Chapter 1

Projections

Projections are special matrices (linear transformations) that are extremely useful in linear models. Most of the theory of linear models has to do with projections.

1.1 General Definition of a Projection

Suppose $\mathcal{M}$ is a vector space and N_1 and N_2 are two subspaces in $\mathcal{M}$, where $N_1 + N_2 = \mathcal{M}$ and $N_1 \cap N_2 = 0$. Consider the unique decomposition $z = x + y$, where $x \in N_1$, and $y \in N_2$. The linear transformation

$$P_{N_1|N_2} z = x$$

is called the projection of z onto the subspace N_1 along the subspace N_2.

Note: The vector x is the result of the projection of z.

Theorem 1.1 *The projection operator onto N_2 along N_1 is given by*

$$P_{N_2|N_1} = I - P_{N_1|N_2} \, .$$

In this course, we will be concerned with projections in $\mathcal{M} = R^n$.

Remark 1.1 1. *If the projection is at a "right angle", then it is called an orthogonal projection and such projections are unique.*

2. *If the projection is not at a right angle, then it is called a projection. There can be many different projections onto a certain subspace.*

Definition 1.1 *Let A be an $n \times n$ matrix. A is said to be a projection operator (matrix) onto $C(A)$ along $C(A)^c$ if for any $v \in C(A)$,*

$$Av = v \ .$$

Definition 1.2 *If $A^2 = A$, then A is said to be an idempotent matrix.*

Theorem 1.2 *$A^2 = A$ if and only if A is a projection matrix.*

Definition 1.3 *M is a perpendicular (orthogonal) projection operator (matrix) onto $C(X)$ if and only if*

$$\begin{array}{ll} i) & v \in C(X) \Rightarrow Mv = v \ \ (\textit{projection}) \\ ii) & w \in C(X)^{\perp} \Rightarrow Mw = 0 \ \ (\textit{orthogonal}). \end{array}$$

Theorem 1.3 *If M is an orthogonal projection operator onto $C(X)$, then $C(M) = C(X)$.*

Proof: We must show that $C(M) \subset C(X)$ and $C(X) \subset C(M)$.

1) $\Rightarrow C(X) \subset C(M)$.
Let $v \in C(X)$, then $Mv \in C(M)$, since $C(M) = \{z : Mt = z, t \in R^n\}$. But $Mv = v$ since M is a projection, thus $v \in C(M)$.
2) $\Leftarrow C(M) \subset C(X)$.

Let $v \in C(M)$. Then there exists a t such that $Mt = v$. Write $t = t_1 + t_2$, where $t_1 \in C(X)$, $t_2 \in C(X)^{\perp}$. Now $Mt = M(t_1 + t_2) = Mt_1 + Mt_2 = v$. Since M is an orthogonal projection operator onto $C(X)$, $Mt_2 = 0$, and thus $Mt_1 = v$. Since $t_1 \in C(X)$, $Mt_1 = t_1$, and this implies $t_1 = v$. Thus for $v \in C(M)$, we have $Mv = v$ which implies $v \in C(X)$.

Theorem 1.4 *M is an orthogonal projection operator onto $C(M)$ if and only if $M = M^2$ and $M = M'$. Thus a matrix is an orthogonal projection operator if it is idempotent and symmetric.*

Proof: 1) " $\Rightarrow$ " Suppose M is an orthogonal projection operator. We want to show that $M^2 = M$ and $M = M'$.
Let $v \in R^n$, and write $v = v_1 + v_2$, where $v_1 \in C(M)$, $v_2 \in C(M)^\perp$. Then

$$\begin{aligned} M^2 v &= M^2(v_1 + v_2) \\ &= M^2 v_1 + M^2 v_2 = M(Mv_1) + M(Mv_2) \\ &= Mv_1 = Mv_1 + Mv_2 = M(v_1 + v_2) = Mv. \end{aligned}$$

Thus $M^2 v = Mv$ for any $v \in R^n$. This implies that $(M^2 - M)v = 0$ for any $v \in R^n$, and thus $M^2 - M = 0 \Rightarrow M^2 = M$.
To see that $M = M'$, let $w = w_1 + w_2$, where $w_1 \in C(M)$, $w_2 \in C(M)^\perp$. Also write $v = v_1 + v_2$, where $v_1 \in C(M)$ and $v_2 \in C(M)^\perp$. So

$$\begin{aligned} (I - M)v &= (I - M)(v_1 + v_2) \\ &= v_1 + v_2 - (Mv_1 + Mv_2) \\ &= v_1 + v_2 - v_1 - Mv_2 \\ &= (I - M)v_2 = v_2\ . \end{aligned}$$

Similarly, $Mw = M(w_1 + w_2) = Mw_1 + Mw_2 = w_1$. Thus we get $w'M'(I - M)v = w_1' v_2 = 0$. This is true for any v and w, so that $M'(I - M) = 0$, which implies $M' = M'M$. Since $M'M$ is symmetric, M' must be symmetric, hence M is symmetric, i.e., $M = M'$.

