A Non-Least Squares Approach to Linear Models

A Non-Least Squares Approach to Linear Models

By

Mike Jacroux

Cambridge Scholars Publishing

A Non-Least Squares Approach to Linear Models

By Mike Jacroux

This book first published 2023

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-9244-8
ISBN (13): 978-1-5275-9244-5

To Justus Seely for his mentoring and friendship over the years.

TABLE OF CONTENTS

PREFACE

I took a Ph.D. level course in linear models from Justus Seely in 1974. At the time Justus was writing a set of notes that he eventually hoped to turn into a textbook on linear models. Unfortunately, Justus died in 2002 before he was able to finalize his notes into a textbook. In addition, the notes that Justus was developing changed over time as the clientele in his class changed. This text is based on a set of notes that Justus used to teach his 1989 class on linear models. While a good deal of what Justus wrote has been modified to include my own views and prejudices on linear models, the current text relies on ideas Justus developed in his notes. With the kind permission of the Seely family, most of what appears in this text is based on or taken directly from the notes that Justus wrote. The examples and problems that are due to Justus are referenced as Seely(1989).

There are two aspects of the current text that the author believes make it significantly different from most texts on linear models. The first and perhaps most meaningful difference is that whereas most textbooks on linear models initially introduce least squares estimation and then use that as the basis for the development of the theory of best linear unbiased estimation, that is not the approach taken here. Rather, the theory of linear estimation developed here is based on a well-known theorem in mathematical statistics which basically says that if an unbiased estimator for a parameter has zero covariance with all unbiased estimators of zero, then the estimator is a minimum variance unbiased estimator. The reasons for this approach are several. First, it is more statistical in nature. Second, this approach easily allows estimation theory to be developed under the more general assumption of a $\sigma^2 V$ covariance structure where σ^2 is an unknown positive constant and V is a known positive definite matrix rather than a $\sigma^2 I_n$ covariance structure which is typically assumed in most linear models texts. Lastly, in the author's view, the approach used here simplifies the proofs of many of the main results given thus making the text easier to read. Thus the approach towards linear estimation used here has the dual benefits of initially allowing for the consideration of a much wider variety of models and at the same time makes the text simpler to read. The second major difference between this and most other texts on linear models is found in chapter 3 of the current text. In this chapter a systematic approach is given for studying relationships between different parameterizations for a given expectation space. While such relationships have been alluded to in other texts, this is the only formal approach to studying such relationships known to the author and is primarily due to Justus Seely (1989).

Usage of the book

The material presented in this book provides a unifying framework for using many types of models arising in applications such as regression, analysis variance, analysis of covariance and variance component models to analyze data generated from experiments. The author has used the material in the current text to teach a semester long course in linear models at Washington State University to both undergraduate and graduate students majoring in mathematics and statistics. The minimal background required by students to read the text includes three semesters of calculus, an introductory course in mathematical statistics, an undergraduate course in linear algebra and preferably a course on regression or analysis of variance. By having this background, the reader should have gained familiarity with basic concepts in probability and mathematical statistics such as multi-dimensional random variables, expectation, covariance, point estimation, confidence interval estimation and hypothesis testing. However, before the reader embarks on studying this text, it is strongly recommended that the linear algebra material that is provided in the appendices (Appendix A1 through A13) be studied in detail because the material presented in the main text is highly dependent and freely uses the results presented there. These appendices contain a great deal of material on linear algebra not usually contained in an undergraduate course on linear algebra, particularly in appendices A5 through A13. In fact, the author typically spends the first three weeks covering topics in the Appendices such as direct sums of subspaces, projection matrices, generalized inverses, affine sets, etc. With this knowledge in hand, the student can then proceed linearly through the book from section to section. The first four chapters of the book comprise what I consider to be a basic course in linear models and can be covered easily in a semester. Chapter 5 presents some additional topics of interest on estimation theory that might be covered if time allows. There are a number of problems at the end of each chapter that an instructor can choose from to make assignments that will enhance the students' understanding of the material in that chapter. The level of difficulty of these problems ranges from easy to challenging. There are also a number of numerical applied type problems in each chapter that can be assigned to give the students a feel for dealing with data. It is my hope that by studying this text the reader will gain an appreciation of linear models and all its applications.

CHAPTER 1

PROBABILITY AND STATISTICAL PRELIMINARIES

1.1 Random Vectors and Matrices

In this chapter, we introduce some of the basic mathematical and statistical fundamentals required to study linear models. We begin by introducing the ideas of a random vector and a random matrix. To this end, let $Y_1,\ldots,Y_n$ be a set of n random variables. In this text we only consider continuous random variables, hence we associate with $Y_1,\ldots,Y_n$ the joint probability density function $f(y_1,\ldots,y_n)$.

Definition 1.1.1. An n-dimensional vector **Y** is called a continuous random vector if the n components of **Y** are all continuous random variables, i.e., $\mathbf{Y}= (Y_1,\ldots,Y_n)'$ is a continuous random vector if $Y_1,\ldots,Y_n$ are all continuous random variables.

Because in this text we only consider continuous random variables, whenever we refer to a random variable or a random vector it will be assumed to be continuous, thus we shall no longer use the term continuous to describe it. If **Y** is a random vector, we can use more concise notation to describe the joint density function of **Y** such as $f(y)$ or $f_{\mathbf{Y}}(y)$ where the subscript on $f_{\mathbf{Y}}(y)$ may be omitted if the random vector **Y** being considered is clear from the context.

More generally, we can extend the idea of a random vector to that of a random matrix.

Definition 1.1.2. An mxn matrix $\mathbf{W} = (W_{ij})_{mxn}$ is called a continuous random matrix if its mxn components are all continuous random variables, i.e., $\mathbf{W}= (W_{ij})_{mxn}$ is a continuous random matrix if W_{ij}, i = 1,…,m, j = 1,…,n are all continuous random variables.

As with random variables and vectors, we shall no longer use the term continuous when referring to a random matrix.

1.2 Expectation Vectors and Matrices

In this section we define what we mean by the expectation of a random vector or matrix. So let **Y** = $(Y_1,\ldots,Y_n)'$ be an n-dimensional random vector with joint density function $f_Y(y)$. Then the expectation of Y_i, denoted by $E(Y_i)$ or μ_i, is computed as

$$E(Y_i) = \mu_i = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} y_i f_Y(y)\, dy_1\ldots dy_n$$

provided the above integral exists.

Definition 1.2.1. Let **Y** = $(Y_1,\ldots,Y_n)'$ be a random vector. Then the expectation vector of **Y**, denoted by $E(\mathbf{Y}) = \mu_Y = \mu$, is defined as

$$E(\mathbf{Y}) = (E(Y_1),\ldots,E(Y_n))' = (\mu_1,\ldots,\mu_n)'$$

provided all expectations exist.

We extend the definition of an expectation vector to the expectation of a random matrix.

Definition 1.2.2. Let **W** = $(W_{ij})_{mxn}$ be a random matrix. The expectation matrix of **W**, denoted by $E(\mathbf{W})$, is defined to be

$$E(\mathbf{W}) = [E(W_{ij})]_{mxn}$$

provided all expectations exist.

