{ are independent and each}$$

is distributed as $N(0, \Sigma)$. The random matrix A has a Wishart distribution with $n = N-1$ degrees of freedom and parameter Σ . The distribution function of A is denoted by $W(\Sigma, n)$.

Suppose now that $Z_{(\alpha)}$ $\alpha = 1, \dots, n$ is partitioned into

$$Z_{(\alpha)} = \begin{pmatrix} Z_{(\alpha)}^{(1)} \\ Z_{(\alpha)}^{(2)} \\ Z_{(\alpha)}^{(3)} \end{pmatrix} \quad \text{where } Z_{(\alpha)}^{(1)}, Z_{(\alpha)}^{(2)} \text{ and } Z_{(\alpha)}^{(3)} \text{ have } q, r \text{ and } s$$

Theorem 4.3.2. (Anderson 1958)

Suppose $Y_{(1)}, Y_{(2)}, \dots, Y_{(m)}$ are independent with $Y_{(\alpha)}$ distributed according to $N(\Gamma\omega_{\alpha}, \Phi)$ where $\omega_{(\alpha)}$ is an r component vector. Let $G = \sum_{\alpha} Y_{(\alpha)}\omega'_{(\alpha)}H^{-1}$ where $H = \sum_{\alpha}\omega_{(\alpha)}\omega'_{(\alpha)}$ and is non-singular. Then

$$\sum_{\alpha=1}^m Y_{(\alpha)}Y'_{(\alpha)} \sim GHG' \quad (1.4.1)$$

is distributed as $\sum_{\alpha=1}^{m-r} U_{\alpha}U'_{\alpha}$, where the U_{α} are independently distributed each according to $N(0, \Phi)$ and independently of G .

Corollary

If $\Gamma = 0$, the matrix GHG' is distributed as

$$\sum_{\alpha=m-r+1}^m U_{\alpha}U'_{\alpha} \quad (1.4.2)$$

where the $U_{(\alpha)}$ are independently distributed each according to $N(0, \Phi)$.

Thus under the conditions of the theorem the matrix (1.4.1) is distributed $W(\Phi, m-r)$ and the matrix GHG' (1.4.2) is $W(\Phi, r)$ if $\Gamma = 0$.

To find the density of A_3 we apply Theorem 4.3.2 with

$$\begin{aligned} \begin{vmatrix} Z_{(\alpha)}^{(1)} \\ Z_{(\alpha)}^{(2)} \\ Z_{(\alpha)}^{(3)} \end{vmatrix} &= Y_{(\alpha)} \\ Z_{(\alpha)}^{(3)} &= \omega_{(\alpha)} \\ N-1 &= m \\ s &= r \\ \beta &= \Gamma \\ \Sigma_3 &= \Phi \end{aligned} \quad (1.4.3)$$

$$\text{Then } \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = \sum_{\alpha} Y_{(\alpha)}Y'_{(\alpha)}$$

$$\begin{vmatrix} A_{13} \\ A_{23} \end{vmatrix} A_{33}^{-1} = G$$

$$A_{33} = H$$

and

$$A_{\cdot 3} = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} - \begin{vmatrix} A_{13} \\ A_{23} \end{vmatrix} A_{33}^{-1} A_{33} A_{33}^{-1} \begin{vmatrix} A_{31} \\ A_{32} \end{vmatrix} = \sum_{\alpha=1}^m Y_{(\alpha)} Y_{(\alpha)}' - GHG' \quad (1.4.3)$$

Thus it follows that conditional on $Z_{(\alpha)}^{(3)} = z_{(\alpha)}^{(3)}$ $A_{\cdot 3}$ is distributed as $\sum_{\alpha=1}^{N-s-1} U_{(\alpha)} U_{(\alpha)}'$ where U_{α} are independent, each distributed as $N(0, \Sigma_{\cdot 3})$.

i.e. conditional on $Z_{(\alpha)}^{(3)} = z_{(\alpha)}^{(3)}$, $A_{\cdot 3}$ is distributed as $W(\Sigma_{\cdot 3}, n-s)$, where $n = N-1$. But since this distribution does not depend on $\{Z_{(\alpha)}^{(3)}\}$ it is also the unconditional distribution of $A_{\cdot 3}$.

Thus we have

Theorem 1.4.1

$A_{\cdot 3}$ is distributed as $W(\Sigma_{\cdot 3}, n-s)$.

Using the fact that $A_{\cdot 3}$ is distributed as $\sum_{\alpha=1}^{n-s} U_{\alpha} U_{\alpha}'$ where the $\{U_{(\alpha)}\}$ are independent $N(0, \Sigma_{\cdot 3})$ variables, we can apply Anderson's Theorem 4.3.2 again to find the distribution of $A_{11 \cdot 23} = A_{11 \cdot 3} - A_{12 \cdot 3} A_{22 \cdot 3}^{-1} A_{21 \cdot 3}$

$$\text{Let } U_{(\alpha)} = \begin{vmatrix} U_{(\alpha)}^{(1)} \\ U_{(\alpha)}^{(2)} \end{vmatrix} \quad \alpha = 1, 2, \dots, n-s$$

where $U_{(\alpha)}^{(1)}$ has q components

$U_{(\alpha)}^{(2)}$ has r components, and the partitioning corresponds to

the partitioning of $\begin{vmatrix} Z_{(\alpha)}^{(1)} \\ Z_{(\alpha)}^{(2)} \end{vmatrix}$.

$$\text{Let } U_{(\alpha)}^{(2)} = u_{(\alpha)}^{(2)}, \quad \alpha = 1, 2, \dots, n-s.$$

Then, as above we have that the variables $(U_{(\alpha)}^{(1)} \mid u_{(\alpha)}^{(2)})$ $\alpha = 1, 2, \dots, n-s$ are independent and each is distributed as $N_q(\beta u_{(\alpha)}^{(2)}, \Sigma_{11 \cdot 23})$

$$\beta = \Sigma_{12 \cdot 3} \Sigma_{22 \cdot 3}^{-1}$$

$$\Sigma_{11 \cdot 23} = \Sigma_{11 \cdot 3} - \Sigma_{12 \cdot 3} \Sigma_{22 \cdot 3}^{-1} \Sigma_{21 \cdot 3} \quad .$$

Applying Theorem 4.3.2 again with

$$U_{(\alpha)}^{(1)} = Y_{(\alpha)}$$

$$u_{(\alpha)}^{(2)} = w_{(\alpha)}$$

$$n-s = m$$

(1.4.4)

$$r = r$$

$$\beta = \Gamma$$

$$\Sigma_{11 \cdot 23} = \Phi$$

$$A_{11 \cdot 23} = \Sigma Y_{(\alpha)} Y_{(\alpha)}'$$

$$A_{12 \cdot 3} A_{22 \cdot 3}^{-1} = G$$

$$A_{22 \cdot 3} = H \quad ,$$

we find that conditional on $U_{(\alpha)}^{(2)} = u_{(\alpha)}^{(2)}$ $\alpha = 1, 2, \dots, n-s$

$$A_{11 \cdot 23} = A_{11 \cdot 3} - (A_{12 \cdot 3} A_{22 \cdot 3}^{-1}) A_{22 \cdot 3} (A_{22 \cdot 3}^{-1} A_{21 \cdot 3})$$

is distributed as

$$\sum_{\alpha=1}^{n-s-r} V_{(\alpha)} V_{(\alpha)}'$$

where the $\{V_{(\alpha)}\}$ are independent, each distributed as $N_q(0, \Sigma_{11 \cdot 23})$.

Since this distribution does not depend on $\{u_{(\alpha)}^{(2)}\}$, it is also the unconditional distribution of $A_{11 \cdot 23}$. Thus we have

Theorem 1.4.2

$A_{11 \cdot 23}$ is distributed as $W(\Sigma_{11 \cdot 23}, n-s-r)$

Using the corollary to Theorem (4.3.2) we have that

$\beta = \Sigma_{12 \cdot 3} \Sigma_{22 \cdot 3}^{-1} = 0$, then

$$(A_{12 \cdot 3} A_{22 \cdot 3}^{-1}) A_{22 \cdot 3} (A_{22 \cdot 3}^{-1} A_{21 \cdot 3})$$

is distributed as

$$\sum_{\alpha=n-s-r+1}^{n-s} V_{(\alpha)} V'_{(\alpha)}$$

where the $\{V_{(\alpha)}\}$ are independent $N_q(0, \Sigma_{11 \cdot 23})$ variables. Thus we have

Theorem 1.4.3.

If $\Sigma_{12 \cdot 3} \Sigma_{22 \cdot 3}^{-1} = 0$, then $A_{12 \cdot 3} A_{22 \cdot 3}^{-1} A_{21 \cdot 3}$ is distributed as

$W(\Sigma_{11 \cdot 23}, r)$ independently of $A_{11 \cdot 23}$.

We note that the distribution of $A_{11 \cdot 23}$ could also be found directly by applying Theorem 4.3.2 to the alternative expression for $A_{11 \cdot 23}$ given by (1.3.5).

We would then have the following correspondences

$$\begin{aligned} Z_{(\alpha)}^{(1)} &= Y_{(\alpha)} \\ \begin{vmatrix} Z_{(\alpha)}^{(2)} \\ Z_{(\alpha)}^{(3)} \\ Z_{(\alpha)} \end{vmatrix} &= \omega_{(\alpha)} \\ N-1 &= m \\ r+s &= r \\ \Sigma_{11 \cdot 23} &= \Phi \\ \begin{vmatrix} \Sigma_{12} \Sigma_{13} \\ \Sigma_{22} \Sigma_{23} \\ \Sigma_{32} \Sigma_{33} \end{vmatrix}^{-1} &= \beta \end{aligned} \tag{1.4.5}$$

$$A_{11} = \sum Y_{(\alpha)} Y_{(\alpha)}'$$

$$\begin{vmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix}^{-1} \begin{vmatrix} A_{31} \\ A_{32} \end{vmatrix} = GHG'$$

and as above we find that $A_{11.23}$ is distributed as $W(\Sigma_{11.23}, n-s-r)$ and if in (1.4.5) $\beta = 0$ then GHG' is $W(\Sigma_{11.23}, n-s-r)$.

We note that the sampling distributions of the partial covariance matrices $A_{.3}$ and $A_{11.23}$ are the same as the distribution of an unconditional sample covariance matrix. The distribution of $A_{.3}$ based on a sample of size N from $N_p(\mu, \Sigma)$ is the same as the distribution of a sample covariance matrix based on a sample of size $N-s$ from a $N_{p-s}(\omega, \Sigma_{.3})$ population. Similarly $A_{11.23}$ is distributed as a sample covariance matrix based on a sample of size $N-s-r$ from a $N_{p-s-r}(\nu, \Sigma_{11.23})$

We would remark here that the condition $\beta = 0$ implies that $\Sigma_{12.3} = 0$

where 0 is a matrix with all its elements zero. Under normal theory this implies that $(X^{(1)} | X^{(3)})$ and $(X^{(2)} | X^{(3)})$ conditionally independent. It does not necessarily imply complete independence of $X^{(1)}$ and $X^{(2)}$.

Suppose $\Sigma_{12} \neq 0$, $\Sigma_{13} \neq 0$ and $\Sigma_{23} \neq 0$. It is still possible for $X^{(1)}$ and $X^{(2)}$ to be conditionally independent. Consider

$\Sigma_{12.3} = \Sigma_{12} - \Sigma_{13} \Sigma_{33}^{-1} \Sigma_{32}$. If $\Sigma_{12} = \Sigma_{13} \Sigma_{33}^{-1} \Sigma_{32}$, then $\Sigma_{12.3} = 0$ and $X^{(1)}$ and $X^{(2)}$ are conditionally independent but not completely independent.

Thus we cannot infer complete independence from conditional independence.

We can illustrate this situation with an example when $p = 3$. Suppose that the covariance matrix of a normally distributed random vector X is

$$\begin{vmatrix} 1 & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{vmatrix}$$

$$\begin{aligned} \text{Then } \Sigma_{12 \cdot 3} &= \sigma_{12} - \frac{\sigma_{13}\sigma_{32}}{\sigma_{33}} \\ &= \frac{1}{4} - \frac{1}{2} \cdot \frac{1}{2} = 0 \end{aligned}$$

and $X^{(1)}$ and $X^{(2)}$ are conditionally independent but not completely independent.

Suppose now that $X^{(1)}$ and $X^{(2)}$ are completely independent ($\Sigma_{12} = 0$). They will only be conditionally independent if at least one of them is independent of $X^{(3)}$, since $\Sigma_{12 \cdot 3} = 0$ if $\Sigma_{12} = 0$ and either $\Sigma_{13} = 0$ or $\Sigma_{23} = 0$ (or both Σ_{13} and Σ_{23} are "zero"). Thus if on the basis of statistical test we decide that $X^{(1)}$ and $X^{(2)}$ are conditionally independent, we have one of the following possibilities.

- (1) $X^{(1)}$ and $X^{(2)}$ are not independent but $\Sigma_{12} = \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32}$
- (2) $X^{(1)}$ and $X^{(2)}$ are independent and at least one of them is independent of $X^{(3)}$.
- (3) $X^{(1)}$ and $X^{(2)}$ are completely independent and both are independent of $X^{(3)}$.

It is also possible for $X^{(1)}$ and $X^{(2)}$ to be completely independent but not conditionally independent. This would occur if both were dependent on $X^{(3)}$. In this case $\Sigma_{12 \cdot 3} = \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{32}$. For example if the covariance matrix of X was

$$\begin{vmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{vmatrix}$$

Then $\sigma_{12 \cdot 3} = 0 - \frac{1}{2} \cdot \frac{1}{2} = -\frac{1}{4}$. $X^{(1)}$ and $X^{(2)}$ are uncorrelated and therefore independent under normal theory, but they are not conditionally independent.

The above discussion illustrates the fact that caution must be observed when statistical inference is made using measures of partial association.

CHAPTER II

11.1 Introduction.

Let X be a p dimensional random vector which is distributed as $N_p(\mu, \Sigma)$ and let X be partitioned into three sets of components

$\left| \begin{array}{l} X^{(1)} \\ X^{(2)} \\ X^{(3)} \end{array} \right|$ with dimensions q, r and s respectively, where $q+r+s = p$.

Measures of the relationships that exist between two of the sets of variables when the effect of the third set has been eliminated, are of considerable practical and theoretical interest. Frequently in an experiment, the battery of observations made on each experimental unit may fall naturally into three sets. For example an experiment may be conducted in which sets of physiological, psychological and physical observations are made on each individual, and the experimenter may wish to investigate the relationships that exist between the physiological and the psychological variables, when the effect of the physical variables has been eliminated. An ideal method of investigating these relationships would be to draw samples for fixed values of the physical variables, but this is often impossible in practice because of the complexity of the situation. The experimenter must therefore try to extract the information from his original sample. To do this he must investigate the relationships that exist in the conditional distribution of two of the sets given the third set.

The most familiar statistical measure of this type is of course the partial correlation coefficient, which measures the correlation between two components of a random vector when the effects of some or all of the other

components have been removed. Other measures which have been defined are

- (i) Partial Multiple Correlation Coefficients
(Roy and Whittlesey (1952), Hooper (1962)).
- (ii) Canonical partial correlation coefficients
(Roy and Whittlesey (1952)).
- (iii) Partial trace correlation coefficient
(Hooper (1962)).
- (iv) The Generalised partial multiple correlation matrix
(Troskie (1969)).

We shall discuss the first three of these measures in this chapter.

The generalised partial multiple correlation matrix, (iv), will be discussed in Chapter III, where it will be shown that (i), (ii) and (iii) are special cases of (iv).