2) " $\Leftarrow$ " If $M^2 = M$ and $v \in C(M)$, then since $v = Mb$ we have $Mv = M\,Mb = Mb = v$. If $M' = M$ and $w \perp C(M)$, then $Mw = M'w = 0$ because the columns of M are in $C(M)$.

Theorem 1.5 *Orthogonal projection operators are unique.*

Proof: Let M and P be two orthogonal projection operators onto some space $\mathcal{M}_1$. Let $v \in R^n$ and write $v = v_1 + v_2$, $v_1 \in \mathcal{M}_1$ and $v_2 \in \mathcal{M}_1^\perp$.

$$Mv = M(v_1 + v_2) = Mv_1 + Mv_2 = Mv_1 = v_1,$$

since $v_1 \in \mathcal{M}_1$ and M is an orthogonal projection operator onto $\mathcal{M}_1$.

Now it must be the case that $Pv = P(v_1 + v_2) = Pv_1 = v_1$, and this implies $Mv = Pv$, which implies $(M - P)v = 0$ for any $v \in R^n$. Thus $M = P$.

Note: Projection operators are not unique in general.

1.2 Illustration of orthogonal projection in linear models

Consider the linear model

$$Y = X\beta + \varepsilon \ ,$$

where $E(\varepsilon) = 0$, and $\text{Cov}(\varepsilon) = \sigma^2 I$. We have $E(Y) = \mu = X\beta$. If X has full rank (i.e., $r(X) = p$), then $X'X$ is invertible and the least squares estimator of β is $\hat{\beta} = (X'X)^{-1}X'Y$. The least squares estimator of $\mu = X\beta$ is $\hat{\mu} = X\hat{\beta} = X(X'X)^{-1}X'Y = MY$, where $M = X(X'X)^{-1}X'$. $\hat{\mu}$ is the orthogonal projection of Y onto $C(X)$.

Theorem 1.6 *Suppose M is an $n \times n$ orthogonal projection operator of rank $r \leq n$. Then*

1) The eigenvalues of M are 0 or 1.

2) $r(M) = tr(M) = r$.

3) M is a positive semidefinite matrix.

Proof of 1:) Since M is an orthogonal projection operator, $M = M^2$ and $M = M'$. Let λ be an eigenvalue of M with eigenvector x. Thus $Mx = \lambda x$ and $M^2x = M(Mx) = M(\lambda x) = \lambda Mx = \lambda^2 x$. But $M = M^2$ implies $Mx = M^2x$, which implies $\lambda x = \lambda^2 x \Rightarrow (\lambda - \lambda^2)x = 0 \Rightarrow \lambda(1 - \lambda) = 0 \Rightarrow \lambda = 0$ or $\lambda = 1$.

Proof of 2:) $r(M) = r$. By definition,

$$\begin{aligned} tr(M) &= \sum_{i=1}^{n} \lambda_i \\ &= \sum_{i=1}^{r} \lambda_i + \sum_{i=r+1}^{n} \lambda_i \\ &= \sum_{i=1}^{r} 1 + \sum_{i=r+1}^{n} 0 = r \end{aligned}$$

since all eigenvalues are either 0 or 1, and exactly r of them are 1.

Proof of 3)

Need to show that for any $x \neq 0$, $x'Mx \geq 0$.

From 1), we know that all the eigenvalues for M are either 0 or 1. By the Spectral Theorem, $M = P\Lambda P'$, where P is an orthogonal matrix and Λ is a diagonal matrix with the diagonal elements to be 0 or 1. Therefore, $x'Mx = x'P\Lambda P'x = (P'x)'\Lambda(P'x) \geq 0$, since the diagonal element in Λ are nonnegative. Therefore, M is positive semidefinite.

Theorem 1.7 *Suppose X is an $n \times p$ of rank $r \leq min(n, p)$. Suppose $\{a_1, \ldots, a_r\}$ is an orthonormal basis for $C(X)$. Let $A = (a_1, \ldots, a_r)$, where a_i is $n \times 1$. Then,*

$$AA' = \sum_{i=1}^{r} a_i a_i'$$

is the unique orthogonal projection operator onto $C(X)$.