The expectation operator associated with random vectors and matrices has some properties that are useful in connection with studying linear models. Some of these properties are given in the following theorems and corollaries.

Theorem 1.2.3. Let $A = (a_{ij})_{lxm}$, $B = (b_{ij})_{nxp}$, and $C = (c_{ij})_{lxp}$ be matrices of real numbers and let $\mathbf{Z}=(Z_{ij})_{mxn}$ be a random matrix. Then $E(A\mathbf{Z}B+C) = AE(\mathbf{Z})B+C$.

Proof. Let $\mathbf{W} = A\mathbf{Z}B + C = (W_{ij})_{lxp}$. Then $W_{ij} = \sum_{r=1}^{m}\sum_{s=1}^{n} a_{ir}Z_{rs}b_{sj}+ c_{ij}$. Thus

$$E(\mathbf{W}) = (E(W_{ij}))_{lxp} = (E[\sum_{r=1}^{m}\sum_{s=1}^{n} a_{ir}Z_{rs}b_{sj} + c_{ij}])_{lxp}$$

$$= ([\sum_{r=1}^{m}\sum_{s=1}^{n} a_{ir}E(Z_{rs})b_{sj} + c_{ij}])_{lxp}$$

$$= [(AE(\mathbf{Z})B)_{ij}]_{lxp} + [(c_{ij})]_{lxp} = AE(\mathbf{Z})B + C.$$

Corollary 1.2.4. If $\mathbf{X} = (X_1,\ldots,X_n)'$ is an n-dimensional random vector and $A = (a_{ij})_{mxn}$ and $C= (c_i)_{mx1}$ are matrices of real numbers, then $E(A\mathbf{X} + C) = AE(\mathbf{X})+C$.

Proof. Let $B = I_n$ in theorem 1.2.3 above.

Example 1.2.5. Let $\mathbf{Y} = (Y_1,\ldots,Y_n)'$ be an n-dimensional random vector where the Y_i's are independent random variables. Let $E(\mathbf{Y}) = 1_n\mu$ where μ is an unknown parameter. Then we can write $E(\mathbf{Y})$ in the form $E(\mathbf{Y}) = X\mu$ where $X = 1_n$. Let $\hat{\mu} = (X'X)^{-1}X'\mathbf{Y} = \sum_{i=1}^{n} Y_i/n$. Then $E(\hat{\mu})=E[(X'X)^{-1}X'\mathbf{Y}]=(X'X)^{-1}X'E(\mathbf{Y})= (X'X)^{-1}(X'X)\mu = \mu$.

Proposition 1.2.6. Let $A = (a_{ij})_{mxn}$ and $B = (b_{ij})_{mxn}$ be matrices of real numbers and let $\mathbf{X}$ and $\mathbf{Y}$ be n-dimensional random vectors. Then $E(A\mathbf{X} + B\mathbf{Y}) = AE(\mathbf{X})+BE(\mathbf{Y})$.

Proof. Let $\mathbf{W} = (W_i)_{mx1} = A\mathbf{X} + B\mathbf{Y}$ where $W_i = \sum_{j=1}^{n} a_{ij}X_j + \sum_{j=1}^{n} b_{ij}Y_j$. Then

$$E(\mathbf{W}) = [E(W_i)]_{mx1} = [E(\sum_{j=1}^{n} a_{ij}X_j + \sum_{j=1}^{n} b_{ij}Y_j)]_{mx1}$$

$$=[\sum_{j=1}^{n} a_{ij}E(X_j) + \sum_{j=1}^{n} b_{ij}E(Y_j)]_{mx1}$$

$$= [\sum_{j=1}^{n} a_{ij}E(X_j)]_{mx1} + [\sum_{j=1}^{n} b_{ij}E(Y_j)]_{mx1} = AE(\mathbf{X}) + BE(\mathbf{Y}).$$

1.3 Covariance Matrices

Let $\mathbf{Y} = (Y_1,\ldots,Y_n)'$ be a random vector with joint density $f_Y(y)$. Assume $E(\mathbf{Y}) = (E(Y_1),\ldots,E(Y_n))' = (\mu_1,\ldots,\mu_n)'$ exists. Then the covariance between Y_i and Y_j, denoted by $cov(Y_i , Y_j) = \sigma_{ij}$, is computed as

$$Cov(Y_i,Y_j) = \sigma_{ij} = E[(Y_i - \mu_i)(Y_j - \mu_j)] = \int_{-\infty}^{\infty} \ldots\int_{-\infty}^{\infty} (y_i - \mu_i)(y_j -\mu_j)\, f_Y(y)\, dy_1\ldots dy_n$$

provided the above integral exists. Also, the variance of Y_i, denoted by $var(Y_i) = \sigma_i^2$, is defined as

$$var(Y_i) = \sigma_i^2 = cov(Y_i ,Y_i) = E[(Y_i - \mu_i)^2].$$

Definition 1.3.1. Let $\mathbf{X} = (X_1,\ldots,X_m)'$ and $\mathbf{Y} = (Y_1,\ldots,Y_n)'$ be random vectors. Then the covariance matrix between $\mathbf{X}$ and $\mathbf{Y}$, denoted by cov($\mathbf{X}$, $\mathbf{Y}$), is defined as

$$\text{cov}(\mathbf{X},\mathbf{Y}) = (\text{cov}(X_i,Y_j))_{mxn} = (\sigma_{ij})_{mxn}$$

provided all covariances exist.

Several properties associated with covariance matrices are given below.

Proposition 1.3.2. Let $\mathbf{X}$ and $\mathbf{Y}$ be mx1 and nx1 random vectors such that $E(\mathbf{X}) = \mu_X$ and $E(\mathbf{Y}) = \mu_Y$. Then

$$\text{cov}(\mathbf{X},\mathbf{Y}) = E[(\mathbf{X} - \mu_X)(\mathbf{Y} - \mu_Y)']$$

Proof. Observe that

$$\text{cov}(\mathbf{X},\mathbf{Y}) = (\text{cov}(X_i,Y_j))_{mxn} = (E[(X_i - \mu_{X_i})(Y_j - \mu_{Y_j})])_{mxn}$$

$$= E[(X_i - \mu_{X_i})(Y_j - \mu_{Y_j})]_{mxn} = E[(\mathbf{X} - \mu_X)(\mathbf{Y} - \mu_Y)'].$$

Definition 1.3.3. Let $\mathbf{Y}$ be an nx1 random vector. Then cov($\mathbf{Y}$,$\mathbf{Y}$), denoted by cov($\mathbf{Y}$)= $V = V_Y$, is called the dispersion or covariance matrix of $\mathbf{Y}$.

Thus if $\mathbf{Y} = (Y_1,\ldots,Y_n)'$ is an nx1 random vector, cov($\mathbf{Y}$) is an nxn matrix having $\text{cov}(Y_i,Y_j)$ as its off-diagonal elements for all $i \neq j$ and $\text{var}(Y_i)$ as it diagonal elements for $i=1,\ldots,n$.