Another topic that we shall discuss briefly in this Chapter is that of using a number of "independent" variables to predict the value of a dependent variable. In this type of problem the experimenter is not interested in the correlations that exist between the dependent variables and the independent variable, but needs a measure of how adequately the independent variables predict the dependent variable, and also some means of deciding which subset of the dependent variables give the best prediction. It is often found in a predication situation, that there are a large number of variables that might be considered for the predication equation (we assume is linear). However prediction equations using all the variables often give worse predictions than equations which use only a subset of the available variables. The choice of the best subset of variables to use in a prediction situation has received much attention in the literature, and we shall not discuss it here except to show that the partial multiple corre-

function^f makes the variance of the residual $(X_i^* - B(X^{(2)} | X^{(3)}))$ a minimum.

lation coefficient can be used to determine whether or not a certain subset of variables should be used in a prediction equation. A detailed discussion of these problems can be found in Linhart (1958,1960) and Browne (1969).

11.2 The Partial Multiple Correlation Coefficient.

The concept of multiple correlation between two sets of variables is familiar; it is the maximum correlation that exists between a single variable from the first set, X_1 say, and a linear combination of the variables from the second set. The linear combination with the greatest correlation is called the Regression function. Multiple correlation is an extremely useful concept since it often happens that although the correlations between X_1 and the individual variables from the second set may not be large, there can exist a very high correlation between X_1 and a linear combination of the variables from the second set. (Keeping, Example 2 page 258). When the concept of multiple correlation is applied to conditional distributions a number of interesting and useful results follow.

Suppose we have two sets of conditional variables $(X^{(1)}|X^{(3)})$ and $(X^{(2)}|X^{(3)})$ with q and r components respectively. We select one variable $X_i^* = (X_i^{(1)}|X^{(3)})$ from the set $(X^{(1)}|X^{(3)})$ where the asterisk is used to distinguish the conditioned variable $(X_i^{(1)}|X^{(3)})$ from the i th component of the original vector X . We now ask, what linear combination of the set $(X^{(2)}|X^{(3)})$ has maximum correlation with X_i^* ? It will be shown that this linear combination is the "partial regression function" and the resulting correlation coefficient is the Partial Multiple correlation coefficient. It will also be shown that the "partial regression function" makes the variance of the residual $(X_i^* - \beta(X^{(2)}|X^{(3)}))$ a minimum.

These results do not depend on the underlying distribution of X ($p \times 1$), the only requirement being that the determinant of the covariance matrix, $\|\Sigma\|$, must be finite and greater than zero. Nevertheless since the distribution of the sample estimate of the partial multiple correlation coefficient, depends on the assumption that the variables concerned are drawn from a normal population, we shall not drop the normality assumption in this thesis. We shall now define the partial multiple correlation coefficient in the population and demonstrate some of its properties.

Theorem 2.1.1.

Let $X = \begin{vmatrix} X^{(1)} \\ X^{(2)} \\ X^{(3)} \end{vmatrix}$ with q , r and s components respectively. Let

X be distributed as $N_p(0, \Sigma)$. Let $(X^{(1)} | x^{(3)})$ and $(X^{(2)} | x^{(3)})$ denote the conditional variables $X^{(1)}$ given $X^{(3)} = x^{(3)}$ and $X^{(2)}$ given $X^{(3)} = x^{(3)}$ respectively.

Let $\Sigma_{.3} = \begin{vmatrix} \Sigma_{11.3} & \Sigma_{12.3} \\ \Sigma_{21.3} & \Sigma_{22.3} \end{vmatrix}$ denote the covariance matrix of $\begin{vmatrix} X^{(1)} \\ X^{(2)} \end{vmatrix} | x^{(3)}$

Let X_i^* be the i th variable of $(X^{(1)} | x^{(3)})$.

Let $\sigma_{(i.3)}$ be the i th row of $\Sigma_{12.3}$.

Let $\sigma_{(ii.3)}$ be the variance of X_i^* .

Then, the linear combination of $(X^{(2)} | x^{(3)})$ that has maximum correlation with X_i^* is

$$\beta(X^{(2)} | x^{(3)})$$

where β is the $(1 \times r)$ vector

$$\beta = \sigma_{(i.3)} \Sigma_{22.3}^{-1}$$

The value of this maximum correlation is

$$\sqrt{\frac{\sigma_{(i.3)} \Sigma_{22.3}^{-1} \sigma'_{(i.3)}}{\sigma_{(ii.3)}}$$

Proof. To find the linear combination with maximum correlation we shall use a Lagrange multiplier λ .

Let $(X^{(2)}|X^{(3)})$ be an arbitrary linear combination of $(X^{(2)}|X^{(3)})$ where α is an $(1 \times r)$ vector. The correlation between X_i^* and $\alpha(X^{(2)}|X^{(3)})$ is

$$\Psi(\alpha) = \frac{\text{Cov}(\alpha(X^{(2)}|X^{(3)}), X_i^*)}{\sqrt{\text{Var}(X_i^*) \text{Var}(\alpha(X^{(2)}|X^{(3)}))}} \quad (2.2.1)$$

The values of the variances and covariances can be obtained from Σ_3

$$\text{Thus} \quad \Psi(\alpha) = \frac{\alpha \sigma'_{(i.3)}}{\sqrt{(\sigma_{ii.3}) (\alpha \Sigma_{22.3} \alpha')}} \quad (2.2.2)$$

Since correlation coefficients are invariant with respect to change of origin and scale, we assume that the variables have been scaled so that

$$\sigma_{ii.3} = 1 \quad \text{and} \quad \alpha \Sigma_{22.3} \alpha' = 1 \quad (2.2.3)$$

To maximise $\Psi(\alpha)$ subject to (2.2.3) we consider the function

$$F(\alpha) = \alpha \sigma'_{(i.3)} - \frac{1}{2} \lambda (\alpha \Sigma_{22.3} \alpha' - 1)$$

where λ is a scalar undetermined multiplier

$$\frac{\partial F}{\partial \alpha} = \sigma'_{(i.3)} - \lambda \Sigma_{22.3} \alpha'$$

For a maximum this expression must be zero. Hence it follows that

$$\alpha' = \frac{1}{\lambda} \Sigma_{22.3}^{-1} \sigma'_{(i.3)}$$

and from the side condition, $\lambda = 1$.

Thus the linear combination of $X^{(2)}|X^{(3)}$ with maximum correlation with X_i^* is

$$\beta(X^{(2)}|X^{(3)}) = \sigma_{(i \cdot 3)} \Sigma_{22 \cdot 3}^{-1} (X^{(2)}|X^{(3)}) \quad (2.2.4)$$

Substituting $\alpha = \beta = \sigma_{(i \cdot 3)} \Sigma_{22 \cdot 3}^{-1}$ in (2.2.2) we obtain the value of the maximum which is given by

$$\begin{aligned} \Psi(\beta) &= \frac{\sigma_{(i \cdot 3)} \Sigma_{22 \cdot 3}^{-1} \sigma_{(1 \cdot 3)}}{\sqrt{(\sigma_{(1 \cdot 3)}) (\sigma_{(i \cdot 3)} \Sigma_{22 \cdot 3}^{-1} \sigma_{(i \cdot 3)})}} \\ &= \sqrt{\frac{\sigma_{(i \cdot 3)} \Sigma_{22 \cdot 3}^{-1} \sigma_{(i \cdot 3)}}{\sigma_{1 \cdot 3}}} \quad (2.2.5) \end{aligned}$$

Definition (2.1).

The linear combination of $(X^{(2)}|X^{(3)})$ that has maximum correlation with $\beta(X_i|X^{(3)})$ is called the Partial Regression Function and is defined as

$$\beta(X^{(2)}|X^{(3)})$$

where $\beta_{r \times 1} = \sigma_{(i \cdot 3)} \Sigma_{22 \cdot 3}^{-1}$

Definition (2.2).

The correlation coefficient between $(X_i|X^{(3)})$ and the partial regression function $\beta(X^{(2)}|X^{(3)})$ is called the Partial Multiple correlation coefficient. The population value of this coefficient is denoted by

$\bar{R}_{i \cdot (r|s)}$ where

$$\bar{R}_{i \cdot (r|s)}^2 = \frac{\sigma_{(i \cdot 3)} \Sigma_{22 \cdot 3}^{-1} \sigma_{(i \cdot 3)}}{\sigma_{1 \cdot 3}} \quad (2.2.6)$$

The corresponding sample value will be denoted by $R_{i \cdot (r|s)}^2$ and will be defined later.

In passing we would explain the notation $\bar{R}_{i \cdot (r|s)}$ more fully.

The subscript i denotes the i th variable of the set $(X^{(1)}|X^{(3)})$.

Thus i locates the dependent variable.

The (\cdot) signifies the correlation.

$(r|s)$ denotes the number of variables in two sets. r is the number of variables in the set $(X^{(2)}|X^{(3)})$ and s is the number of variables in the fixed set $X^{(3)} = x^{(3)}$. The $|$ denotes the conditioning and any variables to the right of $|$ are fixed.

The above notation is most convenient for theoretical considerations, because it is found that the sampling distribution of $\bar{R}_{i \cdot (r|s)}$ depends on the number of variables in each of the sets rather than their actual labelling. In practical applications it will usually be necessary to know exactly which components of the original vector X are fixed and which are considered in the regression relationship. In this case we extend the notation by replacing r and s by the actual components involved in the relationship and dropping the brackets. Thus $\bar{R}_{i \cdot (r|s)}$ then becomes $\bar{R}_{i \cdot q+1, q+2, \dots, q+r | q+r+1, \dots, p}$ where $\{q+1, q+2, \dots, q+r\}$ are the labels of the components of $(X^{(2)}|X^{(3)})$ and $\{q+r+1, \dots, p\}$ are the labels of the fixed set $X^{(3)} = x^{(3)}$.

Suppose for example we have a six component vector $X' = (X_1 X_2 | X_3 X_4 | X_5 X_6)$. Then $\bar{R}_{2 \cdot 3, 4 | 5, 6}$ is the partial multiple correlation coefficient between $(X_2 | X_5 = x_5, X_6 = x_6)$ and $(X_3, X_4 | X_5 = x_5, X_6 = x_6)$. If we are not interested in the individual components we would write the same correlation coefficients as $\bar{R}_{2 \cdot (2|2)}$.

We shall now derive some properties of the partial regression function and the partial multiple correlation coefficient. Firstly $\bar{R}_{i \cdot (r|s)}$,

like all correlation coefficients does not depend on the units in which the variables are measured.

Theorem 2.2.2.

$\bar{R}_{i \cdot (r|s)}$ is invariant with respect to non-singular transformations of X_i and $(X^{(2)} | X^{(3)})$.

Proof. Let a and b be any scalars, $a \neq 0$.

$B_{(rxr)}$ be any non-singular matrix

$d_{(rx1)}$ be any vector

Let $Y_i = a X_i^* + b$

$$Y^{(2)} = B(X^{(2)} | X^{(3)}) + d.$$

The Covariance matrix of $\begin{pmatrix} Y_i \\ Y^{(2)} \end{pmatrix}$ is

$$\begin{vmatrix} a^2 \sigma_{ii \cdot 3} & a \sigma_{(i \cdot 3)} B' \\ B \sigma'_{(i \cdot 3)} a & B \Sigma_{22 \cdot 3} B' \end{vmatrix}$$

The partial regression function of Y_i on $Y^{(2)}$ is

$$a \sigma_{(i \cdot 3)} B' (B \Sigma_{22 \cdot 3} B')^{-1} Y^{(2)}$$

and the square of the partial multiple correlation coefficient is

$$\frac{a \sigma_{(i \cdot 3)} B' (B \Sigma_{22 \cdot 3} B')^{-1} B \sigma'_{(i \cdot 3)} a}{a^2 \sigma_{ii \cdot 3}}$$

which reduces to

$$\frac{\sigma_{i \cdot 3} \Sigma_{22 \cdot 3}^{-1} \sigma'_{(i \cdot 3)}}{\sigma_{ii \cdot 3}} = \bar{R}_{i \cdot (r|s)}^2$$

which is the partial multiple correlation between X_i^* and $X^{(2)} | X^{(3)}$.

Suppose we wish to estimate X_i^* using the partial regression function then the variable $(X_i^* - \beta(X^{(2)} | X^{(3)}))$ can be thought of as the error in

our estimate, or as that part of the variable $(X_i^* | x^{(3)})$ that is not "explained" by $\beta(X^{(2)} | x^{(3)})$. The two properties we give below show that the partial regression function has certain described characteristics.

Theorem 2.2.3.

$(X_i^* - \beta(X^{(2)} | x^{(3)}))$ and $(X^{(2)} | x^{(3)})$ are uncorrelated.

Proof.

$$\begin{aligned} & \text{Cov}(X_i^* - \beta(X^{(2)} | x^{(3)}), X^{(2)} | x^{(3)}) \\ &= \text{Cov}(X_i^*, X^{(2)} | x^{(3)}) - \beta \Sigma_{22 \cdot 3} \\ &= \sigma_{i \cdot 3} - \sigma_{i \cdot 3} \Sigma_{22 \cdot 3}^{-1} \Sigma_{22 \cdot 3} \\ &= 0 \end{aligned}$$

Theorem 2.2.4.

Of all possible linear combinations of $(X^{(2)} | x^{(3)})$, the partial regression function makes the variance of $(X_i^* - \beta(X^{(2)} | x^{(3)}))$ a minimum.

Proof. Let $\alpha(X^{(2)} | x^{(3)})$ be an arbitrary linear combination of $(X^{(2)} | x^{(3)})$.

$$\begin{aligned} \text{Var}(\alpha(X^{(2)} | x^{(3)})) &= \text{Var}(X_i^* - \beta(X^{(2)} | x^{(3)})) + (\beta - \alpha)(X^{(2)} | x^{(3)}) \\ &= \text{Var}(X_i^* - \beta(X^{(2)} | x^{(3)})) + \text{Var}((\beta - \alpha)(X^{(2)} | x^{(3)})) \end{aligned} \quad (2.2.7)$$

the covariance terms vanishing by Theorem 2.2.3. Hence (2.2.7) becomes

$$\text{Var}(X_i^* - \beta(X^{(2)} | x^{(3)})) + (\beta - \alpha) \Sigma_{22 \cdot 3} (\beta - \alpha)' \quad (2.2.8)$$

The second term of (2.2.8) is non-negative since $\Sigma_{22 \cdot 3}$ is positive definite. Hence (2.2.8) attains its minimum when $\beta = \alpha$.

Thus the partial regression function makes the variance of the residual a minimum.

We can easily obtain an expression for the residual variance

$$\text{Var}(X_i^* - \beta(X^{(2)} | x^{(3)})) = \text{Var}(X_i^*) + \text{Var}(\beta(X^{(2)} | x^{(3)})) - 2\beta \text{Cov}(X_i^* X^{(2)} | x^{(3)}) \quad (2.2.9)$$

These terms can be found by taking the appropriate elements of Σ_3 . Thus (2.2.9) becomes

$$\begin{aligned} & \sigma_{ii.3} + \beta \Sigma_{22.3} \beta' - 2\beta \sigma_{(i.3)} \\ &= \sigma_{ii.3} + \sigma_{(i.3)} \Sigma_{22.3}^{-1} \sigma_{(i.3)}' - 2\sigma_{(i.3)} \Sigma_{22.3}^{-1} \sigma_{(i.3)}' \\ &= \sigma_{ii.3} - \sigma_{(i.3)} \Sigma_{22.3}^{-1} \sigma_{(i.3)}' \end{aligned}$$

$$\text{i.e. } \text{Var}(X_i^* - \beta(X^{(2)} | x^{(3)})) = \sigma_{ii.3} - \sigma_{(i.3)} \Sigma_{22.3}^{-1} \sigma_{(i.3)}' \quad (2.2.10)$$

$$\text{But } \sigma_{(i.3)} \Sigma_{22.3}^{-1} \sigma_{(i.3)}' = \sigma_{ii.3} \bar{R}_{i.(rls)}^2$$

Thus we also have the useful relationship

$$\text{Var}(X_i^* - \beta(X^{(2)} | x^{(3)})) = \sigma_{ii.3} (1 - \bar{R}_{i.(rls)}^2) \quad (2.2.11)$$

Since $\sigma_{ii.23} = \sigma_{ii.3} - \sigma_{(i.3)} \Sigma_{22.3}^{-1} \sigma_{(i.3)}'$ it follows from (2.2.10) that

$$\text{Var}(X_i^* - \beta(X^{(2)} | x^{(3)})) = \sigma_{ii.23} \quad (2.2.12)$$

Equating (2.2.11) and (2.2.12) we have

$$\sigma_{ii.3} (1 - \bar{R}_{i.(rls)}^2) = \sigma_{ii.23}$$

$$\text{i.e. } 1 - \bar{R}_{i.(rls)}^2 = \frac{\sigma_{ii.23}}{\sigma_{ii.3}} \quad (2.2.13)$$

From (2.2.13) we see that the complement of the partial multiple correlation coefficient will give us a measure of the reduction in variance that is gained by using the conditional distribution of $\begin{matrix} X_i \\ x^{(2)} \\ x^{(3)} \end{matrix}$ compared with the conditional distribution of $(X_i^{(1)} | x^{(3)})$.