Proof: Clearly AA' is symmetric. We need to show that $(AA')^2 = AA'$ and $C(AA') = C(X)$. Now $(AA')^2 = (AA')(AA') = A(A'A)A' = AI_{r\times r}A' = AA'$.

Note here that

$$A'A = \begin{pmatrix} a_1' \\ \vdots \\ a_r' \end{pmatrix}_{r\times n} (a_1, \ldots, a_r)_{n\times r}$$

$$= \begin{pmatrix} a_1'a_1 & a_1'a_2 & \dots & a_1'a_r \\ a_1'a_2 & a_2'a_2 & \dots & a_2'a_r \\ \vdots & \vdots & \ddots & \vdots \\ a_1'a_r & \dots & \dots & a_r'a_r \end{pmatrix}$$

$$= \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{pmatrix}_{r\times r} = I_{r\times r}.$$

Now we need to show $C(AA') = C(X)$.
1) ”$\Rightarrow$” Let $x \in C(AA')$. Then $x = AA't$ for some t. Let $t^* = A't$. Then $AA't = At^* = x$, and thus $x \in C(X)$, since $C(X) = \{z : At^* = z\}$. Thus $C(AA') \subset C(X)$.

2) ”$\Leftarrow$” let $x \in C(X)$, then $x \in C(A)$ since $C(A) = \mathcal{S}(a_1, \dots, a_r) = C(X)$. Now $x \in C(A)$ implies $x = At$ for some t. Since A has full rank, (i.e.,$r(A) = r$), any vector $t \in R^r$ can be written as $t = A'z$, where $z \in R^n$. Thus $x = At = AA'z$, which implies $x \in C(AA')$. Thus $C(X) \subset C(AA')$, and this completes the proof.

Theorem 1.8 *(special case) Suppose X is an $n \times p$ matrix of rank p. Define $M = X(X'X)^{-1}X'$. Then M is the orthogonal projection operator onto $C(X)$ (along $C(X)^{\perp}$). If X is $n \times 1$ then $M = \frac{XX'}{X'X}$.*

Proof: Since X has full rank p, $(X'X)^{-1}$ exists. Clearly M is symmetric since $\left(X(X'X)^{-1}X'\right)' = (X')'(X'X)^{-1'}X' = X(X'X)^{-1}X'$. Now $M^2 = (X(X'X)^{-1}X')(X(X'X)^{-1}X') = X(X'X)^{-1}(X'X)(X'X)^{-1}X' = X(X'X)^{-1}X' = M$.

Finally, we need to show that $C(M) = C(X)$. $C(X) = \{z : Xt = z\}$ and $C(M) = \left\{z^* : X(X'X)^{-1}X't = z^*\right\} = \{z^* : Xt^* = z^*\} = C(X)$. Thus M is the orthogonal projection onto $C(X)$.

Theorem 1.9 *Suppose $M = X(X'X)^{-1}X'$ as above. Then M can be written $M = QQ' = V_1V_1'$ where Q is based on the spectral decomposition as defined on p. 30, and V_1 is defined as in the SVD decomposition.*

Proof is an exercise.

Theorem 1.10 *$I - M$ is the unique orthogonal projection operator onto $C(X)^{\perp}$.*

<u>Proof:</u>

1) $(I-M)' = I - M' = I - M$, thus $I - M$ is symmetric.

2) $(I-M)(I-M) = I - M - M + M^2 = I - M - M + M = I - M$. Thus $I - M$ is idempotent.

   Therefore 1) and 2) prove that $I - M$ is an orthogonal projection operator.

3) We need to show $C(I-M) = C(X)^{\perp}$. Let $x \in C(M)$. We want to show that for any $y \in C(I-M)$, $x'y = 0$. It is enough to show that $(I-M)x = 0$, where $x \in C(M)$. $(I-M)x = x - Mx = x - x = 0$. Thus $C(X)^{\perp} = C(I-M)$.

Theorem 1.11 *Suppose X is an $n \times p$ matrix of rank p. Write $X = (X_1, X_2)$, where X_1 is $n \times k$, X_2 is $n \times (p-k)$, $r(X_1) = k$, and $r(X_2) = p - k$. Let $M_j = X_j(X_j'X_j)^{-1}X_j'$, $j = 1, 2$, and let $M = X(X'X)^{-1}X'$. Further, let*

$$X_j^* = (I - M_{3-j})X_j, \;\; j = 1, 2 \;,$$

and

$$M_j^* = X_j^*(X_j^{'*}X_j^*)^{-1}X_j^*.$$

Thus X_1^* consists of columns which are orthogonal to X_2 and X_2^* has columns which are orthogonal to X_1. Therefore, $M = M_1 + M_2^*$ and $M = M_2 + M_1^*$.