Proposition 1.3.4. Let $\mathbf{Y}$ be an nx1 random vector such that $E(\mathbf{Y}) = \mu_Y$. Then

$$\text{cov}(\mathbf{Y}) = E[(\mathbf{Y} - \mu_Y)(\mathbf{Y} - \mu_Y)']$$

Proof. This follows directly from Proposition 1.3.2.

Suppose $\mathbf{Y}=(Y_1,\ldots,Y_n)'$ is a random vector. If $a \in R^n$, then $a'\mathbf{Y} = \sum_{i=1}^{n} a_iY_i$ is called a linear combination of $Y_1,\ldots,Y_n$. Random variables of the form $a'\mathbf{Y}$ are fundamental in linear model theory and it is convenient to have matrix expressions for the mean and variance of such random variables.

Suppose $\mathbf{Y}$ is an n-dimensional random vector. If $E(\mathbf{Y})=\mu$ exists and $a \in R^n$, then

$$E(a'\mathbf{Y}) = E(\sum_{i=1}^{n} a_iY_i) = \sum_{i=1}^{n} a_iE(Y_i) = \sum_{i=1}^{n} a_i\mu_i = a'\mu.$$

An expression for the variance can also be obtained whenever $cov(Y_i, Y_j)$ exists for all i,j by observing that

$$var(a'\mathbf{Y}) = var(\sum_{i=1}^{n} a_iY_i) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_ia_jcov(Y_i, Y_j) = a'cov(\mathbf{Y})a .$$

Proposition 1.3.5. Suppose **Y** is an n-dimensional random vector such that $E(Y) = \mu$ and $cov(\mathbf{Y})=V$ exists. Then:

(a) $cov(a'\mathbf{Y}, b'\mathbf{Y}) = a'Vb$ for all $a,b \in R^n$.

(b) V is a positive semi-definite matrix.

(c) V is the only matrix satisfying statement (a).

Proof. (a) To prove (a), observe that

$$cov(a'\mathbf{Y}, b'\mathbf{Y}) = E[(a'\mathbf{Y} - E(a'\mathbf{Y}))(b'\mathbf{Y} - E(b'\mathbf{Y}))]$$

$$=E[(a'\mathbf{Y} - a'E(\mathbf{Y}))(b'\mathbf{Y} - b'E(\mathbf{Y}))]$$

$$= E[(a'\mathbf{Y}-a'\mu)(b'Y- b'\mu)] = E[(a'(\mathbf{Y} - \mu)b'(\mathbf{Y} - \mu)]$$

$$= E[(\sum_{i=1}^{n} a_i(Y_i-\mu_i))(\sum_{j=1}^{n} b_j(Y_j-\mu_j))]$$

$$= E[(\sum_{i=1}^{n} \sum_{j=1}^{n} a_ib_j(Y_i - \mu_i)(Y_j - \mu_j)]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_ib_jE[(Y_i - \mu_i)(Y_j - \mu_j)]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_ib_j\, cov(Y_i , Y_j) = a'Vb.$$

(b) For (b), note that $cov(Y_i,Y_j) = cov(Y_j,Y_i)$ implies $V = V'$ and for any $a \in R^n$, $a'Va=var(a'\mathbf{Y})$ implies $a'Va \geq 0$.

(c) For (c), suppose G also satisfies the condition. Then $a'Vb = a'Gb$ for all $a,b \in R^n$, hence $a'(V - G)b = 0$ for all $a,b \in R^n$ which implies $V = G$.

The covariance matrix is a very useful tool for expressing variances and covariances of linear combinations of random vectors.

Example 1.3.6. Let **Y** and X = 1_n be as in Example 1.2.5 and suppose $Y_1,\dots,Y_n$ have a common variance σ^2. Clearly cov(**Y**) exists and is equal to $\sigma^2 I_n$. Thus, Proposition 1.3.2 (a) implies cov(a'**Y**,b'**Y**) = σ^2a'b for all a,b ϵ R^n. In particular, let $\hat{\mu}$ be as in Example 1.2.5. Then we have that $\hat{\mu}$ = t'**Y** where t= $X(X'X)^{-1}= n^{-1}1_n$ and

$$\text{var}(\hat{\mu}) = \text{var}(n^{-1}1_n'\mathbf{Y}) = n^{-1}1_n'\sigma^2 I_n 1_n n^{-1} = \sigma^2/n.$$

Lemma 1.3.7. Suppose **X** and **Y** are mx1 and nx1 random vectors, A and B are lxm and pxn matrices of real numbers and a and b are lx1 and px1 vectors of real constants. Then

$$\text{cov}(A\mathbf{X} + a, B\mathbf{Y} + b) = A\text{cov}(\mathbf{X},\mathbf{Y})B'.$$

Proof. Let **U** = A**X** + a and let **V** = B**Y** + b. Then by corollary 1.2.4, E(**U**) = AE(**X**)+ a and E(**V**) = BE(**Y**) + b and by proposition 1.3.2 and theorem 1.2.3,

$$\text{cov}[A\mathbf{X} + a, B\mathbf{Y} + b] = \text{cov}[\mathbf{U},\mathbf{V}] = E[(\mathbf{U} - E(\mathbf{U}))(\mathbf{V} - E(\mathbf{V}))']$$

$$= E[(A\mathbf{X} + a - AE(\mathbf{X}) - a)(B\mathbf{Y} + b - BE(\mathbf{Y}) - b))']$$

$$= E[A(\mathbf{X} - E(\mathbf{X}))(B(\mathbf{Y} - E(\mathbf{Y})))'] = E[A(\mathbf{X} - E(\mathbf{X}))(\mathbf{Y} - E(\mathbf{Y}))'B']$$

$$= AE[(\mathbf{X} - E(\mathbf{X}))(\mathbf{Y} - E(\mathbf{Y}))']B' = A\text{cov}(\mathbf{X}, \mathbf{Y})B'.$$

Corollary 1.3.8. Let **Y** be an nx1 random vector, A be an lxn matrix of real numbers and let a be an lx1 vector of real constants. If cov(**Y**) = V, then

$$\text{cov}(A\mathbf{Y} + a) = \text{cov}(A\mathbf{Y}) = AVA'.$$

Proof. By lemma 1.3.7, cov(A**Y** + a) = cov(A**Y** + a,A**Y** + a) = Acov(**Y**,**Y**)A' = Acov(**Y**)A' = AVA' = cov(A**Y**).