Thus we arrive at an alternative definition of the partial multiple correlation coefficient as

$$\bar{R}_{i.(rls)}^2 = 1 - \frac{\sigma_{ii.23}}{\sigma_{ii.3}} \quad (2.2.14)$$

If $\sigma_{ii.23}$ and $\sigma_{ii.3}$ are very nearly equal, their ratio will be close to 1 and $\bar{R}_{i.(rls)}^2$ will be very small. If $\sigma_{ii.23}$ is small compared to $\sigma_{ii.3}$ then their ratio will be small and $\bar{R}_{i.(rls)}^2$ will be large.

Thus a small value of $\bar{R}_{i.(rls)}^2$ will imply that there is very little reduction in variance by using $\left| \begin{matrix} X_i & | & x^{(2)} \\ & & x^{(3)} \end{matrix} \right|$ instead of $(X_i | x^{(3)})$ and a large value of $\bar{R}_{i.(rls)}^2$ will imply that there is considerable reduction in variance using both sets. As will be seen later, a measure of this kind could be useful in deciding whether or not a certain set of variables should be included in a regression analysis.

Using the relationship, (2.2.13), $1 - \bar{R}_{i.(rls)}^2 = \frac{\sigma_{ii.23}}{\sigma_{ii.3}}$ we can find an expression for $\bar{R}_{i.(rls)}^2$ in terms of the elements of $\Sigma_{.3}$ and its inverse $\Sigma_{.3}^{-1}$. We shall demonstrate this for the case where $q = 1$, then

$$\Sigma_{.3} = \begin{vmatrix} \sigma_{11.3} & \sigma_{(1.3)} \\ \sigma_{(1.3)}' & \Sigma_{22.3} \end{vmatrix}$$

Let $\sigma^{11.3}$ denote the $(1,1)^{\text{th}}$ element of $\Sigma_{.3}^{-1}$ (i.e. the upper left hand corner element). By (Anderson 1958, problem 18 chapter 2), we have the relationship

$$\frac{1}{\sigma_{11.3}} = \sigma_{11.3} - \sigma_{(1.3)} \Sigma_{22.3}^{-1} \sigma_{(1.3)}' = \sigma_{11.23} \quad (2.2.15)$$

and by (2.2.13) when $q = 1$ we have

$$1 - \bar{R}_{1.(rls)}^2 = \frac{\sigma_{11.23}}{\sigma_{11.3}}$$

Using (2.2.15) we obtain

$$\bar{R}_{1.(rls)}^2 = 1 - \frac{1}{\sigma_{11.3} \sigma^{11.3}}$$

This result will hold for $q \geq 1$ if instead of $\Sigma_{.3}$ we consider the reduced matrix

$$\Sigma^* = \begin{vmatrix} \sigma_{11.3} & \sigma_{(i.3)} \\ \sigma_{(i.3)}' & \Sigma_{22.3} \end{vmatrix}$$

which is obtained from $\Sigma_{.3}$ by selecting the appropriate rows and columns.

Using (2.2.16) we can find an expression for $1 - \bar{R}_{1.(r|s)}^2$ since

$$\sigma_{11.3} = \frac{\text{Cof} \sigma_{11.3}}{\|\Sigma_{.3}\|} = \frac{\|\Sigma_{22.3}\|}{\|\Sigma_{.3}\|} \quad (2.2.17)$$

where $\|A\|$ denotes the determinant of A .

$$\text{Thus } 1 - \bar{R}_{1.(r|s)}^2 = \frac{\|\Sigma_{.3}\|}{\sigma_{11.3} \|\Sigma_{22.3}\|} \quad (2.2.18)$$

Let us now consider the relationship between the partial multiple correlation coefficient and the "ordinary" multiple correlation coefficient. Let us suppose for convenience that in the original vector X the partitioning is such that $X^{(1)}$ has only one component (i.e. $q = 1$, $r \geq 1$, $s \geq 1$, $q+r+s = p$). The results derived will still hold if $q > 1$, by considering the i^{th} variable of the first set.

Let us denote the population multiple correlation coefficient between X_1 and $(r+s)$ variables X_2, X_3, \dots, X_p by $\bar{R}_{1.(r+s)}$ and let $\bar{R}_{1.s}$ be the population multiple correlation coefficient between X_1 and the last s variables of the above set.

Anderson 1958 page 32 states that the conditional variance of X_1 given the $(r+s)$ variables X_2, \dots, X_p is $\sigma_{11.23} = (1 - \bar{R}_{1.(r+s)}^2) \sigma_{11}$ where σ_{11} is $\text{Var}(X_1)$ (2.2.19)

and similarly the conditional variance of X_1 given the s variables

X_{r+2}, \dots, X_p is

$$\sigma_{11 \cdot 3} = (1 - \bar{R}_{1 \cdot s}^2) \sigma_{11} \quad (2.2.20)$$

By (2.2.14)

$$\begin{aligned} \bar{R}_{1 \cdot (r|s)}^2 &= 1 - \frac{\sigma_{11 \cdot 23}}{\sigma_{11 \cdot 3}} \\ &= 1 - \frac{1 - \bar{R}_{1 \cdot r+s}^2}{1 - \bar{R}_{1 \cdot s}^2} \\ &= \frac{\bar{R}_{1 \cdot r+s}^2 - \bar{R}_{1 \cdot s}^2}{1 - \bar{R}_{1 \cdot s}^2} \end{aligned} \quad (2.2.21)$$

Equation (2.2.21) could also be derived directly from the alternate definition of $\bar{R}_{1 \cdot (r|s)}^2$ given in (2.2.6) and using 2.2.19 and 2.2.20

$$\bar{R}_{1 \cdot (r|s)}^2 = \frac{\sigma_{(1 \cdot 3) \Sigma 22 \cdot 3}^{-1} \sigma_{(1 \cdot 3)}}{\sigma_{11 \cdot 3}} \quad (2.2.22)$$

$$\text{and } 1 - \bar{R}_{1 \cdot (r|s)}^2 = \frac{\sigma_{11 \cdot 3}^{-1} \sigma_{(1 \cdot 3) \Sigma 22 \cdot 3} \sigma_{(1 \cdot 3)}}{\sigma_{11 \cdot 3}}$$

$$= \frac{\sigma_{11 \cdot 23}}{\sigma_{11 \cdot 3}} \quad (2.2.23)$$

Using (2.2.19) and (2.2.20)

$$1 - \bar{R}_{1 \cdot (r|s)}^2 = \frac{1 - \bar{R}_{1 \cdot r+s}^2}{1 - \bar{R}_{1 \cdot s}^2} \quad (2.2.24)$$

From which it follows that

$$\bar{R}_{1 \cdot (r|s)}^2 = \frac{\bar{R}_{1 \cdot r+s}^2 - \bar{R}_{1 \cdot s}^2}{1 - \bar{R}_{1 \cdot s}^2} \quad (2.2.25)$$

which is the same as (2.2.21).

From (2.2.21) and (2.2.24) we have that

$$\frac{\bar{R}_{1 \cdot (r|s)}^2}{1 - \bar{R}_{1 \cdot (r|s)}^2} = \frac{\bar{R}_{1 \cdot r+s}^2 - \bar{R}_{1 \cdot s}^2}{1 - \bar{R}_{1 \cdot r+s}^2} \quad (2.2.26)$$

We can demonstrate the relationship between the partial multiple correlation coefficient and the partial correlation coefficient by considering the case where $(X^{(2)}|X^{(3)})$ consists of a single component (i.e. $r = 1$, $q = 1$ and $s \geq 1$).

The $r \times r$ matrix $\Sigma_{22.3}$ becomes $\sigma_{22.3}$, the conditional variance of $(X^{(2)}|X^{(3)})$ and

$$\begin{aligned} \bar{R}_{1.(1|s)}^2 &= \frac{\sigma_{12.3}^2 \sigma_{22.3}^{-1} \sigma_{21.3}}{\sigma_{11.3}} \\ &= \frac{\sigma_{12.3}^2}{\sigma_{11.3} \sigma_{22.3}} = \rho_{12.34, \dots, p}^2 \end{aligned} \quad (2.2.27)$$

which is the square of the population partial correlation coefficient between X_1 and X_2 keeping the remaining s variables fixed.

III.2 The Partial Multiple Correlation Coefficient in the Sample.

Let us suppose we have a sample of size N from a $N_p(\mu, \Sigma)$ population. From this sample we can calculate the maximum likelihood estimate of $\Sigma_{.3} = \frac{1}{N} A_{.3}$ as given in Chapter I section 3.

Let us suppose that the original vectors of the sample have been partitioned into three sets of components where $q = 1$, $r > 1$, $s \geq 1$. The conditional sample covariance matrix $\hat{\Sigma}_{.3}$ will be partitioned accordingly into

$$\begin{aligned} \hat{\Sigma}_{.3} &= \begin{vmatrix} \hat{\sigma}_{11.3} & \hat{\sigma}_{(1.3)} \\ \hat{\sigma}_{(1.3)}' & \hat{\Sigma}_{22.3} \end{vmatrix} \\ &= \frac{1}{N} \begin{vmatrix} a_{11.3} & a_{(1.3)} \\ a_{(1.3)}' & A_{22.3} \end{vmatrix} \end{aligned} \quad (3.2.1)$$

The maximum likelihood estimate of the square of the partial multiple

correlation coefficient will be given by

$$R_{1 \cdot (rls)}^2 = \frac{\hat{\sigma}_{(1 \cdot 3)} \hat{\Sigma}_{22 \cdot 3}^{-1} \hat{\sigma}'_{(1 \cdot 3)}}{\hat{\sigma}_{11 \cdot 3}} \quad (3.2.2.)$$

$$= \frac{a_{(1 \cdot 3)} A_{22 \cdot 3}^{-1} a'_{(1 \cdot 3)}}{a_{11 \cdot 3}} \quad (3.2.3)$$

It is clear that $R_{1 \cdot (rls)}^2$ has the same properties in the sample as $\bar{R}_{1 \cdot (rls)}^2$ has in the population. Thus the formulas given in the preceding section for $\bar{R}_{1 \cdot (rls)}^2$ will be valid for the sample value if the terms in the formulas are replaced by their sample estimates.

Two of them (2.2.6) and (2.2.21) make it easy to deduce the sampling distribution of $R_{1 \cdot (rls)}$.

By (2.2.1)

$$\begin{aligned} 1 - R_{1 \cdot (rls)}^2 &= \frac{\hat{\sigma}_{11 \cdot 3} - \hat{\sigma}_{(1 \cdot 3)} \hat{\Sigma}_{22 \cdot 3}^{-1} \hat{\sigma}'_{(1 \cdot 3)}}{\hat{\sigma}_{11 \cdot 3}} \\ &= \frac{a_{(11 \cdot 3)} - a_{(1 \cdot 3)} A_{22 \cdot 3}^{-1} a'_{(1 \cdot 3)}}{a_{11 \cdot 3}} \\ &= \frac{a_{11 \cdot 23}}{a_{11 \cdot 3}} \end{aligned} \quad (3.2.4)$$

Thus by (3.2.3) and (3.2.4)

$$\frac{R_{1 \cdot (rls)}^2}{1 - R_{1 \cdot (rls)}^2} = \frac{a_{(1 \cdot 3)} A_{22 \cdot 3}^{-1} a'_{(1 \cdot 3)}}{a_{11 \cdot 23}} \quad (3.2.5)$$

We see that equations (3.2.3), (3.2.4) and (3.2.5) only involve elements of the matrix A_3 . From theorem 1.4.1 Chapter I, we have that A_3 is distributed as $W(\Sigma_3, n-s)$ where $n = N-1$. This implies that the sampling distribution of the partial covariance matrix A_3 is the same as that of a covariance matrix based on a sample of size $N-s$ from a

(p-s) dimensional normal population with covariance matrix Σ_3 . (Since we have p dimensional vectors partitioned into $q = 1, r \geq 1$ and $s \geq 1$ sets and since $q+r+s = p$, we have then $p - s = r + 1$).

More over the functions of A_3 that are used in the definition of the sample multiple correlation coefficient are exactly the same as the functions that would be used in the definition of the sample multiple correlation coefficient based on a sample of size N-s from a normal population with $(p-s-1) = r$ independent variables.

Thus it follows that the distribution of the sample partial multiple correlation coefficient $R_{1.(r|s)}$ is the same as the distribution of the sample multiple correlation coefficient based on a sample of size N-s from an (r+1) dimensional normal population with (squared) population multiple correlation coefficient

$$\bar{R}_{1.(r|s)}^2 = \frac{\sigma_{(1.3)} \Sigma_{22.3}^{-1} \sigma'_{(1.3)}}{\sigma_{11.3}}$$

A detailed derivation of the distribution of the multiple correlation coefficient in both the central and the non-central case is given in Anderson 1958, pages 93-96. A derivation of the distribution of $R_{1.(r|s)}$ would be identical with N-s replacing N, r+1 replacing p and the elements of the conditional matrix A_3 replacing the elements of A.

To illustrate these statements let us derive the distribution of $R_{1.r|s}$ when $\bar{R}_{1.(r|s)} = 0$ (i.e. the central case).

In the population we have by (2.2.3)

$$\bar{R}_{1.(r|s)}^2 = \frac{\beta \Sigma_{22.3} \beta'}{\sigma_{11.3}}$$

where $\beta = \sigma_{(1.3)} \Sigma_{22.3}^{-1}$.

Thus $\bar{R}_{1.(rls)}^2 = 0$ if and only if $\beta = 0$.

From (2.3.5) we have

$$\frac{R_{1.(rls)}^2}{1-R_{1.(rls)}^2} = \frac{a_{(1.3)} A_{22.3}^{-1} a'_{(1.3)}}{a_{11.23}}$$

From Theorem 1.4.3, Chapter I specialised for $q = 1$, and since $\beta = 0$ $a_{(1.3)} A_{22.3}^{-1} a'_{(1.3)}$ is distributed as $\sum_{\alpha=n-s-r+1}^{n-s} V_{\alpha}^2$ where $\{V_{\alpha}\}$ are independent $N(0, \sigma_{11.23})$ variables and $n = N-1$

$$\therefore \frac{a_{(1.3)} A_{22.3}^{-1} a'_{(1.3)}}{\sigma_{11.23}} \text{ is distributed as } \chi_r^2 \quad (2.3.6)$$

By Theorem 1.4.2 Chapter I when $q = 1$ we have $a_{11.23}$ is distributed as $\sum_{\alpha=1}^{n-s-r} V_{\alpha}^2$ where $\{V_{\alpha}\}$ are independent $N(0, \sigma_{11.23})$ variables

$$\therefore \frac{a_{11.23}}{\sigma_{11.23}} \text{ is distributed as } \chi_{n-s-r}^2 \quad (2.3.7)$$

where $n = N-1$.