1.3 Generalized inverses

Singular matrices arise much in the theory of linear models. Suppose X is $n \times p$ of rank $r < \min(n, p)$. Then $(X'X)^{-1}$ does not exist. In these cases, we will need the notion of a generalized inverse of $X'X$ so that estimates can be computed.

Definition 1.4 *Consider the linear transformation $A : R^p \longrightarrow R^n$. A generalized inverse of A is the linear transformation A^- such that*

$$AA^- y = y \ \ \textit{for all} \ \ y \in C(A).$$

Note: Since $A : R^p \longrightarrow R^n$, $A^- : R^n \longrightarrow R^p$.

This definition is equivalent to the following.

Definition 1.5 *Suppose A is an $n \times p$ matrix, then $A^-_{p\times n}$ is a generalized inverse of A if*

$$AA^- A = A \ .$$

Recall that $y \in C(A)$ means $y = Ax$ for some $x \in R^p$. Thus, substituting this into the definition, we get

$$AA^- y = y \ , \quad y \in R^n \ ,$$

and $y = Ax$.

Note: by the definition of a generalized inverse, we have

$$(A^- A)(A^- A) = A^-(AA^- A) = A^- A \ .$$

Thus $A^- A$ is idempotent, and hence a projection.

The generalized inverse is not unique. We want to focus on a specific type of generalized inverse that satisfies additional properties.

Definition 1.6 Moore-Penrose Generalized Inverse

Suppose A is an $n \times p$ matrix. The Moore-Penrose generalized inverse (M-P g-inverse) of A is a $p \times n$ matrix A^+ such that

1) $(AA^+)' = AA^+$ *(AA^+ is symmetric)*

2) $(A^+A)' = A^+A$ *(A^+A is symmetric)*

3) $AA^+A = A$ *(A^+ is a g-inverse of A)*

4) $A^+AA^+ = A^+$ *(A is a g-inverse of A^+)*

From the definition, it is immediate that AA^+ and A^+A are orthogonal projection operators.

Theorem 1.12 *Every matrix A has a Moore-Penrose generalized inverse.*

Proof: Suppose A has rank r. From the SVD of A, we can write $A = V_1 \Delta U_1'$ where Δ is an $r \times r$ diagonal matrix with positive elements, and V_1 and U_1 have orthonormal columns. Define $A^+ = U_1 \Delta^{-1} V_1'$. We must show that A^+ satisfies the four conditions of a MP g-inverse.

1) $AA^+ = V_1 \Delta U_1' U_1 \Delta^{-1} V_1' = V_1 V_1'$ which is clearly symmetric.

2) $A^+A = U_1 \Delta^{-1} V_1' V_1 \Delta U_1' = U_1 U_1'$ which is symmetric.

3) $AA^+A = V_1 \Delta U_1' U_1 \Delta^{-1} V_1' V_1 \Delta U_1' = V_1 \Delta U_1' = A$.

4) $A^+AA^+ = U_1 \Delta^{-1} V_1' V_1 \Delta U_1' U_1 \Delta^{-1} V_1' = A^+$.

Theorem 1.13 *The Moore-Penrose generalized inverse is unique.*

Proof is an exercise.

Examples of Moore-Penrose Generalized Inverses.

1) If M is an orthogonal projection operator then $M^+ = M$.

2) If A is an $n \times n$ non-singular matrix then $A^+ = A^{-1}$

3) If $A = \text{diag}(a_{11}, \ldots, a_{nn})$ then A^+ is a diagonal matrix with elements $1/a_{ii}$ if $a_{ii} \neq 0$ and 0 if $a_{ii} = 0$.

4) Suppose A is an $n \times p$ matrix and $r(A) = n$. Then $A^+ = A'(AA')^{-1}$.

5) Suppose A is an $n \times p$ matrix of rank p, then $A^+ = (A'A)^{-1}A'$.

6) $r(A) = r(A^+)$.

7) For any matrix A, $(A^+)' = (A')^+$.

8) If A is symmetric, then A^+ is symmetric.

9) $(A^+)^+ = A$.

10) Suppose A, B are square and either A or B is singular. Then $(AB)^+ \neq B^+A^+$ in general.

11) $(AB)^+ = B^+A^+$ if A and B are both non-singular.

12) If A and B are not square, then $(AB)^+ \neq B^+A^+$ in general.