The covariance matrix of a random vector has many similarities with the variance of a random variable. For example, it is nonnegative in the sense of Proposition 1.3.5. As another example, if **Y** and **Z** are independent nx1 vectors, then cov $\begin{pmatrix}\mathbf{Y}\\ \mathbf{Z}\end{pmatrix}$ = $\begin{pmatrix}\text{cov}(\mathbf{Y}) & 0_{nn}\\ 0_{nn} & \text{cov}(\mathbf{Z})\end{pmatrix}$. Now, if α is a real number and c ϵ R^n, using corollary 1.3.8, we have

$$\text{cov}(\alpha\mathbf{Y} + \mathbf{Z} + c) = \text{cov}(\alpha\mathbf{Y} + \mathbf{Z}) = \text{cov}\left[(\alpha I_n, I_n)\begin{pmatrix}\mathbf{Y}\\ \mathbf{Z}\end{pmatrix}\right] = (\alpha I_n , I_n)\ \text{cov}\begin{pmatrix}\mathbf{Y}\\ \mathbf{Z}\end{pmatrix} (\alpha I_n , I_n)'$$

$$=\alpha^2 \operatorname{cov}(\mathbf{Y}) + \operatorname{cov}(\mathbf{Z}) \quad (1.3.9)$$

as long as cov(**Y**) and cov(**Z**) both exist. Additional properties of cov(.) are discussed in the problems at the end of this chapter.

Example 1.3.10. (Two variance component model) Suppose $Y_{ij} = \mu + b_i + e_{ij}$, i,j=1,2, where μ is an unknown parameter and $b_1,b_2,e_{11},\dots,e_{22}$ are independent random variables having zero means. Also assume the b_i have variance σ_b^2 and the e_{ij} have variance σ^2. Set $\mathbf{Y} = (Y_{11},Y_{12},Y_{21},Y_{22})'$, $X = (1,1,1,1)'$, $\mathbf{e} = (e_{11},e_{12},e_{21},e_{22})'$,

$$B' = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

and $\mathbf{b} = (b_1,b_2)'$. Then **Y** can be expressed in matrix form as $\mathbf{Y} = X\mu + B\mathbf{b} + \mathbf{e}$. Notice that **b** and **e** are independent random vectors, that $X\mu$ is a constant vector, that $\operatorname{cov}(\mathbf{b}) = \sigma_b^2 I_2$ and $\operatorname{cov}(\mathbf{e}) = \sigma^2 I_4$. From (1.3.9) we conclude that

$$\operatorname{cov}(\mathbf{Y}) = \operatorname{cov}(X\mu + B\mathbf{b} + \mathbf{e}) = \operatorname{cov}(B\mathbf{b} + \mathbf{e}) = \operatorname{cov}(B\mathbf{b}) + \operatorname{cov}(\mathbf{e})$$

$$= B\operatorname{cov}(\mathbf{b})B' + \sigma^2 I_4 = \sigma^2 I_4 + \sigma_b^2 BB'.$$

Also notice that $E(\mathbf{Y}) = X\mu$ since **b** and **e** both have expectation zero. In this example, σ^2 and σ_b^2 are called variance components.

The expression given for cov(**Y**) in proposition 1.3.4 in terms of a random matrix also leads to a convenient form for the expectation of a quadratic form.

Definition 1.3.11. Suppose $\mathbf{Y} = (Y_1,\dots,Y_n)'$ is a random vector and $A=A'=(a_{ij})_{nxn}$. Then

$$\mathbf{Y}'A\mathbf{Y} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} Y_i Y_j$$

Is called a quadratic form in **Y.**

Proposition 1.3.12. Let $\mathbf{Y} = (Y_1,\dots,Y_n)'$ be a random vector with $E(Y) = \mu$ and $\operatorname{Cov}(\mathbf{Y}) = V$. Let $A = A' = (a_{ij})_{nxn}$. Then

$$E(\mathbf{Y}'A\mathbf{Y}) = \mu' A \mu + \operatorname{tr} AV.$$

Proof. To begin, observe that $\mathbf{Y}'A\mathbf{Y} = (\mathbf{Y}-\mu)'A(\mathbf{Y}-\mu) + \mu'A\mathbf{Y} + \mathbf{Y}'A\mu - \mu'A\mu$. Now observe that since $A=A'$, $\mathbf{Y}'A\mu = (\mathbf{Y}'A\mu)' = \mu'A'\mathbf{Y} = \mu'A\mathbf{Y}$ and $E(\mathbf{Y}'A\mu) = E(\mu'A\mathbf{Y}) = \mu'AE(\mathbf{Y}) = \mu'A\mu$, we have that

$$E(\mathbf{Y}'A\mathbf{Y}) = E[(\mathbf{Y}-\mu)'A(\mathbf{Y}-\mu) + \mu'A\mathbf{Y} + \mathbf{Y}'A\mu - \mu'A\mu]$$

$$= E[(\mathbf{Y}-\mu)'A(\mathbf{Y}-\mu)] + E(\mu'A\mathbf{Y}) + E(\mathbf{Y}'A\mu) - E(\mu'A\mu)$$

$$= E[(\mathbf{Y}-\mu)'A(\mathbf{Y}-\mu)] + \mu'A\mu + \mu'A\mu - \mu'A\mu$$

$$= E[\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}(Y_i - \mu_i)(Y_j - \mu_j)] + \mu'A\mu$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}E(Y_i - \mu_i)(Y_j - \mu_j) + \mu'A\upsilon$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}\operatorname{cov}(Y_i, Y_j) + \mu'A\mu = \operatorname{tr}AV + \mu'A\mu.$$

Corollary 1.3.13. Let $\mathbf{W} = \mathbf{Y} - b$ where $E(\mathbf{Y}) = \mu$, $\operatorname{cov}(\mathbf{Y}) = V$ and $b \in R^n$. If $A = A'$ is a matrix of real numbers, then $\operatorname{cov}(\mathbf{W}) = \operatorname{Cov}(\mathbf{Y})$ and

$$E(\mathbf{W}'A\mathbf{W}) = \operatorname{tr}AV + (\mu - b)'A(\mu - b).$$

The problems at the end of this chapter related to this section explore other aspects of the covariance operator. We strongly suggest that the reader go through these problems and at the very least become familiar with the properties discussed.

1.4 The Multivariate Normal Distribution

In this section, we introduce the multivariate normal distribution and investigate some of its properties. For a brief review of some of the distribution theory used in this section, the reader should consult Appendix A14.

Let Z be a standard normal random variable and recall the following properties associated with Z:

(1) The probability density function for Z is, for all $z \in R^1$,

$$f(z) = (1/2\pi)^{(1/2)}\exp[(-1/2)z^2].$$

(2) $E(Z) = 0$.

(3) $Var(Z) = 1$.

(4) The moment generating function (m.g.f.) for Z is, for all $t \in R^1$,

$$M_z(t) = \exp[(1/2)t^2].$$

Now let $Z_1,\ldots,Z_n$ be mutually independent standard normal random variables and let $\mathbf{Z} = (Z_1,\ldots,Z_n)'$. Then some easily established facts concerning **Z** are the following:

(1) The joint density function for **Z** is, for all $z \in R^n$,

$$f_Z(z) = \prod_{i=1}^{n} f_{Z_i}(z_i) = \prod_{i=1}^{n} (1/2\pi)^{(1/2)}\exp((-1/2)z_i^2)$$

$$= (1/2\pi)^{(n/2)}\exp((-1/2)(\sum_{i=1}^{n} z_i^2))$$

$$= (1/2\pi)^{(n/2)}\exp((-1/2)(z'I_n z)).$$

(2) $E(\mathbf{Z}) = 0_n$.