But (2.3.6) and (2.3.7) are independent, so it follows that

$$\frac{R_{1.(rls)}^2}{1-R_{1.(rls)}^2} \frac{n-s-r}{r} \text{ is distributed as } F_{r, n-s-r} \quad (2.3.8)$$

We can now change the variable in the density function of $F_{r, n-s-r}$ into

$$R = \sqrt{\frac{\frac{r}{N-s-r-1} F_{r, N-s-r-1}}{1 + \frac{r}{N-s-r-1} F_{r, N-s-r-1}}} \quad \text{and let } N = n+1$$

and multiply by the Jacobian. We obtain the density function of the partial multiple correlation coefficient under the hypothesis $\bar{R}_{1.(rls)} = 0$

i.e.

$$f(R) = \frac{2\Gamma(\frac{1}{2}(N-s-1))}{\Gamma(\frac{1}{2}r)\Gamma(\frac{1}{2}(N-s-r-1))} R^{(r-1)} (1-R^2)^{\frac{1}{2}(N-s-r-1)-1} \quad (2.3.9)$$

$$0 \leq R \leq 1$$

which we see is the same as the density function of the sample multiple correlation coefficient between a dependent variable and independent variables when the population multiple correlation coefficient is zero. The sample multiple correlation being based on sample of size $N-s$ from an $(r+1)$ dimensional normal population.

An alternative derivation of the distribution of the partial multiple correlation coefficient has been given by Linhart (1958, 1960b) and another derivation will be given in Chapter III.

We shall now consider some applications of the partial multiple correlation coefficient. The partial multiple correlation coefficient can be used to determine whether or not certain variables should be excluded in a regression analysis. In a regression analysis we have a single variable, the criterion, and a number of other variables which the experimenter proposes to use in a linear equation which will predict the value of the criterion. In problems of this type it often happens that more variables are available than are actually needed to give a good prediction of the criterion. The difficulty arises because the "true" coefficients of the prediction equation are unknown and have to be estimated from sample values of the criterion and the predictor variables. If the true coefficients were known the inclusion of extra variables would always improve the prediction but in practice, especially if the calibration sample is small compared to the number of predictor variables, it is possible that the prediction of the criterion will be very poor for individuals not included

in the calibration sample. Much attention has been given to the problem of choosing the best subset of the available predictor variables and a complete discussion is beyond the scope of this thesis, but we would mention that a very lucid discussion of this problem can be found in Browne (1969). We shall only discuss the role of the partial multiple correlation coefficient in prediction.

Let us suppose that we wish to predict X_1 using either the set of s variables $X^{(3)}$ or the $(r+s)$ variables $X^{(2)}, X^{(3)}$. We shall also assume that the components of $X^{(2)}$ and $X^{(3)}$ have been decided before a sample has been drawn from the population. Let us first consider what happens in the population. One possible measure of the accuracy of the prediction of X_1 by linear combination of $X^{(3)}$ is given by $\text{Var}(X_1 - \beta X^{(3)}) = \sigma_{11.3}$ (Rao 1965 page 225) where

$$\sigma_{11.3} = \sigma_{11}(1 - \bar{R}_{1.s}^2) \quad (2.3.10)$$

Similarly using $X^{(2)}$ and $X^{(3)}$, the accuracy of the prediction can be measured by

$$\sigma_{11.23} = \sigma_{11}(1 - \bar{R}_{1.r+s}^2) \quad (2.3.11)$$

The reduction in conditional variance using both sets will be

$$\begin{aligned} & \sigma_{11}(1 - \bar{R}_{1.s}^2) - \sigma_{11}(1 - \bar{R}_{1.r+s}^2) \\ & = \sigma_{11}(\bar{R}_{1.r+s}^2 - \bar{R}_{1.s}^2) \end{aligned} \quad (2.3.12)$$

The proportional reduction in conditional variance by using both sets will be

$$\frac{\bar{R}_{1.r+s}^2 - \bar{R}_{1.s}^2}{1 - \bar{R}_{1.s}^2} \quad (2.3.13)$$

which is by ((2.2.21) page 32) is the square of the partial multiple correlation coefficient $\bar{R}_{1 \cdot (r|s)}^2$. Thus the partial multiple correlation coefficient measures the contribution of $X^{(2)}$ to the prediction equation and a test based on the sample partial multiple correlation coefficient would indicate if $X^{(2)}$ should be included in the prediction equation. If the inclusion of $X^{(2)}$ in the regression equation improves the prediction of the criterion, one would intuitively expect the conditional variance $\sigma_{11 \cdot 23}$ to be less than $\sigma_{11 \cdot 3}$.

Suppose we wish to test if there is any significant difference between the two conditional variances $\sigma_{11 \cdot 3}$ and $\sigma_{11 \cdot 23}$ i.e. we wish to test

$$H_0 \quad \sigma_{11 \cdot 3} = \sigma_{11 \cdot 23}$$

against the alternative

$$H_1 \quad \sigma_{11 \cdot 3} > \sigma_{11 \cdot 23} \quad (2.3.14)$$

This test is equivalent to testing if there is any significant difference between the two multiple correlation coefficients $\bar{R}_{1 \cdot s}$ and $\bar{R}_{1 \cdot r+s}$ and the test statistic for this hypothesis based on a sample of size N from the multivariate normal population of $(X_1, X^{(2)}, X^{(3)})$ will be given by (H.E. Anderson Jr. and Fruchter 1960)

$$\frac{R_{1 \cdot r+s}^2 - R_{1 \cdot s}^2}{1 - R_{1 \cdot r+s}^2} \quad \frac{N - (r+s) - 1}{r} \quad (2.3.15)$$

which under H_0 is distributed as F with r and $N-(r+s)-1$ degrees of freedom. But comparing (2.3.15) with (2.2.26) page 32 we see that (2.3.15) is exactly the ratio

$$\frac{R_{1 \cdot (r|s)}^2}{1 - R_{1 \cdot (r|s)}^2} \quad \frac{N - (r+s) - 1}{r}$$

So hypothesis (2.3.14) is equivalent to testing if the partial multiple correlation coefficient is zero.

Linhart (1958, 1960a) using a different approach modifies the above test since, as he remarks, it is not sufficient that the multiple correlation merely *increases* when extra variables are included but it must increase by more than a certain positive constant to insure that the prediction is improved. The amount that the multiple correlation must increase depends chiefly on the size of the calibration sample. Linhart uses the expected length of the confidence interval as a measure of predictive precision. A calibration sample is drawn from the joint distribution of the criterion and the predictor variables and the coefficients of the prediction equation are estimated from this sample. Then a further observation is made on the predictor variables. A confidence interval on the criterion can be determined from this observation. The length of this confidence interval, ℓ say, is a random variable and its expected value taken over all possible prediction sets and over all possible samples of fixed size can be used as a measure of the accuracy of the prediction. Suppose $E_{r+s}(\ell)$ is the expected length of the confidence interval using both sets of variables (i.e. $X^{(2)}$ and $X^{(3)}$) and $E_S(\ell)$ is the expected length using only the set $X^{(3)}$. Linhart tests the hypothesis

$$H_0 \quad E_{r+s}(\ell) \geq E_S(\ell)$$

against the alternative

$$H_1 \quad E_{r+s}(\ell) < E_S(\ell)$$

and recommends that the set $X^{(2)}$ be included in the prediction equation

only if H_0 can be rejected. The test of this hypothesis is based on the

(2.3.16)

estimated ratio $\frac{\hat{E}_{r+s}(\lambda)}{\hat{E}_{r+s}(\lambda)}$ and the test statistic is (Linhart 1960a eq.4)

$$\sqrt{\frac{R_{1 \cdot r+s}^2 - R_{1 \cdot s}^2}{1 - R_{1 \cdot s}^2}} \quad (2.3.17)$$

Comparing this formula with (2.2.25) page 32 we see that this is exactly the partial multiple correlation coefficient. Thus Linhart's test is equivalent to testing if the partial multiple correlation coefficient is greater than a certain value, a value which depends chiefly on the sample size. This value can be approximated by (Linhart 1960a Eq.15)

$$\sqrt{\frac{r}{N-s-1}} \quad (2.3.18)$$

This test uses the non-central distribution of the (partial) multiple correlation coefficient and we also note that Linhart's derivation of the distribution of his statistic is an alternative proof of the non-central distribution of R . We can state Linhart's hypotheses as

$$\begin{aligned} H_0: R_{1 \cdot (r|s)}^2 &\leq \frac{r}{N-s-1} \\ H_1: R_{1 \cdot (r|s)}^2 &> \frac{r}{N-s-1} \end{aligned} \quad (2.3.19)$$

The additional set of variables $X^{(2)}$ will be included only if H_0 is rejected i.e. if H_1 is "proved". Using the test discussed earlier $X^{(2)}$ is included if $\bar{R}_{1 \cdot (r|s)}^2$ is significantly different from zero. With Linhart's test $X^{(2)}$ will only be included if $R_{1 \cdot r|s}^2$ is greater than $\frac{r}{N-s-1}$. For very large samples these two tests become equivalent. The essential difference between the

two tests discussed is this. The usual test is to test if the partial coefficient multiple correlation/differs from zero. Linhart tests if the partial multiple correlation is greater than a given positive constant. The first hypothesis is a point hypothesis whereas Linhart's is an interval hypothesis. Linhart's test will be much stricter especially if the size of the calibration sample, N , is small compared to the number of predictor variables. Linhart's test will tend to include far fewer variables in the prediction equation. It is when the calibration sample is small that the number of predictor variables has greatest influence on the accuracy of the prediction, and it is usually found in such situations that equations with a large number of variables give worse predictions than equations with fewer variables. Thus it would appear that Linhart's test has distinct advantages over the usual procedure and deserves more recognition than it has hitherto received. We note that in the special case when we wish to decide whether or not to include a single extra variable in the prediction equation, the test reduces to testing

$$H_0 \quad \rho_{12 \cdot s}^2 \leq \frac{1}{N-s-1}$$

where $\rho_{12 \cdot s}$ is the partial correlation coefficient between X_1 and X_2 , say, for fixed values of the remaining variables. Linhart gives a detailed description of how to apply the test in (Linhart 1960a).

A third measure of predictive precision is given by the expected value of the mean square error of prediction (Kerridge 1967). If X_1 is predicted with the s variables of $X^{(3)}$, the expected value of the mean square error of prediction will be given by

$$\delta_{(s)}^2 = \frac{N+1}{N} \left(\frac{N-2}{N-s-2} \right) \sigma_{11 \cdot 3} \quad N > s + 2 \quad (2.3.21)$$

This quantity can be estimated in the sample by (Darlington 1968 Eq.14)

$$\delta_{(s)}^2 = \frac{(N-2)(N+1)}{(N-s-1)(N-s-2)} s_{11.3} \quad (2.3.22)$$

where $s_{11.3}$ is the appropriate element of $\frac{1}{N} A_{.3}$.

Similarly using both $X^{(2)}$ and $X^{(3)}$ to predict X_1 , the expected mean square error of prediction will be

$$\delta_{r+s}^2 = \frac{(N-2)(N+1)}{(N-r-s-1)(N-r-s-2)} \sigma_{11.23} \quad (2.3.23)$$

The decision whether or not to include $X^{(2)}$ in the prediction set can be based on

$$H_0 \quad \delta_{(s)}^2 \leq \delta_{(r+s)}^2$$

against the alternative

$$H_1 \quad \delta_{(s)}^2 > \delta_{(r+s)}^2$$

(2.3.24)

Browne (1969) shows that this hypothesis is equivalent to testing

$$H_0 \quad \bar{R}_{1.(r|s)}^2 \leq \frac{r}{(N-s-2)}$$

$$H_1 \quad \bar{R}_{1.(r|s)}^2 > \frac{r}{(N-s-2)}$$

(2.3.25)

(2.3.25) differs from Linhart's only in the constant term. The test procedure and the decision rule would be the same as in Linhart's test. Both tests involve the non-central distribution of the multiple correlation coefficient. The percentage points of this distribution have not been tabulated but graphs of the lower 95% confidence limits for the population multiple correlation coefficient are given in (Linhart 1958 and Ezekiel 1929) detailed description of their use to test the hypotheses is given in a later paper by the same author (Linhart 1960). When a decision about a single variable is required, tables of the percentage points of the correlation coefficient (David 1968) may be used since in this case the multiple correla-

tion coefficient is the same as the absolute value of the correlation coefficient. An approximate normalisation of the multiple correlation coefficient has been given by Hodgson 1968. If the multiple correlation coefficient has been based on a sample of size N from a p variate normal population and if

$$U = \frac{R}{(1-R^2)^{\frac{1}{2}}}, \quad \bar{U} = \frac{\bar{R}}{(1-\bar{R}^2)^{\frac{1}{2}}}$$

when R and \bar{R} are the sample and population multiple correlations respectively. Between one variable and $p-1$ other variables then,

$$\frac{U(N-p-\frac{1}{2})^{\frac{1}{2}} - (p-2+(N-\frac{3}{2})\bar{U}^2)^{\frac{1}{2}}}{(1+\frac{1}{2}U^2+\frac{1}{2}\bar{U}^2)^{\frac{1}{2}}}$$

is approximately a standardised normal variate.

A similar normalisation exists for the correlation coefficient.

11.4. The Partial Trace Correlation Coefficient.

The partial multiple correlation coefficient gives a measure of the relationship between a single conditional variable and a set of conditional variables. Suppose now we wish to measure the relationship between two sets of conditional variables. $(X^{(1)} | X^{(3)})$ and $(X^{(2)} | X^{(3)})$ with q and r components respectively. This problem is more complex than that of Section 3 since we need to measure relationships between vectors, or if we consider the associated covariance matrices, we must measure relationships between matrices. Such a measure must be defined in two stages. Firstly the relationship between the vectors (or matrices) must be defined and then a scalar function of the relationship must be found, in order for it to be of use for statistical treatment. Hooper (1963) has suggested just such a measure by considering the matrix

$$T = \Sigma_{11.3}^{-1} \Sigma_{12.3} \Sigma_{22.3}^{-1} \Sigma_{21.3} \quad (2.4.1)$$

and its complement

$$(I-T) = I - \Sigma_{11.3}^{-1} \Sigma_{12.3} \Sigma_{22.3}^{-1} \Sigma_{21.3} \quad (2.4.2)$$

(2.4.1) is a matrix generalisation of the partial correlation coefficient. This can easily be seen by considering the case where each set consists of a single conditional variable. Under these conditions (2.4.1) becomes

$$\sigma_{11.3} \sigma_{22.3}^{-1} \sigma_{21.3} \quad (2.4.3)$$

which is the square of the partial correlation coefficient between X_1 and X_2 when the remaining variables are held fixed.

(2.4.1) is a $(q \times q)$ matrix and in order for it to be mathematically tractable we require some function $f(T)$ say, that will carry the matrix into a scalar. In order for the measure to have the usual properties of a correlation coefficient we further require that

- (i) The measure be independent of the units in which $(X^{(1)} \quad X^{(3)})$ and $(X^{(2)} \quad X^{(3)})$ are expressed i.e.

$$f(PTP^{-1}) = f(T) \text{ where } P \text{ is nonsingular.}$$

- (ii) The measure of what has been explained of the variation in the conditioned set plus what has not been explained should sum to unity i.e.

$$f(T) + f(I-T) \equiv 1.$$

- (iii) The measure itself must lie between 0 and 1 i.e.

$$0 \leq f(T) \leq 1.$$

A function that will carry a matrix into a scalar is the trace. The trace of a matrix is defined as the sum of the diagonal elements and it can also be easily shown that this is the sum of the characteristic roots. Since

the characteristic roots of a matrix are invariant under a transformation such as (i), the trace is also invariant under such transformations; so property (i) is satisfied. Properties (ii) and (iii), in addition to (i), will be satisfied if we define the population partial trace correlation coefficient as

$$\begin{aligned} \bar{r}_{qr \cdot s}^2 &= \frac{1}{q} \text{tr}(T) \\ &= \frac{1}{q} \sum_{i=1}^q \lambda_i^2 \quad \text{where} \end{aligned} \quad (2.4.2)$$

$\{\lambda_i\}$ are the characteristic roots of T . Thus the partial trace correlation coefficient is the average value of the characteristic roots of T , and it is easily verified that $\bar{r}_{qr \cdot s}^2$ has properties (i), (ii) and (iii) above.