Example

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad AB = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$(AB)^+ = (1, 0) \text{ and } B^+ = (\frac{1}{2}, \frac{1}{2}).$$

$$A^+ = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \; B^+A^+ = \left(\frac{1}{2}, 0\right) \neq (AB)^+.$$

There are two important special cases for which the equality will hold.

1) For any A, $(A'A)^+ = A^+(A')^+$.

2) If A is $n \times p$ of rank p and B is $p \times r$ of rank p, then $(AB)^+ = B^+A^+$.

Theorem 1.14 *$M = XX^+$ is the unique orthogonal projection onto $C(X)$.*

Proof: Suppose X is $n \times p$ of rank r. We see that $M' = (XX^+)' = XX^+$ and $M^2 = (XX^+)(XX^+) = X(X^+XX^+) = XX^+$. Thus M is an orthogonal projection. Finally we need to show that $C(XX^+) = C(X)$.

1) "$\Rightarrow$" We first show that $C(XX^+) \subset C(X)$. Let $v \in C(XX^+)$, then $v = XX^+t$ for some $t \in R^n$. Now $v = XX^+t = Xt^*$. Thus $v = Xt^*$, and therefore $v \in C(X)$. Thus $C(XX^+) \subset C(X)$.

2) "$\Leftarrow$" Instead of showing that $C(X) \subset C(XX^+)$, we can show that $r(XX^+) = r(X)$. This along with 1) will complete the proof. Now $r(XX^+) = tr(XX^+)$ since XX^+ is an orthogonal projection operator. Thus $r(XX^+) = tr(XX^+) = tr(V_1\Delta U_1'U_1\Delta^{-1}V_1') = tr(V_1V_1') = tr(V_1'V_1) = tr(I_{r\times r}) = r$.

Theorem 1.15 *Let X^- be any generalized inverse of an $n \times p$ matrix X. Then*

$$X^* = X^-XX^- + (I - X^-X)A + B(I - XX^-)$$

is also a generalized inverse of X for any $p \times n$ matrices A and B. If $\dot{X}$ is another generalized inverse of X, then there is a choice of A and B for which $X^* = \dot{X}$.

Proof: To show the first part, we need to show $XX^*X = X$ for any A and B. To show the second part, take $X^* = \dot{X}X\dot{X} + (I - \dot{X}X)A + B(I - X\dot{X})$ and set $A = \dot{X}$ and $B = \dot{X}X\dot{X}$.

Theorem 1.16 *If G_1 and G_2 are generalized inverses of A, then so is G_1AG_2.*

Proof:

$$\begin{aligned} A(G_1AG_2)A &= (AG_1A)G_2A \\ &= AG_2A \\ &= A. \end{aligned}$$

Theorem 1.17 *If A is symmetric, then there exists a g-inverse of A that is symmetric, i.e , $(A^-)' = A^-$.*

Note: The MP g-inverse is symmetric.

Definition 1.7 *A generalized inverse A^- for a matrix A that has the property*

$$A^-AA^- = A^-$$

is said to be reflexive.

Note: MP g-inverse is reflexive.

Theorem 1.18 *Suppose X is an $n \times p$ matrix of rank r. Consider the matrix $X'X$. (note that $X'X$ is also of rank r). If G and H are generalized inverses of $X'X$, then*

i) $XGX'X = XHX'X = X$.

ii) $XGX' = XHX'$.

Proof: i) Let $v \in R^n$, and write $v = v_1 + v_2$, where $v_1 \in C(X)$, $v_2 \in C(X)^\perp$. Since $v_1 \in C(X)$, $v_1 = Xb$ for some $b \in R^p$. Then

$$\begin{aligned} v'XGX'X &= (v_1' + v_2')XGX'X \\ &= v_1'XGX'X \quad \text{since} \quad v_2'X = 0 \\ &= b'X'(XGX'X) \;=\; b'(X'X)G(X'X) \\ &= b'(X'X) \;=\; v_1'X \;=\; v'X\ . \end{aligned}$$

Thus, $v'XGX'X = v'X$. Since v and G are arbitrary, this implies $XGX'X = X$ for any G.

ii) Let $v \in R^n$, $v = v_1 + v_2$, with $v_1 \in C(X)$ and $v_2 \in C(X)^\perp$. Then $v_1 = Xb$ for some $b \in R^p$. Thus

$$\begin{aligned} XGX'v &= XGX'(v_1 + v_2) \;=\; XG(X'v_1 + X'v_2) \\ &= XGX'v_1 \;=\; XGX'Xb \\ &= XHX'Xb \quad \text{(by part i) of this theorem)} \\ &= XHX'v. \end{aligned}$$

Since v is arbitrary, we have $XGX' = XHX'$ for any choice of G and H.