(3) cov $(\mathbf{Z}) = I_n$.

(4) The m.g.f. for **Z** is, for all $t \in R^n$,

$$M_Z(t) = \prod_{i=1}^{n} M_{Z_i}(t_i) = \prod_{i=1}^{n} \exp[(1/2)t_i^2]$$

$$= \exp[(1/2)(t_1^2 +\ldots+ t_n^{2)})] = \exp[(1/2)t'I_n t].$$

Definition 1.4.1. We say the random vector $\mathbf{X} = (X_1,\ldots,X_n)'$ follows an n-dimensional multivariate normal distribution of rank p if **X** has the same distribution as

$$A\mathbf{Z} + b$$

where A is some nxn real matrix with $r(A) = p$, $b \in R^n$ and $\mathbf{Z} = (Z_1,\ldots,Z_n)'$ is an n-dimensional random vector whose components Z_i are independent standard normal random variables.

Proposition 1.4.2. Suppose **X** satisfies definition 1.4.1. Then

(a) $E(\mathbf{X}) = b$.

(b) $\mathrm{cov}(\mathbf{X}) = V$ where $V = AA'$ and $r(V) = r(A) = p$.

Proof. Since **X** has the same distribution as $A\mathbf{Z} + b$, it follows that $E(\mathbf{X}) = E(A\mathbf{Z}+ b)$ and that $\mathrm{cov}(\mathbf{X}) = \mathrm{cov}(A\mathbf{Z} + b)$. These results now follow after applying corollary 1.2.4 and corollary 1.3.8 to $A\mathbf{Z} + b$.

If **X** satisfies definition 1.4.1, we denote it by "$\mathbf{X}\sim N_n(b,V)$ of rank p" where $V = AA'$ and $p = r(A) = r(V)$. If $V > 0$, we will generally omit the rank portion of the preceding statement. We note that if **Z** is as in definition 1.4.1, then $\mathbf{Z}\sim N(0_n, I_n)$. We now investigate some of the properties associated with multivariate normal distributions.

Proposition 1.4.3. An nx1 random vector $\mathbf{X}\sim N_n(b,V)$ of rank p if and only if its m.g.f. has the form

$$M_{\mathbf{X}}(t) = \exp(t'b + (1/2)t'Vt)$$

where $b \in R^n$, $V \geq 0$ and $r(V) = p$

Proof. Suppose $\mathbf{X}\sim N_n(b, V)$. Then **X** satisfies definition 1.4.1 and has the same distribution as $A\mathbf{Z} +b$ where A is an nxn real matrix of rank p, $b \in R^n$ and **Z** is an n-dimensional random vector whose components are independent standard normal random variables. Since **Z** has a m.g.f, so does $A\mathbf{Z} + b$ and since **X** and $A\mathbf{Z} + b$ have the same distributions, they have the same m.g.f. But the m.g.f. of **Z** is

$$M_{\mathbf{Z}}(t) = E(\exp(t'\mathbf{Z})) = \exp((1/2)t'I_nt).$$

Thus,

$$M_{\mathbf{X}}(t) = E[\exp(t'\mathbf{X})] = M_{A\mathbf{Z}+b}(t) = E[\exp(t'(A\mathbf{Z}+b))] = E[\exp(t'A\mathbf{Z} + t'b)]$$

$$= \exp(t'b)E[\exp((A't)'\mathbf{Z})] = \exp(t'b)M_{\mathbf{Z}}(A't)$$

$$= \exp(t'b)\exp[(1/2)t'AA't] = \exp[t'b+(1/2)t'Vt]$$

where V=AA'.

Conversely, suppose X has m.g.f. $M_X(t)$ = exp[t'b + (1/2)t'Vt] where r(V)= p. Since V ≥0, we can find an nxn real matrix A of rank p such that V = AA'. Now, as in the proof above, it follows that the mgf of A**Z** + b is

$$\exp[t'b + (1/2)t'Vt]$$

where V = AA', the same as **X**. Because the m.g.f. uniquely determines the distribution (when the m.g.f. exists in an open n-dimensional neighborhood containing 0_n), **X** has the same distribution as A**Z** + b.

Proposition 1.4.4. Let $\mathbf{X} \sim N_n(\mu, V)$, let C be a qxn real matrix and let a ϵ R^q. Then

$$\mathbf{Y} = C\mathbf{X} + a \sim N_q(C\mu + a,\ CVC') \text{ of } r(CVC').$$

Proof. Observe that

$$M_Y(t) = E[\exp(t'\mathbf{Y})] = E[\exp(t'(C\mathbf{X} + a))] = E[\exp(t'C\mathbf{X} + t'a)]$$

$$= \exp(t'a)E[\exp((C't)'\mathbf{X})]$$

$$= \exp(t'a)M_X(C't) = \exp(t'a)\exp[(C't)'\mu + (1/2)(C't)'V(C't)]$$

$$= \exp[(C't)'\mu + t'a + (1/2)t'CVC't]$$

$$= \exp[t'(C\mu + a) + (1/2)t'CVC't].$$

Now observe that this last expression for $M_Y(t)$ is the same as that of a random vector $Y \sim N_q(C\mu + a, CVC')$ of r(CVC').

Corollary 1.4.5. Let $\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim N_n(\mu, V)$ where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are n_1x1 and n_2x1 random vectors, respectively. Correspondingly, let

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ and } V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

where $E(\mathbf{Y}_1) = \mu_1$, $E(\mathbf{Y}_2) = \mu_2$, $cov(\mathbf{Y}_1) = V_{11}$, $cov(\mathbf{Y}_2) = V_{22}$ and $cov(\mathbf{Y}_1,\mathbf{Y}_2) = V_{12} = V_{21}'$. Then $\mathbf{Y}_1 \sim N_{n_1}(\mu_1, V_{11})$.

Proof. In Proposition 1.4.4, take $C=(I_{n_1}, 0_{n_1n_2})$.

Proposition 1.4.6. Suppose $\mathbf{Y}\sim N_n(\mu,V)$ where $\mathbf{Y}$, μ and V are partitioned as in corollary 1.4.5. Then $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent if and only if $cov(\mathbf{Y}_1,\mathbf{Y}_2) = V_{12} = 0_{n_1n_2}$.