The partial trace correlation is defined in the sample as

$$\begin{aligned} \bar{r}_{qr \cdot s}^2 &= \frac{1}{q} \text{tr} (A_{11 \cdot 3}^{-1} A_{12 \cdot 3} A_{22 \cdot 3}^{-1} A_{21 \cdot 3}) \\ &= \frac{1}{q} \sum_{i=1}^q \ell_i^2 \quad \text{where } \{\ell_i\} \text{ are the sample} \end{aligned} \quad (2.4.3)$$

characteristic roots.

As will be seen in the next section these characteristic roots are the sample partial canonical correlations and the roots in (2.4.2) are the population partial canonical correlations.

For different values of q, r and s the partial trace correlation coefficient takes on familiar forms. For $q = r = 1, s \geq 1$, the partial trace correlation coefficient is the partial correlation coefficient mentioned above. When $s = 0$, the partial trace correlation is the trace correlation. For a full discussion of this correlation see Hooper (1959). Under the condition, $s = 0$, when $q = 1, r = 1$, we have disregarding

sign, the correlation coefficient between X_1 and X_2 and for $q = 1$ $r \geq 1$, the multiple correlation coefficient. For $q = 1$ $r > 1$ $s \geq 1$, we have the partial multiple correlation coefficient.

The partial trace correlation coefficient could be used as an aid to selecting the best subset of variables in a multi-equation regression analysis. Since it gives a measure of the extent to which a given subset, $X^{(2)}$ say, of the independent variables $X^{(2)}$ and $X^{(3)}$ account for the variation in the dependent variables when the effect of $X^{(3)}$ has been eliminated. The best subset would be the one with the highest partial trace correlation.

We remark for any one wishing to refer to Hooper's original paper that our matrix T is equivalent to Hooper's I-F.

11.5. Partial Canonical Correlation Coefficients

The partial canonical correlation coefficients are the zero order canonical correlations between the two sets of conditioned variables $(X^{(1)} | X^{(3)})$ and $(X^{(2)} | X^{(3)})$ with q and r components respectively (Roy and Whittlesey 1952). In the population, these are derived in the same way as the canonical correlations except that the conditional covariance matrix $\Sigma_{\cdot 3} = \begin{vmatrix} \Sigma_{11 \cdot 3} & \Sigma_{12 \cdot 3} \\ \Sigma_{21 \cdot 3} & \Sigma_{22 \cdot 3} \end{vmatrix}$ is used.

$$\Sigma_{\cdot 3} = \begin{vmatrix} \Sigma_{11 \cdot 3} & \Sigma_{12 \cdot 3} \\ \Sigma_{21 \cdot 3} & \Sigma_{22 \cdot 3} \end{vmatrix} \text{ is used.}$$

Thus the squares of the partial canonical correlations are the non-zero roots of the q order determinantal equation

$$|\Sigma_{12 \cdot 3} \Sigma_{22 \cdot 3}^{-1} \Sigma_{21 \cdot 3} - \lambda \Sigma_{11 \cdot 3}| = 0 \quad (2.5.1)$$

or equivalently the non-zero roots of the r order equation

$$|\Sigma_{21 \cdot 3} \Sigma_{11 \cdot 3}^{-1} \Sigma_{12 \cdot 3} - \lambda \Sigma_{22 \cdot 3}| = 0 \quad (2.5.2)$$

The non-zero roots are the same and the two equations differ only in the multiplicity of the zero roots. Let $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_t^2$, $t \leq \min(q, r)$ be the non-zero roots. As with canonical correlation theory, using these roots, we can determine linear combinations $\alpha^{(i)}_i(X^{(2)} | X^{(1)})$, $\gamma^{(i)}_i(X^{(2)} | X^{(3)})$ which for the same fixed $i = 1, \dots, t$, have the greatest partial correlation. These scalar variables are called the partial canonical variables. Thus the partial canonical variables are the variables U_1, U_2, \dots, U_t , V_1, V_2, \dots, V_t such that

$$(1) \begin{aligned} U_i &= \alpha^{(i)}_i(X^{(1)} | X^{(3)}) & i &= 1, \dots, t \\ V_i &= \gamma^{(i)}_i(X^{(2)} | X^{(3)}) & i &= 1, \dots, t \end{aligned}$$

$$(2) \begin{aligned} \text{Cov}(U_i, U_j) &= 1 & \text{for } i &= j \\ &= 0 & \text{for } i &\neq j \end{aligned}$$

i.e. Each U_i has unit variance and is uncorrelated with all the other U 's.

Similarly

$$(3) \begin{aligned} \text{Cov}(V_i, V_j) &= 1 & \text{for } i &= j \\ &= 0 & \text{for } i &\neq j \end{aligned}$$

$$(4) \begin{aligned} \text{Cov}(U_i, V_j) &= 0 & \text{for } i &\neq j \\ &= \lambda_i & \text{for } i &= 1, \dots, t \end{aligned}$$

where λ_i is the positive square root of the i^{th} canonical correlation and t is the rank of $\Sigma_{12 \cdot 3}$.

(4) implies that the $\{U_i\}$ and $\{V_j\}$ are uncorrelated except for t cases. In these cases the correlations between them are the partial canonical correlations. Since we have arranged the $\{\lambda_i^2\}$ in descending order, it

follows that the first pair of partial canonical variates are the linear combinations of $(X^{(1)}|X^{(3)})$ and $(X^{(2)}|X^{(3)})$ that have maximum correlation. In the case where the set $(X^{(1)}|X^{(3)})$ has only one component $q = 1$, the first canonical variate will be the linear combination of $(X^{(2)}|X^{(3)})$ that has maximum correlation with $(X_1|X^{(3)})$ and in this case the first canonical correlation, λ_1 will be the population partial multiple correlation coefficient.

In the sample the partial canonical correlations are found either from $\hat{\Sigma}_{.3} = \frac{1}{N} A_{.3}$ or its unbiased estimate $\frac{1}{N-s-1} A_{.3}$. They could also be computed from the matrix of sample partial correlation coefficients

$$R = \begin{vmatrix} 1 & r_{12(3)} & \cdots & r_{1q+r(3)} \\ r_{21(3)} & 1 & & r_{2q+r(3)} \\ & & \ddots & \\ r_{q+r(3)} & \cdots & & 1 \end{vmatrix}$$

where $r_{ij(3)}$ is the partial correlation between the i^{th} and j^{th} variables in $\begin{vmatrix} X^{(1)} \\ X^{(2)} \end{vmatrix} | X^{(3)}$. R is of course partitioned according to $\Sigma_{.3}$.

A difficulty arises with the sample estimates of the partial canonical correlations. In the sample, the rank of $\hat{\Sigma}_{12.3}$ will always be $\min(q,r)$ even in cases where the true rank of $\Sigma_{12.3}$ is $< \min(q,r)$. Thus we shall obtain non-zero estimates of population zero roots. In practice these estimates are likely to be very small and would fail to pass any test of their significance. It is usually found that most of the partial canonical correlation structure can be explained by the first few partial canonical correlations. If we add the assumption that the variables $(X^{(1)}|X^{(3)})$ and $(X^{(2)}|X^{(3)})$ have a multivariate normal distribution, it is shown in

(Roy 1957 Page 41) that the joint distribution of sample partial canonical correlations is exactly the same as the joint distribution of the canonical correlations, with N replaced by $N-s$, where s is the number of components of $x^{(3)}$ and N the sample size. Let $\lambda_1 > \lambda_2 > \dots > \lambda_q$ be the roots of

$$|A_{12 \cdot 3} A_{22 \cdot 3}^{-1} A_{21 \cdot 3} - \lambda A_{11 \cdot 3}| = 0$$

where $A_{ij \cdot 3}$ is as defined in Chapter 1, Section 3 and is based on a sample of size N . Then under the hypothesis

$$H_0 \quad \Sigma_{12 \cdot 3} = 0$$

the joint distribution of the squares of the partial canonical correlations will be given by

$$C \prod_{i=1}^q \lambda_i^{\frac{1}{2}(r-q-1)} (1-\lambda_i)^{\frac{1}{2}(N-s-r-q-2)} \prod_{i < j}^q (\lambda_i - \lambda_j)$$

where

$$C = \pi^{\frac{1}{2}q} \prod_{i=1}^q \frac{\Gamma(\frac{1}{2}(N-s-i))}{\Gamma(\frac{1}{2}(N-s-r-i))\Gamma(\frac{1}{2}(q+1-i))\Gamma(\frac{1}{2}(r+1-i))} \quad (2.5.4)$$

We remarked earlier that when $q = 1$, there is only a single partial canonical correlation and this is of course equivalent to the partial multiple correlation coefficient. Thus we can specialise (2.5.4)

for $q = 1$ and obtain the density function of the square of the partial multiple correlation coefficient.

$$f(R^2) = \frac{\Gamma(\frac{1}{2}(N-s-1))}{\Gamma(\frac{1}{2}(N-s-r-1))\Gamma(\frac{1}{2}r)} R^{r-2} (1-R^2)^{\frac{1}{2}(N-s-r-3)} \quad (2.5.5)$$

CHAPTER III

The Generalised Partial Correlation Matrix

III.1. The Square root of a Matrix.

Before discussing the Generalised Partial Correlation Matrix we shall give two definitions of the square root of a matrix which we shall use in the next section.

Definition I.

Let B be any positive definite matrix with dimension $q \times q$.

Let Γ be an orthogonal matrix such that

$$\Gamma' B \Gamma = D$$

where $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_q)$

i.e. D is a diagonal matrix with the q characteristic roots of B down the main diagonal.

$$\text{Then } B^{\frac{1}{2}} = \Gamma D^{\frac{1}{2}} \Gamma'$$

where

$$D^{\frac{1}{2}} = \text{diag}(\lambda_1^{\frac{1}{2}}, \lambda_2^{\frac{1}{2}}, \dots, \lambda_q^{\frac{1}{2}}).$$

This definition gives a symmetric square root

$$\text{i.e. } (B^{\frac{1}{2}})' = B^{\frac{1}{2}}.$$

Definition II.

Let B be a positive definite matrix, then B can be expressed as

$$B = T T'$$

where T is nonsingular and may be triangular.

The square root of B is then defined as

$$B^{\frac{1}{2}} = T.$$

This square root is not necessarily symmetric.

In an expression such as

$$B^{\frac{1}{2}}AB^{\frac{1}{2}}$$

we make the convention that the post multiplier is $(B^{\frac{1}{2}})'$.

We also note that $B^{-\frac{1}{2}} = (B^{\frac{1}{2}})^{-1}$, i.e. we find the square root first and then take the inverse.

III.2. The Generalised Partial Correlation Matrix in the Population.

An extremely elegant and general matrix from which all the types of partial correlations discussed in the previous sections can be found was suggested by (Troskie, 1968). This matrix is an extension of the Generalised Multiple Correlation Matrix defined by Khatri in 1964. (Khatri, 1964, Troskie, 1968).

As in Chapter I let us suppose that X is a $p \times 1$ random vector which is distributed as $N_p(\mu, \Sigma)$, and let X be partitioned into three sets of components $X^{(1)}$, $X^{(2)}$ and $X^{(3)}$, with q, r , and s components respectively. Let $q \leq r$ and $q+r+s = p$. The covariance matrix of Σ is partitioned accordingly. Suppose now that we keep $X^{(3)}$ fixed at $x^{(3)}$. The covariance matrix of the conditional distribution is $\begin{vmatrix} X^{(1)} & | & x^{(3)} \\ X^{(2)} & & \end{vmatrix}$ is

$$\Sigma_{\cdot 3} = \begin{vmatrix} \Sigma_{11 \cdot 3} & \Sigma_{12 \cdot 3} \\ \Sigma_{21 \cdot 3} & \Sigma_{22 \cdot 3} \end{vmatrix}$$

where $\Sigma_{11 \cdot 3}$ is $q \times q$, $\Sigma_{12 \cdot 3}$ is $r \times q$ and $\Sigma_{22 \cdot 3}$ is $r \times r$, $q \leq r$ (see Chapter I).

The Partial Generalised Correlation Matrix is defined in the population as

$$P_{.3} = \Sigma_{11.3}^{-\frac{1}{2}} \Sigma_{12.3} \Sigma_{22.3}^{-1} \Sigma_{21.3} \Sigma_{11.3}^{-\frac{1}{2}} \quad (3.2.1)$$

where the square root used can be either the symmetric square root (Definition I) or the non-symmetric square root (Definition II) and the post-multiplier $\Sigma_{11.3}^{-\frac{1}{2}}$ is $(\Sigma_{11.3}^{-\frac{1}{2}})'$.

To illustrate the very general properties of $P_{.3}$, let us consider its form for different values of q, r , and s .

For $q = 1$, $r = 1$ and $s = p-2$

$$\begin{aligned} P_{.3} &= \sigma_{11.3}^{-\frac{1}{2}} \sigma_{12.3} \sigma_{22.3}^{-1} \sigma_{21.3} \sigma_{11.3}^{-\frac{1}{2}} \\ &= \frac{\sigma_{12.3}^2}{\sigma_{11.3} \sigma_{22.3}} \end{aligned}$$

which is the square of the partial correlation coefficient between X_1 and X_2 keeping the s variables X_3, X_4, \dots, X_p fixed.

For $q = 1$, $r \geq 1$ and $s \geq 1$

$$\begin{aligned} P_{.3} &= \sigma_{11.3}^{-\frac{1}{2}} \sigma_{(1.3)} \Sigma_{22.3}^{-1} \sigma_{(1.3)}' \sigma_{11.3}^{-\frac{1}{2}} \\ &= \frac{\sigma_{(1.3)} \Sigma_{22.3}^{-1} \sigma_{(1.3)}'}{\sigma_{11.3}} = \bar{R}_{1.r|s}^2 \end{aligned}$$

i.e. $P_{.3}$ is the square of the partial multiple correlation coefficient discussed in (Chapter II, Sections 2 and 3).

If $s = 0$, then the original vector X is partitioned into only two sets of components. $\Sigma_{.3}$ then becomes

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

and $P_{.3}$ becomes P the population Generalised Multiple Correlation matrix

$$R = \Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-\frac{1}{2}}.$$

The properties of this matrix are discussed fully in Troskie(1968 and

1969) and are not of direct concern here, but the analogy between the "unconditional case" when $s = 0$ and the "conditional" case when $s > 0$ is very clear and bears the same type of relationship as exists between the correlation coefficient and the partial correlation coefficient.

The characteristic roots of $P_{.3}$ are the solutions of the determinantal equation

$$|\Sigma_{11.3}^{-\frac{1}{2}} \Sigma_{12.3} \Sigma_{22.3}^{-1} \Sigma_{21.3} \Sigma_{11.3}^{-\frac{1}{2}} - \lambda I| = 0$$

But these are the same as the roots of

$$|\Sigma_{11.3}^{-1} \Sigma_{12.3} \Sigma_{22.3}^{-1} \Sigma_{21.3} - \lambda I| = 0$$

Thus the characteristic roots of $P_{.3}$ are the squares of the partial canonical correlations.