Remark 1.2 *1) Part ii) of this theorem ($XGX' = XHX'$) says that $X(X'X)^{-}X'$ is invariant with respect to the choice of generalized inverse.*

2) It also follows that $X(X'X)^{-}X'$ is symmetric for any choice of generalized inverse, because by a previous theorem, we know that there exists a generalized inverse of $X'X$ that is symmetric. This, together with part ii) of the theorem, implies that $X(X'X)^{-}X'$ is symmetric for any choice of generalized inverse.

This now leads to the following result.

Theorem 1.19 *Suppose X is an $n \times p$ matrix of rank r. Let $M = X(X'X)^{-}X'$, where $-$ denotes any generalized inverse. Then M is the unique orthogonal projection operator onto $C(X)$.*

Proof: We prove this theorem by using the definition of an orthogonal projection.

1) Let $v \in C(X)$. Then $v = Xb$ for some $b \in R^p$. Thus

$$\begin{aligned} Mv &= X(X'X)^{-}X'v \\ &= X(X'X)^{-}X'Xb \\ &= Xb \quad \text{by part i) of the previous theorem} \\ &= v\,. \end{aligned}$$

Thus for any $v \in C(X)$, $Mv = v$, which means that M is a projection operator onto $C(X)$ by its definition.

2) To show orthogonality, we consider $w \in C(X)^{\perp}$. Then $w'x_j = 0$, $j = 1, \ldots, p$, where x_j is the jth column of X. By taking transposes, we have $x_j'w = 0$, $j = 1, \ldots, p$. Thus

$$\begin{aligned} Mw &= X(X'X)^{-}X'w \\ &= X(X'X)^{-}0 \\ &= 0 \end{aligned}$$

Thus M is the unique orthogonal projection operator onto $C(X)$.

Remark 1.3 Different forms for M
Suppose X is an $n \times p$ matrix of rank r. Let M denote the orthogonal projection operator onto $C(X)$. Then

i) *$M = QQ'$ where Q is the matrix from the QR decomposition of $X = QR$.*

ii) *$M = V_1V_1'$ where V_1 comes from the SVD of $X = V_1\Delta U_1'$.*

iii) *$M = XX^+$ where X^+ is the MP g-inverse of X.*

iv) *$M = X(X'X)^-X'$ where $X'X^-$ is any g-inverse of $X'X$.*

Theorem 1.20 *Suppose M and M_0 are orthogonal projection operators with $C(M_0) \subset C(M)$. Then*

i) $MM_0 = M_0M = M_0$.

ii) $M - M_0$ is an orthogonal projection operator.

iii) $C(M_0) \perp C(M - M_0)$.

iv) $C(M - M_0) = C(M) \cap C(M_0)^\perp$.

v) $M - M_0$ is the orthogonal projection operator onto $C(M - M_0)$.

Proof: i) Since $C(M_0) \subset C(M)$, we have $MM_0 = M_0$. Taking transpose on both sides $M_0'M' = M_0'$. By the symmetry property of the orthogonal projectction operator, $M_0' = M_0$ and $M' = M$. Therefore $M_0M = M_0$.

Remark 1.4 *For any two orthogonal projection operators M_1 and M_2, $C(M_1) = C(M_2)$ if and only if $M_1 = M_2$. This follows from the fact that orthogonal projection operators onto the same space are unique.*

ii)

$$\begin{aligned}(M - M_0)^2 &= (M - M_0)(M - M_0)\\ &= M^2 - M_0M - MM_0 + M_0^2\\ &= M - M_0 - M_0 + M_0\\ &= M - M_0\end{aligned}$$

Thus $M - M_0$ is a projection operator.

$$(M - M_0)' = M' - M_0' = M - M_0$$

and thus $M - M_0$ is an orthogonal projection operator.

iii) $(M - M_0)M_0 = MM_0 - M_0^2 = M_0 - M_0 = 0.$

iv) Let $x \in C(M - M_0)$. Then $(M - M_0)x = x$, and thus $Mx - M_0x = x$. $M_0x = 0$ since $C(M - M_0) \perp C(M_0)$. This implies that $Mx = x$, so that $x \in C(M)$. Hence $x \in C(M) \cap C(M_0)^\perp$.

Now let $x \in C(M) \cap C(M_0)^\perp$. Then $x \in C(M)$ and $x \in C(M_0)^\perp$. Now $(M - M_0)x = Mx - M_0x = x - 0 = x$. Thus, $x \in C(M - M_0)$. Therefore $C(M - M_0) = C(M) \cap C(M_0)^\perp$.

v) $(M - M_0)x = x$ for any $x \in C(M - M_0)$ by part iv). Thus $M - M_0$ is the orthogonal projection operator onto $C(M - M_0)$.