Proof. By Corollary 1.4.5, $\mathbf{Y}_1 \sim N_{n_1}(\mu_1,V_{11})$ and $\mathbf{Y}_2 \sim N_{n_2}(\mu_2,V_{22})$ and $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are mutually independent if and only if $M_Y(t) = M_{Y_1}(t_1)M_{Y_2}(t_2)$. Now, observe that by Proposition 1.4.3,

$M_{Y_1}(t_1) = \exp[t_1'\mu_1 + (1/2)t_1'V_{11}t_1]$ and $M_{Y_2}(t_2) = \exp[t_2'\mu_2 + (1/2)t_2'V_{22}t_2]$ and

$$M_Y(t) = M_Y\left(\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}\right) = \exp\left[(t_1', t_2')\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + (1/2)(t_1', t_2')\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}\right]$$

$$= \exp[t_1'\mu_1 + t_2'\mu_2 + (1/2)(t_1'V_{11}t_1 + t_1'V_{12}t_2 + t_2'V_{21}t_1 + t_2'V_{22}t_2)].$$

Therefore, $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent if and only if $t_1'V_{12}t_2 + t_2'V_{21}t_1 = 0$ for all possible values of t_1 and t_2. But since $t_2'V_{21}t_1 = (t_2'V_{21}t_1)' = t_1'V_{12}t_2 = (t_1'V_{12}t_2)'$, the condition is that $t_1'V_{12}t_2 = 0$ for all t_1 and t_2, i.e., that $V_{12} = 0_{n_1n_2}$.

Proposition 1.4.7. Suppose $\mathbf{Y}\sim N_n(\mu,V)$ where $V>0$. Then

$$f_Y(y) = (2\pi)^{-(n/2)}|V|^{-(1/2)}\exp[-(1/2)(y - \mu)'V^{-1}(y - \mu)]$$

for all $y \in R^n$ and where $|V|$ denotes the determinant of V.

Proof. Since $\mathbf{Y}\sim N_n(\mu,V)$, it has the same distribution as $A\mathbf{Z} + \mu$ where $\mathbf{Z}=(Z_1,\ldots,Z_n)'$ and the Z_i are all independent standard normal random variables, $\mu \in R^n$, A is an nxn real matrix with $r(A) = n$ and $V = AA'$. Now recall that, for all $z \in R^n$,

$$f_Z(z) = \prod_{i=1}^{n} f_{Z_i}(z_i) = \prod_{i=1}^{n} (2\pi)^{-(1/2)}\exp[-(1/2)z_i^2]$$

$$=(2\pi)^{-(n/2)}\exp[(1/2)z'I_nz].$$

As indicated in Appendix A14

$$f_Y(y) = (1/|A|\quad)f_Z(A^{-1}(y-\mu))$$

$$= (2\pi)^{-(n/2)}(1/|A|\quad)\exp[-(1/2)(y-\mu)'A^{-1'}I_nA^{-1}(y-\mu)]$$

$$= (2\pi)^{-(n/2)}\,|V|^{-(1/2)}\exp[-(1/2)(y-\mu)'V^{-1}(y-\mu)].$$

Note. When n = 1, the expression given above for the density function of **Y** reduces to the density function of a 1-dimensional normal random variable.

1.5 The Chi-Squared Distribution

In this section we consider properties of central and non-central chi-squared distributions which play fundamental roles in tests of hypotheses in linear models.

Definition 1.5.1. Let $\mathbf{Y} \sim N_v(\mu, I_v)$. Then the random variable $\mathbf{Y'Y} = \sum_{i=1}^{v} Y_i^2$ is said to follow a non-central chi-squared distribution with v degrees of freedom (d.f.) and non-centrality parameter $\lambda = \mu'\mu$.

The reader should observe that when $\lambda = \mu'\mu = 0$, we get the well - known central chi-squared distribution with v d.f which can be expressed as the sum of v squared independent standard normal random variables. We shall use notation such as $X \sim \chi^2(v,\lambda)$ to indicate that the random variable X follows a chi-squared distribution with v d.f. and non-centrality parameter λ and use the notation $X \sim \chi^2(v)$ to denote a random variable X which follows a central chi-squared distribution with v d.f..

Proposition 1.5.2. If $Y \sim \chi^2(v,\lambda)$, then the moment generating function for Y is $M_Y(t) = (1/(1-2t))^{(v/2)}\exp[\lambda t/(1-2t)]$ for all $t<1/2$.

Proof. Let $Z \sim N(0,1)$ and let $\mu \in R^1$. Then by definition 1.5.1, $(Z+\mu)^2 \sim \chi^2(1,\lambda)$ where $\lambda = \mu^2$. The m.g.f. of $(Z+\mu)^2$ is

$$M_{(Z+\mu)^2}(t) = E[\exp(t(z+\mu)^2)] = \int_{-\infty}^{\infty} \exp[t(z+\mu)^2]\, f_Z(z)\, dz$$

$$= (2\pi)^{-(1/2)} \int \exp\,[t(z + \mu)^2] \exp[-z^2/2]\, dz$$

$$= (2\pi)^{-(1/2)} \int \exp\,[-((1/2) - t)z^2 + 2\mu tz + \mu^2 t] dz$$

$$= (2\pi)^{-(1/2)} \int \exp\,[-((1/2) - t)(z - Q)^2 + \mu^2 t + (2\mu^2 t^2/(1 - 2t))] dz$$
(with $Q = 2\mu t/(1 - 2t)$)

$$= \exp[\mu^2 t + (2\mu^2 t^2/(1 - 2t))]\, (2\pi)^{-(1//2)} \int \exp\,[-(z - Q)^2/2(1 - 2t)^{-1}]\, dz$$

$$= \exp[\mu^2 t + (2\mu^2 t^2/(1 - 2t))](1 - 2t)^{-(1/2)}$$
$$\times\, (2\pi)^{-(1/2)}(1 - 2t)^{(1/2)} \int \exp\,[-(z - Q)^2/2(1 - 2t)^{-1}]\, dz$$

$$= \exp[\mu^2 t + (2\mu^2 t^2/(1 - 2t))](1 - 2t)^{-(1/2)} \times 1 = \exp[\mu^2 t/(1 - 2t)](1 - 2t)^{-(1/2)}.$$

Thus again by definition 1.5.1, if $X_i = (Z_i + \mu_i)^2$, $i = 1,\ldots,v$ and the Z_i's are all independent standard normal random variables, then a $\chi^2(v,\lambda)$ random variable having $\lambda = \sum_{i=1}^{v} \mu_i^2$ is given by $\chi^2(v,\lambda) = \sum_{i=1}^{v} X_i = \sum_{i=1}^{v} (Z_i+\mu_i)^2$. Because all terms are mutually independent and using the expression for $M_{(z+\mu)2}(t)$ obtained above, we have that

$$M_{\chi2(v,\lambda)}(t) = \prod_{i=1}^{v} M_{(Z_i+\mu_i)^2}(t) = \prod_{i=1}^{v} (1 - 2t)^{-(1/2)} \exp[\mu_i^2 t/(1 - 2t)]$$

$$= (1 - 2t)^{-(v/2)} \exp[(t/(1 - 2t)) \sum_{i=1}^{v} \mu_i^2] = (1 - 2t)^{-(v/2)} \exp\,[t\lambda/(1 - 2t)]$$

where $\lambda = \sum_{i=1}^{v} \mu_i^2$. Since $M_{\chi2(v,\lambda)}(t)$ has the form of a r.v. $W \sim\chi^2(v,\lambda)$, we have the desired result.