The trace of the generalised partial correlation matrix is

$$\text{tr } P_{.3} = \text{tr } \Sigma_{11.3}^{-\frac{1}{2}} \Sigma_{12.3} \Sigma_{22.3}^{-1} \Sigma_{21.3} \Sigma_{11.3}^{-\frac{1}{2}} = \text{tr } T$$

where T is defined in Chapter II section 4.

Thus $\frac{1}{q} \text{tr } P_{.3}$ is exactly the partial trace correlation coefficient.

Hence we see that each type of partial correlation can be derived from $P_{.3}$ by considering special cases.

III.3. The Generalised Partial Correlation Matrix in the Sample.

The sample estimate of $P_{.3}$ is obtained from the conditional Wishart matrix $A_{.3}$. Let $X_{(\alpha)}$, $\alpha = 1, \dots, N$ be a random sample from $N_p(\mu, \Sigma)$ and let $X_{(\alpha)}$'s be partitioned as the vector X . Then as in Chapter I Section 3 we find the maximum likelihood estimate of the covariance matrix of the conditional distribution of $\begin{vmatrix} X^{(1)} & | & x^{(3)} \\ X^{(2)} & & \end{vmatrix}$ is given by

$$\hat{\Sigma}_{.3} = \frac{1}{N} A_{.3} \quad \text{where}$$

$$A_{\cdot 3} = \begin{vmatrix} A_{11 \cdot 3} & A_{12 \cdot 3} \\ A_{21 \cdot 3} & A_{22 \cdot 3} \end{vmatrix}$$

It is clear that

$$R_{\cdot 3} = A_{11 \cdot 3}^{-\frac{1}{2}} A_{12 \cdot 3}^{-1} A_{22 \cdot 3}^{-\frac{1}{2}}$$

is the maximum likelihood estimate of $P_{\cdot 3}$ and that $R_{\cdot 3}$ has the same properties in the sample as $P_{\cdot 3}$ has in the population. For example, if $q = 1$, $r \geq 1$ and $s \geq 1$,

$$R_{\cdot 3} = a_{11 \cdot 3}^{-\frac{1}{2}} a_{(1 \cdot 3)}^{-1} a_{22 \cdot 3}^{-\frac{1}{2}} = R_{1 \cdot r | s}^2$$

which is the square of the sample partial multiple correlation coefficient.

From the properties of $P_{\cdot 3}$ and its sample estimate $R_{\cdot 3}$, it is evident that if the sampling distribution of $R_{\cdot 3}$ is known, the distributions of the other partial correlations will follow as special cases. If the distribution of the roots of $R_{\cdot 3}$ is known, we shall also have the distributions of the partial canonical correlations and the partial trace correlation.

We shall derive the distribution of $R_{\cdot 3}$ for the central case, i.e. when $P_{\cdot 3} = 0$, where 0 is a $q \times q$ zero matrix. Now $P_{\cdot 3} = 0$ if and only if $\Sigma_{12 \cdot 3} = 0$. In this case $(X^{(1)} | X^{(3)})$ and $(X^{(2)} | X^{(3)})$ are conditionally independent. We note that this does not necessarily imply complete independence of $X^{(1)}$ and $X^{(2)}$.

We shall state the central distribution of $R_{\cdot 3}$ as theorem.

Theorem 3.3.1.

If $R_{\cdot 3}$ is the maximum likelihood estimator of $P_{\cdot 3}$ based on a sample of size $N > p$ from $N_p(\mu, \Sigma)$ then, under the hypothesis $P_{\cdot 3} = 0$, the

distribution of R_3 is multivariate Beta Type I with parameters $\frac{1}{2}r$, $\frac{1}{2}(n-s-r)$, and dimension q where

$$n = N-1$$

$$q = \text{number of components of } X^{(1)}$$

$$r = \text{number of components of } X^{(2)}$$

$$s = \text{number of components of the fixed subvector } X^{(3)}$$

$$q+r+s = p \quad \text{and} \quad q \leq r.$$

Proof. Let $B = A_{12 \cdot 3} A_{22 \cdot 3}^{-1} A_{21 \cdot 3}$

$$E = A_{11 \cdot 3} - A_{12 \cdot 3} A_{22 \cdot 3}^{-1} A_{21 \cdot 3}$$

Then

$$R_3 = (E+B)^{-\frac{1}{2}} B (E+B)^{-\frac{1}{2}}$$

If $P_3 = 0$, $\Sigma_{12 \cdot 3} = 0$ and it follows from Theorem 1.4.2 and Theorem 1.4.3 of Chapter I Section 4 that B and E are independent and both have a Wishart distribution, where

$$E \text{ is } W(\Sigma_{11 \cdot 23}, n-s-r)$$

$$B \text{ is } W(\Sigma_{11 \cdot 33}, r)$$

Hence R_3 has a multivariate Beta Type I distribution (Olkin and Rubin 1964, Troskie 1966) with parameters $\frac{1}{2}r$ and $\frac{1}{2}(n-s-r)$ and dimension q . i.e. R_3 is $B_1(\frac{1}{2}r, \frac{1}{2}(n-s-r))$.

The density function of R_3 will be

$$f(R_3) = \frac{\Gamma_q(\frac{1}{2}(n-s))}{\Gamma_q(\frac{1}{2}r) \Gamma_q(\frac{1}{2}(n-s-r))} |R_3|^{\frac{1}{2}(r-q-1)} |I-R_3|^{\frac{1}{2}(n-s-r-q-1)} \quad (3.3.1)$$

$$R_3 > 0$$

where

$$\Gamma_q(\frac{1}{2}n) = \pi^{\frac{1}{4}q(q-1)} \prod_{i=1}^q \Gamma(\frac{1}{2}(n-i+1))$$

From the density function of R_3 we can easily find the moments of the determinant $|R_3|$.

Theorem 3.3.2.

The h^{th} moment of $|R_3|$ is given by

$$E|R_3|^h = \frac{\Gamma_q(\frac{1}{2}(n-s))\Gamma_q(\frac{1}{2}(r+2h))}{\Gamma_q(\frac{1}{2}r)\Gamma_q(\frac{1}{2}(n-s-2h))} \quad (3.3.2)$$

Proof.

$$E|R_3|^h = \frac{\Gamma_q(\frac{1}{2}(n-s))}{\Gamma_q(\frac{1}{2}r)\Gamma_q(\frac{1}{2}(n-s-r))} \int_0^1 \dots \int_0^1 |R_3|^h |R_3|^{\frac{1}{2}(r-q-1)} |I-R_3|^{\frac{1}{2}(n-s-r-q-1)} dR_3$$

where dR_3 are the differentials of the $\frac{1}{2}q(q+1)$ different elements of the determinant $|R_3|$.

$$\begin{aligned} &= \frac{\Gamma_q(\frac{1}{2}(n-s))}{\Gamma_q(\frac{1}{2}r)\Gamma_q(\frac{1}{2}(n-s-r))} \int \dots \int |R_3|^{\frac{1}{2}(r-q+2h-1)} |I-R_3|^{\frac{1}{2}(n-s-r-q-1)} dR_3 \\ &= \frac{\Gamma_q(\frac{1}{2}(n-s))}{\Gamma_q(\frac{1}{2}r)\Gamma_q(\frac{1}{2}(n-s-r))} \cdot \frac{\Gamma_q(\frac{1}{2}(r+2h))\Gamma_q(\frac{1}{2}(n-s-r))}{\Gamma_q(\frac{1}{2}(n-s+2h))} \\ &= \frac{\Gamma_q(\frac{1}{2}(n-s))\Gamma_q(\frac{1}{2}(r+2h))}{\Gamma_q(\frac{1}{2}r)\Gamma_q(\frac{1}{2}(n-s+2h))} \end{aligned}$$

In the same way we can also derive the moments $|I-R_3|$ and obtain

Theorem 3.3.3.

$$E|I-R_3|^h = \frac{\Gamma_q(\frac{1}{2}(n-s))\Gamma_q(\frac{1}{2}(n-s-r+2h))}{\Gamma_q(\frac{1}{2}(n-s-r))\Gamma_q(\frac{1}{2}(n-s+2h))}$$

Proof. The same as the proof of Theorem 3.3.2.

For certain values of q, r , and s , the density function of $|R_3|$ takes on some interesting forms. Substituting $q = 1, r = 1, s \geq 1$ in formula 3.3.1 we obtain

$$f(r^2) = \frac{\Gamma(\frac{1}{2}(n-s))}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}(n-s-1))} r^{-1} (1-r^2)^{\frac{1}{2}(n-s-3)} \quad 0 < r^2 < 1$$

which is the density function of the square of the sample partial correlation coefficient $r_{12 \cdot (s)}^2$, based on a sample of size $N = n+1$ drawn from a $N_p(\mu, \Sigma)$ population, when the population partial correlation coefficient $\rho_{12 \cdot (s)}^2$ is zero.

We can also express this density function in terms of $r_{12 \cdot (s)}$ and the sample size $N = n+1$.

$$f(r) = \frac{\Gamma(\frac{1}{2}(N-s-1))}{\sqrt{\pi}\Gamma(\frac{1}{2}(N-s-2))} (1-r^2)^{\frac{1}{2}(N-s-4)} \quad -1 < r < 1$$

The moments of $r_{ij \cdot (s)}$ can be obtained from (3.3.2) by setting $q = 1$ and $r = 1$ and we obtain

$$E(r^{2h}) = \frac{\Gamma(\frac{1}{2}(n-s))\Gamma(\frac{1}{2}(2h+1))}{\Gamma(2)\Gamma(\frac{1}{2}(n-s+2h))}$$

Since the density function of the partial correlation coefficient is even, it follows that all the odd order moments are zero. Thus setting $h = 1$, we obtain the variance of the partial correlation coefficients

$$\begin{aligned} \text{Var}(r^2) &= E(r^2) = \frac{\Gamma(\frac{1}{2}(n-s))\Gamma(3/2)}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}(n-s-2))} \\ &= \frac{1}{n-s} = \frac{1}{N-s-1} \end{aligned} \quad (3.3.3)$$

If we set $q = 1$, $r \geq 1$, $s \geq 1$ we obtain the density function of the square of the partial multiple correlation coefficient $R_{1 \cdot (r|s)}^2$

$$f(R^2) = \frac{\Gamma(\frac{1}{2}(n-s))}{\Gamma(\frac{1}{2}r)\Gamma(\frac{1}{2}(n-s-r))} R^{(r-2)} (1-R^2)^{\frac{1}{2}(n-s-r-2)} \quad (3.3.4)$$

We know that the density function of $R_{1 \cdot (r|s)}^2$ has the same form as that of the square of sample multiple correlation coefficient between a dependent variable and r independent variables based on a sample of size $(N-s)$.

Using formula (3.3.2) we can obtain the even moments of $R_{1 \cdot (r|s)}^2$

$$E(R^{2h}) = \frac{\Gamma(\frac{1}{2}(N-s-1))\Gamma(\frac{1}{2}r+h)}{\Gamma(\frac{1}{2}r)\Gamma(\frac{1}{2}(N-s-1)+h)} \quad (3.3.5)$$

Setting $h = 1$ and using the fact that $\Gamma(z+1) = z\Gamma(z)$ we have

$$E(R^2) = \frac{r}{N-s-1} \quad (3.3.6)$$

In Chapter 2 we showed that Linhart's criterion for determining the number of variables to use in a regression analysis depended on the test of the hypothesis $H_0: \bar{R}_{1 \cdot (r|s)}^2 \leq \frac{r}{N-s-1}$. We now see that $\frac{r}{N-s-1}$ is exactly $E(R_{1 \cdot (r|s)}^2)$. We can offer no theoretical reason for this interesting coincidence.

Conclusion.

It can be seen that the Partial Multiple Correlation Matrix summarizes the theory of conditional association in a concise manner. The practical application of this matrix to tests of hypotheses about conditional independence will be discussed in the next chapter.

CHAPTER IV

IV.1 Introduction.

We shall now construct some examples to illustrate some of the statistical techniques discussed in the previous chapters. The data used are the results of physical fitness tests performed by A.W. Sloan (1966a) on more than 6,000 children at High Schools in the Cape Peninsula. The purpose of Sloan's original study was to determine the influence of sex, age and race on physical fitness. The schools were selected to include children from the upper, middle and lower income groups of the White and Coloured communities. At each school every child between the ages of 12 and 18 who was not exempted from physical activity on medical grounds was included in the tests. Physical fitness was measured by the test battery of the American Association for Health, Physical Education and Recreation (AAPHER, 1958). The AAPHER test battery consists of the following seven tests:- pull up (with a modified pull up for girls), sit up, shuttle run, standing broad jump, 50 yard dash, soft ball throw for distance and 600 yard run-walk. In addition sex, age, height and weight were recorded. Detailed instructions for the performance and scoring of each test may be found in Sloan (1966a). The data have been analysed by Sloan (1966b) who made interracial comparisons and compared the South African school children's fitness scores with those of American and British children. With data of this type one would not be interested in predicting the value of a dependent variable using a subset of the independent variables and so the problems presented are essentially those of correlation and not prediction.

We have constructed artificial examples from some of the data to illustrate the use of the partial multiple correlation coefficient and the generalised partial correlation matrix.

We emphasise that the selection of the variables and interpretation of the results are not based on any physiological theory but are simply used to illustrate the application of the statistical techniques involved.

IV.2 An Application of the Partial Multiple Correlation Coefficient.

We have the following variables

X_1 = weight (pounds)

X_2 = height (inches)

X_3 = pull ups (number)

X_4 = sit ups (number)

X_5 = shuttle run (seconds)

X_6 = standard broad jump (feet)

X_7 = 50 yard dash (seconds)

X_8 = soft ball throw (feet)

X_9 = 600 yard run-walk (seconds)

Let us suppose that all the physical fitness variables and height are correlated with weight. Let us further assume that the variables X_3 , X_4 and X_8 measure "muscular power" and that the variables X_2 , X_5 , X_6 , X_7 and X_9 measure "agility". Suppose we wish to investigate if there is any correlation between weight and "agility" when the effect of "muscular power" has been removed.

We partition the variables into three sets

$$X^{(1)} = X_1 \text{ (weight)} \quad q = 1$$

$$X^{(2)} = (X_2, X_5, X_6, X_7, X_9) \quad r = 5$$

$$X^{(3)} = (X_3, X_4, X_8) \quad s = 3.$$

We now ask: Is there any significant correlation between weight (X_1) and the variables of $X^{(2)}$ when the effect of $X^{(3)}$ has been removed?

The partial multiple correlation coefficient $R_{1.(r|s)}$ will give us the answer to this question. Since $X^{(2)}$ has 5 components and $X^{(3)}$ has 3 components, we set up the hypothesis

$$H_0: \bar{R}_{1.(5|3)} = 0$$

$$H_1: \bar{R}_{1.(5|3)} > 0$$

By (2.3.8) page 36 the test statistic will be

$$\frac{R_{1.(5|3)}^2}{1 - R_{1.(5|3)}^2} \frac{N-3-5-1}{5} = F_{5, N-9}$$

The partial multiple correlation coefficient is found by computing the ordinary multiple correlation coefficients $R_{1.8}$ and $R_{1.3}$ where $R_{1.8}$ is the multiple correlation between weight and all the other variables and $R_{1.3}$ is the multiple correlation between weight and the "excluded variables", X_3 , X_4 and X_8 . The partial multiple correlation coefficient is then found using formula (2.2.25) page 32

$$R_{1.(5|3)}^2 = \frac{R_{1.8}^2 - R_{1.3}^2}{1 - R_{1.3}^2}$$

The computation of the test statistic is simplified by using the relation (2.2.26) page 32

$$\frac{R_{1.(5|3)}^2}{1 - R_{1.(5|3)}^2} = \frac{R_{1.8}^2 - R_{1.3}^2}{1 - R_{1.3}^2}$$

The computations of the multiple correlation coefficients were performed on the University's IBM 1130 computer using the standard multiple linear regression programme. The programme allows the selection of different sets of the independent variables and selections can be made as many times as desired. In addition to the two multiple correlation coefficients required to calculate partial multiple correlation coefficient, the ordinary multiple correlation coefficient between X_1 and $X^{(2)}$, $R_{1.5}$, was also calculated so that a comparison could be made between the partial and the "ordinary" case.