Theorem 1.21 *Suppose M and M_0 are orthogonal projection operators with $C(M_0) \subset C(M)$. Then $C(M - M_0)$ is the orthogonal complement of $C(M_0)$ with respect to $C(M)$.*

Proof: We see that $C(M - M_0) \perp C(M_0)$ from iii) above. This implies that $C(M - M_0)$ is contained in the orthogonal complement of $C(M_0)$ with respect to $C(M)$. If $x \in C(M)$ and $x \in C(M_0)^\perp$, then $Mx = x = (M - M_0)x + M_0x = (M - M_0)x$, and thus $x \in C(M - M_0)$. Therefore, the orthogonal complement of $C(M_0)$ with respect to $C(M)$ is contained in $C(M - M_0)$.

This theorem now implies that

$$C(M) = C(M_0) + C(M - M_0)$$

and thus $r(M) = r(M_0) + r(M - M_0)$.

Theorem 1.22 *Suppose M_1 and M_2 are two orthogonal projection operators in R^n. Then $M_1 + M_2$ is the orthogonal projection operator onto $C(M_1, M_2)$ if and only if $C(M_1) \perp C(M_2)$.*

Remark 1.5 *$C(M_1) \perp C(M_2)$ if and only if $M_1M_2 = M_2M_1 = 0$.*

Theorem 1.23 *If M_1 and M_2 are symmetric matrices, with $C(M_1) \perp C(M_2)$ and $M_1 + M_2$ is an orthogonal projection operator, then M_1 and M_2 are orthogonal projection operators.*

1.4 Solutions to Systems of Linear Equations

Consider the matrix equation

$$Y = X\beta \tag{1.1}$$

where $Y_{n\times 1}$, $X_{n\times p}$, and $\beta_{p\times 1}$.

We ask the following question: For a given X and Y, does there exist a solution β to the equation in (1.1)?

Characterization of Solution

1) If $p = n$ and X is nonsingular, the answer is **YES** and the unique solution is

$$\beta = X^{-1}Y \ .$$

In general, the solution depends on Y.

2) Suppose $p \leq n$. If $Y \in C(X)$, the answer to the question above is **YES** again, since Y can be expressed as a linear combination of the columns of X. In fact,

$$\beta = X^{-}Y \tag{1.2}$$

is a solution, since by the definition of a generalized inverse

$$X\beta = XX^{-}Y = Y \quad \text{for all} \ \ Y \in C(X) \ .$$

Is the solution in (1.2) unique?

This depends on $r(X)$. If X has full rank, (i.e., $r(X) = p$), then the columns of X form a basis for $C(X)$ and the coordinates of Y relative to that basis are unique and therefore the solution is unique.

However, the solution is not unique if $r(X) < p$. Thus, if β^* is a solution to $Y = X\beta$, then so is $\beta^* + w$, where $w \in \mathcal{N}(X)$. Thus the set of all solutions is of the form

$$X^{-}Y + (I - X'(XX')^{-}X)z\ ,\ \ z \in R^p\ .$$

Note: $X'(XX')^{-}X$ is the orthogonal projection operator onto $C(X')$ and $I - X'(XX')^{-}X$ is the orthogonal projection operator onto $C(X')^{\perp} = \mathcal{N}(X)$.

3) If $Y \notin C(X)$, and $p < n$, then no solution exists. This is the usual situation in linear models. In this case, we look for a vector in $C(X)$ that is "closest" to Y and solve the system above with that vector instead of Y.

Let $M = X(X'X)^{-}X'$. We know that

$$Y = MY + (I - M)Y$$

MY is the orthogonal projection of Y onto $C(X)$ and thus MY is the closest vector to Y in $C(X)$. We now solve the system

$$MY = X\beta\ .$$

We know the general solution to this system is

$$X^{-}MY + (I - X'(XX')^{-}X)z\ . \tag{1.3}$$

We can simplify this general solution in (1.3) a bit more.