Note: (1) If $Y\sim\chi^2(v,\lambda)$, one can use the m.g.f. given in Proposition 1.5.2 to show that $E(Y) = v + \lambda$ and that $var(Y) = 2v + 4\lambda$.

(2) When $\lambda = 0$, the m.g.f. of the non-central chi-squared distribution reduces to that of $Y\sim\chi^2(v)$ which is $M_Y(t) = (1/(1 - 2t))^{(v/2)}$.

Proposition 1.5.3. Suppose $Q_1,\ldots,Q_r$ are mutually independent and $Q_i\sim\chi^2(v_i,\lambda_i)$ for $i = 1,\ldots,r$. If $Q = \sum_{i=1}^{r} Q_i$, then $Q\sim\chi^2(v,\lambda)$ where $v = \sum_{i=1}^{r} v_i$ and $\lambda = \sum_{i=1}^{r} \lambda_i$.

Proof. Since $Q = Q_1+\ldots+Q_r$ and the Q_i are all independent,

$$M_Q(t) = \prod_{i=1}^{r} M_{Q_i}(t) = \prod_{i=1}^{r} (1/(1\text{-}2t))^{(v_i/2)} \exp[t\lambda_i/(1 - 2t)]$$

$$= (1/(1 - 2t))^{(v/2)} \exp[t\lambda/(1 - 2t)]$$

where $v = \sum_{i=1}^{r} v_i$ and $\lambda = \sum_{i=1}^{r} \lambda_i$. Since $M_Q(t)$ has the form of a $\chi^2(v,\lambda)$ r.v., we have the desired result.

1.6 Quadratic Forms in Normal Random Vectors

Suppose $\mathbf{X} = (X_1,\ldots,X_n)'$ is a random vector and $A = A' = (a_{ij})_{nxn}$ is a matrix of real constants. Then, as defined in Section 1.3, $\mathbf{X}'A\mathbf{X} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}X_iX_j$ is called a quadratic form. In this section, we consider quadratic forms in which $\mathbf{X}\sim N_n(\mu, V)$ as well as the properties of such quadratic forms.

Proposition 1.6.1. Suppose $\mathbf{U}\sim N_p(\mu,\Sigma)$ and let $A = A'$ satisfy $A\Sigma A = \pi A$ where $\pi > 0$ is a scalar. Then

$$(1/\pi)\, \mathbf{U}'A\mathbf{U}\sim\chi^2(r(A),\lambda)$$

where $\lambda = (1/\pi)\mu' A\mu$

Proof. Let $s = r(A)$ and observe that $A\geq 0$. So by theorem A10.12, we can find a pxs matrix H such that $A = HH'$ and $r(H) = s$. Then

$$\mathbf{U}'A\mathbf{U} = \mathbf{U}'HH'\mathbf{U} = (H'\mathbf{U})'(H'\mathbf{U}) = \mathbf{Y}'\mathbf{Y}$$

where $\mathbf{Y} = H'\mathbf{U}\sim N_s(H'\mu, H'\Sigma H)$. Let $L = (H'H)^{-1}H'$ and observe that

$$H'\Sigma H = (H'H)^{-1}H'HH'\Sigma H(H'H)(H'H)^{-1} = LA\Sigma AL' = L\pi AL'$$

$$= \pi(H'H)^{-1}H'HH'H(H'H)^{-1} = \pi I_s$$

which implies $\mathbf{Y}=H'\mathbf{U}\sim N_s(H'\mu,\pi I_s)$ and $(1/\pi)^{(1/2)}\mathbf{Y}\sim N_s((1/\pi)^{(1/2)}H'\mu,I_s)$. Hence, by definition 1.5.1,

$$(1/\pi)^{(1/2)}\mathbf{Y}'(1/\pi)^{(1/2)}\mathbf{Y} = (1/\pi)\mathbf{Y}'\mathbf{Y} = (1/\pi)\mathbf{U}'HH'\mathbf{U} = (1/\pi)\mathbf{U}'A\mathbf{U}\sim\chi^2(s,\lambda)$$

where $s = r(A)$ and $\lambda = (1/\pi)^{(1/2)}\mu' HH'\mu(1/\pi)^{(1'2)} = (1/\pi)\mu' A\mu$.

Corollary 1.6.2. Let $\mathbf{Y}\sim N_n(\mu,\pi I_n)$ where $\pi > 0$. If $A = A' = A^2$, then

$$(1/\pi)\mathbf{Y}'A\mathbf{Y}\sim\chi^2(k,\lambda)$$

where $k = trA = r(A)$ and $\lambda = (1/\pi)\mu' A\mu$.

Proof. This follows from proposition 1.6.1 since $A\pi I_n A = \pi A^2 = \pi A$ and from proposition A11.11 since $A^2 = A$ implies $k = trA = r(A)$.

Proposition 1.6.3. Let $A = A'$ and $V = BB'$ where B is an nxn matrix. Set $T = B'AB$. Then the following statements hold:

(a) $trT = tr(VA)$.

(b) If $T^2 = T$, then $tr\, T = r(T)$.

(c) $r(T) = r(VAV)$.

(d) $T^2 = T$ if and only if $V(AVA-A)V = 0_{nn}$.

Proof. (a) By proposition A10.14 $trT = tr(B'AB) = tr(ABB') = tr(AV) = tr(VA)$.

(b) This is part of proposition A11.11.

(c) Using proposition A8.2 and proposition A6.2, we have that

$$r(VAV) = r(BB'AV) = r(B'AV) - \dim(R(B'AV) \cap N(B))$$

$$= r(B'AV) - \dim(R(B'AV) \cap R(B')^{\perp}) = r(B'AV) - 0$$

$$= r(VAB) = r(BB'AB) = r(B'AB) - \dim[R(B'AB) \cap N(B)]$$

$$= r(B'AB) - \dim[R(B'AB) \cap R(B')^{\perp}] = r(B'AB) - 0 = r(T).$$

(d) Assume $T^2 = T$. Then $(B'AB)(B'AB) = B'AB$ and $BB'ABB'ABB' = BB'ABB'$ which implies that $VAVAV = VAV$ which yields the desired result. Conversely, assume $VAVAV = VAV$. Then $BB'AVAV-BB'AV = 0_{nn}$ which implies that $B[B'AVAV-B'AV] = 0_{nn}$ and

$$R(B'AVAV-B'AV) \subset N(B) \cap R(B') = R(B')^{\perp} \cap R(B') = 0_n.$$

Hence $B'AVAV - B'AV = 0_{nn}$. But this implies that

$$VAVAB - VAB = BB'AVAB - BB'AB = B(B'AVAB-B'AB) = 0_{nn}.$$

Hence, as above, that

$$R(B'AVAB - B'AB) \subset N(B) \cap R(B') = R(B')^{\perp} \cap R(B') = 0_n$$

and that $B'AVAB - B'AB = B'ABB'AB - B'AB = T^2 - T = 0_{nn}$.