The calculations were performed for eight groups of children and for each group we calculated the following multiple correlations.

TABLE I

<u>Coefficient</u>	<u>Dependent Variable</u>	<u>Independent Variable</u>
$R_{1.8}$	1	2,3,4,5,6,7,8,9
$R_{1.3}$	1	3,8,4
$R_{1.5}$	1	2,5,6,7,9.

The results of the computations are given in the table below.

Identification code for the groups

W = White

C = Coloured

Number = groups age in years

B = boy

G = girl

N = number in the sample.

TABLE II

Group	N	$R_{1.8}$	$R_{1.3}$	$R_{1.5}$
WG16	43	.73576	.22540	.70096
WG17	27	.80061	.43956	.76460
WB15	36	.82071	.45741	.77429
WB16	35	.75665	.41690	.71164
WB17	54	.76241	.27127	.75442
CB15	39	.84060	.57333	.83129
CB16	36	.79409	.58504	.65932
CB17	24	.65490	.39575	.59349

The next table gives the information required for the computation of $R_{1.(rls)}$ and the test statistic.

TABLE III

Group	$R_{1.8}^2$	$R_{1.3}^2$	$R_{1.8}^2 - R_{1.3}^2$	$1 - R_{1.8}^2$	$1 - R_{1.3}^2$
WG16	.54134	.05801	.48333	.45866	.94199
WG17	.64098	.19321	.44777	.35902	.80679
WB15	.67356	.20922	.46434	.32644	.79078
WB16	.57252	.17381	.39871	.42748	.82619
WB17	.58127	.07359	.50768	.41873	.92641
CB15	.70661	.32871	.37790	.29339	.67129
CB16	.63058	.34227	.28831	.36942	.65773
CB17	.42889	.15662	.27227	.57111	.84338

The partial multiple correlation coefficient can be calculated from Table III by dividing the elements of column 4 by column 6 and taking the square root. The results are shown in Table IV where in addition in the

last column we give the ordinary multiple correlation coefficient between X_1 and X_2, X_5, X_6, X_7 and X_9 . Except in the two cases marked with an asterisk the partial multiple correlation is smaller than the multiple correlation, so that it would appear that the variables X_3, X_4, X_8 do exert an influence on the regression relationship.

TABLE IV

Group	$R_{1.(513)}^2$	$R_{1.(513)}$	$R_{1.5}$
WG16	.51309	.71631	.70096*
WG17	.55500	.74498	.76460
WB15	.58719	.76628	.77429
WB26	.48259	.69469	.71164
WB17	.54801	.74027	.75442
CB15	.56295	.75030	.83129
CB16	.43834	.66207	.65932*
CB17	.32283	.56818	.59349

We now calculate the test statistic
$$\frac{R_{1.(rls)}^2}{1-R_{1.(rls)}^2} \frac{N-s-r-1}{r} = F_{r, N-s-r-1}$$

where N = sample size, r = No. of variables in the partial regression relationship, s = number of excluded variables. In our case $r = 5$ and $s = 3$.

All the necessary information is in Tables II and III.
$$\frac{R_{1.(513)}^2}{1-R_{1.(513)}^2}$$

can be found from Table III by finding the quotient of corresponding elements in Column 4 and Column 6. The constant term will be $\frac{N-9}{5}$.

The results of the computations are displayed in Table V. All tests are made at the 5% level.

TABLE V

Group	$\frac{R_{1 \cdot (513)}^2}{1 - R_{1 \cdot (513)}^2}$	N-9	$\frac{N-9}{5}$	F	Upper 5% Point $F_{5, N-9}$	Conclusion
WG16	1.05379	34	6.80	7.16	2.49	Reject H_0 :
WG17	1.24720	18	3.60	4.48	2.77	"
WB15	1.42244	27	5.40	7.68	2.57	"
WB16	.93270	26	5.20	4.58	2.59	"
WB17	1.21243	45	9.00	10.91	2.43	"
CB15	1.28805	30	6.00	7.72	2.53	"
CB16	.78044	27	5.40	4.21	2.57	"
CB17	.47674	15	3.00	1.43	2.90	Accept H_0 :

The upper 5% points of the F distribution are shown in the last column of Table V. Comparing the observed value of F with the critical value for each group we see that, except for the CB17 group, H_0 is rejected in every case. Thus we conclude that except for the 17 year old Coloured boys, there is a positive correlation between weight and "agility" when the effects of "muscular power" have been removed.

Having established that the correlation exists we would now be interested in finding out how much correlation, could it possibly be as high as .7?

We set up the hypothesis

$$H_0: \bar{R}_{1 \cdot (r1s)} = .7$$

$$H_1: \bar{R}_{1 \cdot r1s} < .7$$

The sampling distribution of $R_{1 \cdot (r1s)}$ is the same as the distribution of a multiple correlation coefficient based on a sample of size N-s from

an $r+1$ dimensional normal population, where s is the number of excluded variables. Thus we can use Hodgson's approximation to test H_0 . The formula on page 44 we have that

$$Z = \frac{U(N-p-\frac{1}{2})^{\frac{1}{2}} - (p-2 + (N-\frac{3}{2})\bar{U}^2)^{\frac{1}{2}}}{(1 + \frac{1}{2}U^2 + \frac{1}{2}\bar{U}^2)^{\frac{1}{2}}} \quad \text{is } N(0,1)$$

where $U^2 = \frac{R^2}{1-R^2}$ and $\bar{U}^2 = \frac{\bar{R}^2}{1-\bar{R}^2}$ and R and \bar{R} are the sample and population multiple correlation coefficients, p is the dimension of the normal population and N is the sample size. To test the partial multiple correlation coefficient we use the approximation with N replaced by $N-s$ and $p = r+1$. The value of Z is computed and compared with the appropriate percentage point of the normal distribution.

Under $H_0: \bar{R}_1(.513) = .7$ we use the approximation with $\bar{R} = .7$ and $p = 6$ with a sample size of $N-3$. For each group, $\bar{U} = \frac{.7}{(1-.49)^{\frac{1}{2}}} = .9802$ and the appropriate value of U is found by taking the positive square root of the entries in Table V column 1. The computations are summarized in Table VII and all tests are made at the 5% level.

TABLE VII

A = Numerator of Z B = Denominator of Z

Group	N-3	A	B	Z = $\frac{A}{B}$
WG16	40	-.4609	1.4168	-.3253
WG17	24	-.3892	2.1116	-.1839
WB15	33	.2864	1.4803	.6948
WB16	32	-.8937	1.9467	-.4590
WB17	51	-.1651	1.4444	.1143
CB15	36	.0694	1.4574	.0047
CB16	31	-1.3058	1.3676	-.9548

Lower 5% point of Z = -1.64.

In every case we accept H_0 and conclude there is no reason to reject the hypothesis that the partial multiple correlation coefficient is .7 for all the groups.

It would appear that there is very little difference between racial or age groups except in the case of 17 year old Coloured boys. In the other groups there is a high correlation between weight and the "agility" variables when the effects of the "muscular power" have been eliminated. Since these examples are highly artificial there is no point in discussing possible reasons for this.

IV.3 An application of the Generalised Partial Correlation Matrix.

Suppose it is known that the physical fitness variables (X_3, X_4, \dots, X_9) are related to weight and height (X_1, X_2) . Let us assume that the variables (X_3, X_4, X_8) measure "muscular power" and the variables (X_5, X_6, X_7, X_9) measure "agility". Suppose that "muscular power" and "agility" are two different types of physical fitness and we are interested in determining if there is any relationship between the two sets when the effects of weight and height have been removed. A problem of this type could be tackled using the Generalised Partial Matrix, $P_{.3}$. The hypothesis we wish to test is $H_0: P_{.3} = 0$ where 0 is a zero matrix. This hypothesis is equivalent to testing $H_0: \Sigma_{12.3} = 0$ and also by Roy's Union-Intersection principle (Morrison 1967 page 208) to $H_0: \lambda_1^2 = 0$ where λ_1^2 is the largest characteristic root of $P_{.3}$ which is the same as the square of the largest canonical correlation.

If two sets of multivariate normal variables are independent then all

the canonical correlations are zero. Similarly if two sets of variables are conditionally independent all the partial canonical correlations are zero, so a test of conditional independence will be equivalent to testing if all the partial canonical correlations are zero. Four tests of the hypothesis of independence have been proposed in the literature all of which are based on some function of the canonical correlations. Since Roy (1957 page 41) has shown that the joint distribution of the squares of the partial canonical correlations is the same as the distribution of the squares of the canonical correlations with N (the sample size) replaced by $N-s$ where s is the number of variables held fixed, it follows that the same four tests could be used to investigate conditional independence with appropriate adjustment of the parameters. Thus tests of conditional independence will be based on the characteristic roots of $R_{.3}$, the sample generalised partial correlation matrix.

Suppose we have a sample of size N from a p dimensional normal population. Let $X = X^{(1)}, X^{(2)}, X^{(3)}$, be the partitioning of X into three subvectors with q, r , and s components respectively where $q \leq r$ and $q+r+s = p$. Let $r_1^2 \geq r_2^2 \geq \dots \geq r_q^2$ be the characteristic roots of the sample generalised partial correlation matrix $R_{.3}$. Suppose we wish to test the hypothesis $H_0: P_{.3} = 0$. Any one of the following four tests could be used.

(i) Wilk's Likelihood Ratio Criterion

$$W^{(q)} = \prod_{i=1}^q (1-r_i^2) = |1-R_{.3}|$$

(ii) Roy's Largest Root Criterion which is based on the largest (partial) canonical correlation.

(iii) Pillai's Criterion

$$U^{(q)} = \sum_1^q \frac{r_i^2}{1-r_i^2} = \text{tr } R_{.3} (I - R_{.3})^{-1}$$

(iv) Pillai's Criterion

$$V^{(q)} = \sum_1^q r_i^2 = \text{tr } R_{.3}$$

A discussion of these tests (in the unconditional case) and a comparison of their powers when $q=2$ is given in Pillai and Jayachandran (1967).

For $q=2$ the hypothesis of independence is

$$H_0: \lambda_1^2 = \lambda_2^2 = 0$$

Some of the conclusions Pillai and Jayachandran draw are

(i) For small deviations from H_0 , $V^{(2)}$ seems to have greater power than $W^{(2)}$ which in turn has more power than $U^{(2)}$.

(ii) Suppose $\lambda_1^2 + \lambda_2^2 = \text{constant}$. Then the power of $V^{(2)}$ increases as the two roots tend to become equal, but under the same conditions the powers of $U^{(2)}$ and $W^{(2)}$ decrease.

(iii) When the values of the roots are far apart, and there are large deviations from H_0 , the powers of $U^{(2)}$ and $W^{(2)}$ are sometimes greater than $V^{(2)}$ but the power of $V^{(2)}$ is always greater than $U^{(2)}$ and $W^{(2)}$ when the roots tie close together.

(iv) The power of Roy's largest root criterion is below those of the other three tests and if $\lambda_1^2 + \lambda_2^2 = \text{constant}$ the power of the largest root test tends to decrease as the values of the roots tend to equality.

(v) None of the powers of the four criterion have any monotonicity properties with respect to the sum or product of the roots.

Thus $U^{(2)}$, $V^{(2)}$ and $W^{(2)}$ are all good tests of the hypothesis of independence (and conditional independence). For small deviations from the hypothesis there is very little difference in their powers but $V^{(2)}$ has greater power than the rest for large deviations when the roots are close together. The power of the largest root test stays below those of the other three except in the case of large deviations when there is only one non-zero root. In this case power of the largest root tests exceeds those of the other three.

Tabulations of some of the upper percentage points of $V^{(2)}$, $U^{(2)}$, $W^{(2)}$ can be found in Pillai and Jayachandran (1967). Tables of $V^{(q)}$ for $q = 1-50$ have been prepared by Mijares (1967) who also explains their use in detail. If these tables are available, and all other considerations are equal, $V^{(q)}$ would perhaps be the most convenient test since it would not be necessary to compute the characteristic roots of $R_{.3}$. The upper percentage points of the largest characteristic root test have been extensively tabulated and Heck (1960) has given some charts of the upper 5, 2.5 and 1% points. Some of these charts have been reproduced in Morrison (1967) who also explains the test procedure in detail (page 210). Tabulations of the upper percentage points of the largest root distribution, $V^{(q)}$ and $U^{(q)}$ are also given by Pillai (1960). These tabulations are with respect to the parameters m and n . For a test of independence between a q -set and an r -set of variables in a $(q+r)$ dimensional normal population where $q \leq r$, the parameters m and n are $m = \frac{1}{2}(r-q-1)$ $n = \frac{1}{2}(N-q-r-2)$ where N is the sample size. For a test of conditional independence between q and r variables, keeping s

variables fixed, the parameter n must be modified by replacing N by $N-s$.

The distribution of $W^{(q)}$, the Likelihood Ratio Criterion, can be approximated by the Chi-Square distribution (Anderson 1958). For a $(q+r)$ dimensional normal population, as above, it can be shown that $-\log W^{(q)}$ is approximately distributed as χ_f^2 where

$$a = N - \frac{3}{2} - \frac{q+r}{2}$$

$$f = qr.$$

For a test of conditional independence, N would be replaced by $N-s$ in the expression for a . The hypothesis of conditional independence would be rejected if the observed value, χ_f^2 , was greater than $\chi_{\alpha, f}^2$, the upper $\alpha\%$ significance point of χ_f^2 .

The distribution of $V^{(s)}$ can be approximated by the F distribution (Pillai 1960). The statistic

$$F = \frac{(2n+q+1)}{(2m+q+1)} \frac{V^{(q)}}{q-V^{(q)}} \text{ where } m \text{ and } n \text{ are as defined}$$

above, is approximately distributed as F with $f_1 = q(2m+q+1)$ and $f_2 = q(2n+q+1)$ degrees of freedom and the approximation holds good even for small values of n and m .

We shall now illustrate some of these results with an example. Measurements of physical fitness were performed by Sloan (1966) on a sample of 35, seventeen old White boys. In addition their height and weight were recorded.

Is there any relationship between "muscular power", as measured by variables (X_3, X_4, X_8) and "agility", as measured by variables

(X_5, X_6, X_7, X_9) , when the effects of height and weight have been removed? (For a definition of these variables see page 61).

If the two sets of variables are conditionally independent, the generalised partial correlation matrix will be "zero". Thus we set up the hypothesis $H_0: P_{.3} = 0$ and calculate the sample generalised partial correlation matrix between the "muscular power" variables and the "agility" variables. Details of the computation of $R_{.3}$ are given in the appendix.

We obtain the following results

Total number of variables $p = 9$

Number of fixed variables $s = 2$

Number of variables in first set $q = 3$

Number of variables in second set $r = 4$.

Sample size = $N = 35$

Effective sample size = $N-s = 33$.

$$R_{.3} = \begin{vmatrix} .58261 & .23282 & -.04815 \\ .23282 & .22535 & .04146 \\ -.04815 & .04146 & .23279 \end{vmatrix}$$

Characteristic roots of $R_{.3}$ are

$$r_1^2 = .69878 \quad r_2^2 = .25506 \quad r_3^2 = .08691.$$

From which it follows

$$V^{(3)} = \text{tr } R_{.3} = 1.04075$$

$$W^{(3)} = |1 - R_{.3}| = .20898$$

$$U^{(3)} = \text{tr } R_{.3}(1 - R_{.3})^{-1} = 2.75554$$

Largest root = .69878

Percentage points for $V^{(3)}$ and the largest root were obtained from Statistical Tables for tests of Multivariate Hypotheses. (Pillai 1960).