Proposition 1.6.4. Suppose $\mathbf{Y} \sim N_v(\mu, \pi V)$ where $\pi > 0$, $V \geq 0$ and $\mu \in R(V)$. If $A = A'$ and $V(AVA-A)V = 0_{vv}$, then $(1/\pi)\mathbf{Y}'A\mathbf{Y} \sim \chi^2(k,\lambda)$ where $k = tr(VA)$ and $\lambda = (1/\pi)\mu' A\mu$.

Proof. Since $V \geq 0$, by theorem A10.10, we can find a vxv matrix B such that $V = BB'$. Set $T = B'AB$. Then $T = T'$ and by proposition 1.6.3, it follows that $T^2 = T$. Also, since $\mu \in R(V) = R(BB') = R(B)$, we can find δ such $\mu = B\delta$. Let $\mathbf{X} \sim N_v(\delta, \pi I_n)$. Then $B\mathbf{X} \sim N_v(B\delta, \pi BB') = N_v(\mu, \pi V)$. Now let $\mathbf{Y} = B\mathbf{X}$. Then, $(1/\pi)\mathbf{Y}'A\mathbf{Y} = (1/\pi)\mathbf{X}'B'AB\mathbf{X} = (1/\pi)\mathbf{X}'T\mathbf{X}$ where $T = T' = T^2$. Thus, by corollary 1.6.2,

$$(1/\pi)\mathbf{Y}'A\mathbf{Y} = (1/\pi)\mathbf{X}'T\mathbf{X} \sim \chi^2(k,\lambda)$$

where $k = tr(T) = tr(VA)$ and $\lambda = (1/\pi)\delta' T\delta = (1/\pi)\delta' B'AB\delta = (1/\pi)\mu' A\mu$.

Corollary 1.6.5. The preceding proposition remains true if $A = A' \neq 0_{vv}$ and $(VA)^2 = VA$.

Proof. Since $(VA)^2 = VA$, we have that $VAVAV = VAV$ and the result follows by proposition 1.6.4.

Proposition 1.6.6. Suppose $\mathbf{Y} \sim N_p(\mu, V)$ where $V \geq 0$. Let $Q_i = \mathbf{Y}'A_i\mathbf{Y}$ for $i = 1,\ldots,t$ where $A_i = A_i'$ and let $\mathbf{U_0} = B'\mathbf{Y}$ where B is a pxs matrix. If $A_iVA_j = 0_{pp}$ and $A_iVB = 0_{ps}$ for all $i \neq j$, $i,j = 1,\ldots,t$, then $\mathbf{U_0}, Q_1,\ldots,Q_t$ are all mutually independent.

Proof. For $k = 1,\ldots,t$ observe that $Q_i = \mathbf{Y}'A_i\mathbf{Y} = \mathbf{Y}'A_iA_i^{-}A_i\mathbf{Y} = \mathbf{U}_i'A^{-}\mathbf{U}_i$ where $\mathbf{U_i} = A_i\mathbf{Y}$, thus each Q_i is a function of $\mathbf{U_i}$ for $i=1,\ldots,t$. So to show that the Q_i are all independent, it is enough to show that the $\mathbf{U_i}$ are all independent. To this end, let $\mathbf{U_0}$ be as defined above and let $\mathbf{U_k} = A_k\mathbf{Y}$ for $k=1,\ldots,t$. Now consider the matrix $M=(B,A_1',\ldots,A_t')'$ and let $\mathbf{W} = M\mathbf{Y} = (\mathbf{Y}'B,\mathbf{Y}'A_1',\ldots,\mathbf{Y}'A_t')'$. By proposition1.4.4, $W = (\mathbf{U_0}',\mathbf{U}_1',\ldots,\mathbf{U}_t')' \sim N_n(M\mu, MVM')$ where $n = tp + s$. But also from proposition 1.4.6, we know that the $\mathbf{U_i}$, $i = 0,\ldots,t$ are independent if and only if $cov(\mathbf{U}_i,\mathbf{U}_j) = 0_{pp}$ for all $i \neq j$, $i,j = 1,\ldots,t$ and $cov(\mathbf{U}_0,\mathbf{U}_i) = 0_{sp}$ for $i = 1,\ldots,t$. But by corollary 1.3.7, $cov(\mathbf{U}_i,\mathbf{U}_j) = cov(A_i\mathbf{Y},A_j\mathbf{Y}) = A_iVA_j$, $i \neq j$, $i,j = 1,\ldots,t$ and $cov(\mathbf{U}_0,\mathbf{U_i}) = cov(B'\mathbf{Y},A_i\mathbf{Y}) = B'VA_i$ for $i = 1,\ldots t$. Thus by proposition 1.4.6, the $\mathbf{U}_i$, i = 0,…,t are all independent if and only if the conditions of the proposition hold.

We now momentarily diverge and give two lemmas concerning linear algebra which are needed to prove the main result of this section.

Lemma 1.6.7. Suppose H and G are nxn symmetric matrices such that $r(H+G) = r(H) + r(G)$ and $(H + G)^2 = H + G$. Then $H^2 = H$, $G^2 = G$ and $HG = GH = 0_{nn}$.

Proof. Note that

$$r(H) + r(G) = r(H + G) = \dim R(H + G) \leq \dim[R(H) + R(G)]$$

$$= \dim[R(H)] + \dim[R(G)] - \dim[R(H)\cap R(G)] \leq r(H) + r(G)$$

which implies that $\dim[R(H)\cap R(G)] = 0$, hence that $R(H)\cap R(G) = 0_n$. Now consider $(H + G)^2 = H^2+HG+GH+G^2 = H+G$ which we can rewrite as

$$H^2 + HG - H = G - GH - G^2.$$

This latter expression implies that $R(H^2+HG-H) \subset R(H)\cap R(G) = 0_n$ which implies that $H^2 - H = HG$ and that, since H and G are symmetric, $H^2 - H = GH$. Thus $R(H^2-H) \subset R(H)\cap R(G) = 0_n$ and $H^2-H = 0_{nn}$. But we also have $R(GH) \subset R(H)\cap R(G) = 0_n$, hence $GH = 0_{nn}$. Similarly, it can be shown that $G^2 = G$ and $HG = 0_{nn}$.

Lemma 1.6.8. Suppose $A_1,\ldots,A_t$ are nxn symmetric matrices with ranks $r_1,\ldots,r_t$, respectively. If $I_n = A_1+\ldots+A_t$, then the following are equivalent:

(a) $A_iA_j = 0_{nn}$ for all $i \neq j$.

(b) $A_i^2 = A_i$ for i = 1,…,t.

(c) $\sum_{i=1}^{t} r_i = n$.

Proof. (a) ⇒ (b) Note that $A_iI_n = A_i(A_1+\ldots+A_t) = A_i^2$, hence $A_i^2 = A_i$.

(b) ⇒ (c) By lemma A11.11, $A_i^2 = A_i$ implies that $trA_i = r_i$. So

$$n = tr(I_n) = tr(\sum_{i=1}^{t} A_i) = \sum_{i=1}^{t} tr(A_i) = \sum_{i=1}^{t} r_i.$$