Arguments of Tabulations.

$$m = \frac{1}{2}(r-q-1) = 0 \quad n = \frac{1}{2}(N-s-q-r-2) = 12.$$

Test of $H_0: P_3 = 0$ using $V^{(3)}$.

Upper 5% point of $V^{(3)}$ for $m = 0, n = 12 = .628$ (by Interpolation)

Upper 1% point of $V^{(3)}$ for $m = 0, n = 12 = .749$ (by interpolation)

The observed value of $V^{(3)}$ is greater than both critical values.

Therefore H_0 can be rejected at both the 5 and 1% levels.

Test of H_0 : using Roy's largest root criterion.

Upper 5% point of largest root distribution for $m = 0, n = 12 = .42885$

Upper 1% point of largest root distribution for $m = 0, n = 12 = .51015$.

The observed value is greater than the critical value in both cases.

H_0 can be rejected at both the 5 and 1% levels.

Test of H_0 : using the approximation, $-\log W^{(3)}$.

$$a = 33 - \frac{3}{2} - \frac{7}{2} = 28$$

$$f = 12$$

$$-\log W^{(3)} = 28 \times 1.56542 = 43.83176$$

Upper 5% point of $\chi_{12}^2 = 21.026$

Upper 1% point of $\chi_{12}^2 = 26.217$

Observed value is greater than the critical value in both cases. H_0 can be rejected at both the 5 and 1% levels.

The critical value of $U^{(3)}$ for $m = 0$ and $n = 12$ has not been tabulated. Approximations to the distribution of $U^{(3)}$ are given by

Pillai and Sampson (1959), but can only be used for values of $n \geq 40$ at the 5% level and for $n \geq 50$ at the 1% level. Since in our case $n = 12$, we cannot use the approximation.

All the other test criteria give consistent results. In every case H_0 was rejected, moreover, the value of the test statistic lay well beyond the 1% critical value in all the tests. We may therefore conclude that there is strong evidence of a relationship between "muscular power" and "agility" when the effects of height and weight have been removed.

Concluding Remarks.

The Generalised partial correlation matrix embodies the measures of partial association in a concise way and can be used for practical applications. However, all the tests of hypotheses depend on the distribution of the roots of the matrix. As has been shown these roots are the partial canonical correlations. These roots could also be found by solving the determinantal equation

$$|A_{11 \cdot 3}^{-1} A_{12 \cdot 3} A_{22 \cdot 3}^{-1} A_{21 \cdot 3} - \lambda I| = 0$$

Since many computer systems have a library programme for canonical correlations, the roots could as well be obtained by using the standard programme on the partitioned conditional Wishart matrix. This would make the tests available without the necessity of having to write a long programme, an important consideration when the statistician has a client who needs the results quickly.

A few other thoughts arise as the result of this thesis. There are no tables of the percentage points of the non-central distribution of the multiple correlation coefficient. Much of the tedium of calculation of the percentage points would be relieved by modern high-speed electronic computers and the compilation of such tables for various values of \bar{R} and different sample sizes, would probably present interesting problems in numerical integration. The compilation of tables seems to be a woman's prerogative in statistics and we hope to embark upon such a project in the near future. A knowledge of the percentage points would also give some idea of the precision of Hodgson's approximate normalisation.

Throughout this thesis we have assumed that the underlying distribution of population was normal. The assumption of normality is only needed when the sampling distribution of the various statistics are derived. It would be interesting to see what results are obtained if these ideas are applied to non-normal populations. The multivariate normal distribution has received so much attention in multivariate analysis that other multivariate distributions appear to be neglected. It would seem that there is much interesting work that could be done in this field.

A P P E N D I X

The Computation of the Generalised Partial Correlation Matrix.

The generalised Partial Correlation Matrix was computed on the IBM 1130 computer of the University of Cape Town. The compilation of the programme was greatly facilitated by the large number of subroutines available with this system. The language of the programme is FORTRAN IV. There are six main stages in the programme which we shall discuss in the notation used in the thesis and not use the curious names which appeared when the programme was being written.

Stage I. The Wishart matrix of the observation, A_0 , is calculated.

Stage II. The conditional Wishart matrix $A_{.3}$ is computed.

$A_{.3}$ was computed using the following facts

(Anderson 1958 page 42 problem 18, Morrison 1967 page 65).

Let B be any non-singular partitioned matrix

$$B = \begin{vmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{vmatrix} \quad \text{with inverse} \quad C = \begin{vmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{vmatrix}$$

where B_{11} , B_{22} , C_{11} , C_{22} are all square and the partitioning of C corresponds to that of B . Then $C_{11}^{-1} = B_{11} - B_{12}B_{22}^{-1}B_{21}$.

If the Wishart matrix A is partitioned like B with B_{22} corresponding to the covariance matrix of the variables we wish to keep fixed, then C_{11}^{-1} corresponds to $A_{.3}$ the conditional Wishart matrix. The conditional Wishart matrix could also be found by calculating $A_{11.3} = A_{11} - A_{13}A_{33}^{-1}A_{31}$ etc. This would only involve the inversion of a matrix of small order and would possibly result in greater numerical accuracy. This matter is still being investigated.

Stage III. The Wishart matrix $A_{.3}$ is partitioned into $A_3 = \begin{vmatrix} A_{11.3} & A_{12.3} \\ A_{21.3} & A_{22.3} \end{vmatrix}$

and the square root of $A_{11.3}$ is found.

Stage IV. The square root employed was Choleski's square root i.e.

If A is any positive definite matrix then

$$A = LL'$$

where L is a lower triangular matrix.

The elements of L can be computed from the elements of A and $L = A^{\frac{1}{2}}$. (This type of computation is also used in the solution of systems of linear equations). At this stage the inverse of $A^{\frac{1}{2}}$ was also found.

Stage V. The Generalised partial correlation matrix was computed by

$$R_{.3} = A_{11.3}^{-\frac{1}{2}} A_{12.3} A_{22.3} A_{11.3}^{-\frac{1}{2}}$$

Stage VI. The characteristic roots of R_3 were found.

The programme is shown overleaf. I am greatly indebted to Mrs. B. Nash-Webber and Mr. W.B. de V. Smit for their invaluable assistance in the compilation of this programme and especially for writing the subroutine to compute the matrix square root.

PROGRAMME

```
SUBROUTINE DATA(M,D)
DIMENSION D(9)
READ(2,3)D
03 FORMAT(9F6.0)
RETURN
END
```

```
REAL LB(7),LS(3),LC(4)
DIMENSION XBAR(9),STD(9),WISH(9,9),B(9),D(9),T(9),WISHI(9,9),
1B1(9,7),B11(7,7),B12(2,7),CWISH(7,7),TB(7),CW1(7,3),CW2(7,4),
2CW113(3,3),CW213(4,3),CW123(3,4),CW223(4,4),SCW13(3,3),SCW1I(3,3),
3TS(3),TC(4),CW23I(4,4),AR(3,4),BR(4,3),CR(3,3),R3(3,3),
4CR3(3,3),CVR(3,3),SCWIT(3,3)
EQUIVALENCE (WISH(1,1),WISHI(1,1), B1(1,1))
EQUIVALENCE(B11(1,1),CWISH(1,1))
EQUIVALENCE(CW113(1,1),SCW13(1,1),SCW1I(1,1))
EQUIVALENCE(CW223(1,1),CW23I(1,1))
EQUIVALENCE(R3(1,1),CR3(1,1))
```

```
READ(2,1)N
1 FORMAT(I3)
CALL CORRE(N,9,0,0.0,XBAR,STD,WISH,R,B,D,T)
WRITE(3,30) XBAR
30 FORMAT(1H 'XBAR IS' 9F10.5)
WRITE(3,40) STD
40 FORMAT(1H 'STD IS' 9F10.5)
WRITE(3,50)
50 FORMAT(1H 'THE WISHART MATRIX IS ')
DO 60 I=1,9
60 WRITE(3,70) (WISH(I,J),J=1,9)
70 FORMAT(9F12.3)
```

REARRANGE THE WISHART MATRIX INTO A CONVENIENT FORM FOR PARTITIONING
THIS STEP IS NOT NECESSARY IF THE DATA IS PUNCHED IN THE REQUIRED ORDER

```
CALL RINT(WISH,9,9,8,2)
CALL CINT(WISH,9,8,2)
CALL RINT(WISH,9,9,9,1)
CALL CINT(WISH,9,9,1)
CALL RINT(WISH,9,9,1,4)
CALL CINT(WISH,9,1,4)
WRITE(3,80)
80 FORMAT('1THE REARRANGED WISHART MATRIX IS ')
DO 81 I=1,9
81 WRITE(3,82) (WISH(I,J),J=1,9)
82 FORMAT(9F12.3)
```

CFIND THE INVERSE OF THE WISHART MATRIX. THE INVERSE IS THEN PARTITIONED
CTHE INVERSE OF THE UPPER LEFT HAND SUBMATRIX WILL BE THE CONDITIONAL
CWISHART MATRIX, CALLED CWISH

CALL MINV(WISH,9,DWISH,D,T)

WRITE(3,90)

90 FORMAT (1H,'THE INVERSE OF THE REARRANGED WISHART MATRIX IS')

DO 91 I=1,9

91 WRITE(3,92)(WISHI(I,J),J=1,9)

92 FORMAT(9F12.3)

CALL RCUT(B1,8,B11,B12,9,7,0)

CALL MINV(B11,7,DB11,LB,TB)

WRITE(3,100)

100 FORMAT(1H 'THE CONDITIONAL WISHART MATRIX IS')

DO 101 I=1,7

101 WRITE(3,102) (CWISH(I,J),J=1,7)

102 FORMAT(7F15.5)

CPARTITION THE CONDITIONAL WISHART MATRIX

CALL CCUT(CWISH,4,CW1,CW2,7,7,0)

CALL RCUT(CW1,4,CW113,CW213,7,3,0)

CALL RCUT(CW2,4,CW123,CW223,7,4,0)

CCALCULATE THE SQUARE ROOT OF CW113 USING CHOLESKI'S SQUARE ROOT

CALL MSQRT(CW113,3)

WRITE(3,110)

110 FORMAT(1H 'THE SQUARE ROOT OF CW113 IS')

DO 111 I=1,3

111 WRITE(3,112) (SCW13(I,J),J=1,3)

112 FORMAT(7F15.5)

CFIND THE INVERSE OF THE SQUARE ROOT

CALL MINV(SCW13,3,DSCW,LS,TS)

CFIND THE TRANSPOSE OF THE SQUARE ROOT

CALL MTRA(SCW1I,SCWIT,3,3,0)

CALL MINV(CW223,4,DCW2,LC,TC)

C CALCULATE THE GENERALISED PARTIAL CORRELATION MATRIX

CALL GMPRD(SCW1I,CW123,AR,3,3,4)

CALL GMPRD(CW23I,CW213,BR,4,4,3)

CALL GMPRD(AR,BR,CR,3,4,3)

CALL GMPRD(CR,SCWIT,R3,3,3,3)

WRITE(3,130)

130 FORMAT(1H 'THE GENERALISED PARTIAL CORRELATION MATRIX IS')

DO 131 I=1,3

131 WRITE(3,132) (R3(I,J),J=1,3)

132 FORMAT(3F15.5)

CFIND THE CHARACTERISTIC ROOTS OF R3

CALL MSTR(R3,CR,3,0,1)

CALL EIGEN(CR,CVR,3,0)

WRITE(3,140)

140 FORMAT(1H 'THE CHARACTERISTIC ROOTS OF R3 ARE')

DO 141 I=1,3

141 WRITE(3,142) (CR(I,J),J=1,3)

142 FORMAT(3F15.5)

CALL EXIT

END

BIBLIOGRAPHY

- American Association for Health, Physical Education and Recreation (1958). *Youth Fitness Manual*. Washington, D.C. AAPHER.
- Anderson, Jr. H.E. and Fruchter, B. (1960). Some Multiple Correlation and predictor selection methods. *Psychometrika*. Vol. 25. No. 1 59-76.
- Anderson T.W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.
- Browne, M.W. (1969). Unpublished Ph.D. thesis of the University of South Africa.
- Darlington, R.B. (1968). Multiple regression in research and practice. *Psychological Bulletin*. Vol. 69. No. 3 161-182.
- Ezekial, M.A. and Fox, K.A. (1959). *Methods of correlation and regression analysis*. Wiley, New York.
- Heck, D.L. (1960). Charts of some upper percentage points of the distribution of the largest characteristic root. *Annals of Mathematical Statistics*. Vol. 31. 625-642.
- Hodgson, V. (1968). On the sampling distribution of the multiple correlation coefficient (Abstract). *Annals of Mathematical Statistics*. Vol. 39, 307.
- Hooper, J.W. (1959). Simultaneous equations and canonical correlation theory. *Econometrica*. Vol. 27. 245-256.
- Hooper, J.W. (1962). Partial trace correlations. *Econometrica*. Vol. 30. No. 2. 324-331.
- Keeping, E.S. (1962). *Introduction to Statistical Inference*. D. Van Nostrand, New York.
- Kerridge, D. (1967). Errors of prediction in multiple regression with stochastic regressor variables. *Technometrics*. Vol. 9. No. 2. 309-311
- Linhart, H. (1958). Critère de sélection pour le choix des variables dans l'analyse de régression. *Revue Suisse d'Economie politique et de Statistique*. Vol. 94. 202-232.
- Linhart, H. (1960). A criterion for selecting variables in a regression analysis. *Psychometrika*. Vol. 25. 45-58.
- Linhart, H. (1960). A measure of predictive precision in regression analysis. *Annals of Mathematical Statistics*. Vol. 31. 399-404.

- Mijares, T.A. (1964). *Percentage points of the sum $V_1^{(s)}$ of s roots (s=1-50). A unified table for tests of significance in various univariate and multivariate hypotheses.* Statistical Centre, University of the Philippines, Manila.
- Morrison, D.F. (1967). *Multivariate Statistical Methods.* McGraw-Hill, New York.
- Olkin, I. and Rubin, H. (1964). Multivariate beta distributions and independence properties of the Wishart distribution. *Annals of Mathematical Statistics.* Vol. 35. 261-269.
- Pillai, K.C.S. and Jayachandran, K. (1967). Power comparisons of tests of equality of two multivariate hypotheses based on four criteria. *Biometrika.* Vol. 54. 195-210.
- Pillai, K.C.S. (1960). *Statistical tables for tests of Multivariate Hypotheses.* Statistical Centre, University of the Philippines, Manila.
- Pillai, K.C.S. and Sampson, P. (1959). On Hotelling's Generalisation of T^2 . *Biometrika.* Vol. 46. 160-168.
- Rao, C.R. (1965). *Linear Statistical Inference and its Applications.* Wiley, New York.
- Roy, S.N. (1952). A useful transformation in the case of canonical partial correlations. (Abstract). *Annals of Mathematical Statistics.* Vol. 23. 643.
- Roy, S.N. (1957). *Some aspects of Multivariate Analysis.* Wiley, New York.
- Roy, S.N. and Whittlesey, J. (1952). Canonical Partial Correlations. (Abstract). *Annals of Mathematical Statistics.* Vol. 23. 642.
- Sloan, A.W. (1966a). Physical fitness of Cape Town high school children. *South African Medical Journal.* Vol. 40. 682-687.
- Sloan, A.W. (1966b). Physical fitness of South African compared with British and American high school children. *South African Medical Journal.* Vol. 40. 688-690.
- Troskie, C.G. (1968). The Generalised Multiple Correlation Matrix. To be published in the *South African Statistical Journal.*
- Troskie, C.G. (1969). The Generalised Multiple Correlation Matrix and its applications in Biostatistics. (Paper read at the Symposium on Cross-Disciplinary Sciences in Biomedical Research in April 1969.)