Consider the SVD of X and write $X = V_1 \Delta U_1'$. The MP g-inverse of X is $X^{+} = U_1 \Delta^{-1} V_1'$. Choose the MP g-inverse in (1.3). Thus

$$\begin{aligned}
X^+MY &= X^+X(X'X)^+X'Y \\
&= (U_1\Delta^{-1}V_1')(V_1\Delta U_1')(U_1\Delta^{-2}U_1')(U_1\Delta V_1')Y \\
&= U_1\Delta^{-1}V_1'Y \\
&= X^+Y
\end{aligned}$$

Thus, the general solution can be written as

$$X^+Y + (I - X'(XX')^+X)z\ ,\ \ z \in R^p.$$

If $r(X) = p$, then $(I - X'(XX')^+X)z = 0$ for any $z \in R^p$, and $X^+ = (X'X)^{-1}X'$ so that

$$X^+Y = (X'X)^{-1}X'Y\ .$$

Chapter 2

RANDOM VECTOR, MATRICES AND THEIR DISTRIBUTIONS

2.1 RANDOM VECTORS AND MATRICES

Suppose

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{pmatrix}$$

is a random vector with $E(Y_i) = \mu_i$, Var$(Y_i) = \sigma_{ii}$, and Cov$(Y_i, Y_j) = \sigma_{ij}$.
The expectation of Y is defined as

$$E(Y) = \begin{pmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \mu \ .$$

Definition 2.1 *Suppose Z is an $n \times p$ matrix of random variables. Then*

$$E(Z) = \begin{pmatrix} E(Z_{11}) & \dots & E(Z_{1p}) \\ \vdots & \dots & \vdots \\ E(Z_{np}) & \dots & E(Z_{np}) \end{pmatrix}$$

The expectation of a random matrix is the matrix of the expectations.
Suppose Y is an $p \times 1$ vector of random variables. The covariance matrix of Y is

defined as

$$\begin{aligned} Cov(Y) &= E\left[(Y-\mu)(Y-\mu)'\right] \\ &= \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \dots & \dots & \sigma_{pp} \end{pmatrix} \end{aligned}$$

where $\sigma_{ij} = E\left[(Y_i-\mu_i)(Y_j-\mu_j)\right]$*,* $i, j = 1, \dots, p$*.*

Theorem 2.1 *Suppose* Y *is a random* $n \times 1$ *vector with mean* $\mu = E(Y)$ *and covariance matrix* $\Sigma = Cov(Y)$*. Moreover, suppose* A *is an* $r \times n$ *matrix of constants, and* b *is an* $r \times 1$ *vector of constants. Then*

$$E(AY + b) = AE(Y) + b = A\mu + b$$

and

$$Cov(AY + b) = ACov(Y)A' = A\Sigma A' \ .$$

We leave the proof as an exercise (use definition to prove).

Definition 2.2 *Let* Y *be an* $s \times 1$ *random vector and* W *is an* $r \times 1$ *random vector, with* $E(Y) = \mu$ *and* $E(W) = \gamma$*. We define covariance matrix of* W *and* Y*,* $Cov(W, Y)$*, as*

$$Cov(W, Y) = E\left[(W-\gamma)(Y-\mu)'\right] \ .$$

Note: Cov(W, Y) is an $r \times s$ matrix of covariances with $ijth$ element Cov(W_i, Y_j).

Theorem 2.2 *Let* Y *be an* $s \times 1$ *random vector and* W *is an* $r \times 1$ *random vector with* $Cov(W) = \Sigma_w$*,* $Cov(Y) = \Sigma_y$*,* $Cov(W, Y) = \Sigma_{wy}$*, and* $Cov(Y, W) = \Sigma_{yw}$*. Moreover, let* A *be an* $p \times r$ *matrix of constants, and* B *is an* $p \times s$ *matrix of constants. Then*

$$Cov(AW + BY) = A\Sigma_w A' + B\Sigma_y B' + A\Sigma_{wy} B' + B\Sigma_{yw} A' \ .$$

Theorem 2.3 *Covariance matrix for any random vector is always positive semidefinite.*

Proof: For $Y_{n\times 1}$, Cov$(Y) = E\left[(Y-\mu)(Y-\mu)'\right]$, where $\mu = E(Y)$. To show that Cov(Y) is positive semidefinite, we need to show that for any vector $x \in R^p$, x'Cov$(Y)x \geq 0$. Let $Z = Y - \mu$ for convenience. Then

$$\begin{aligned} x'\text{Cov}(Y)x &= x'E(ZZ')x \;=\; E(x'ZZ'x) \\ &= E(w'w) \;\text{ where }\; w = Z'x \\ &= E(\sum_{i=1}^{p} w_i^2) \;=\; \sum_{i=1}^{p} E(w_i^2) \;\geq 0\,, \end{aligned}$$

since the expectation of a positive random variable is always nonnegative.

Definition 2.3 *The correlation matrix of Y is defined as*

$$\mathit{Corr}(Y) = (\rho_{ij}) = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \sigma_{21} & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \dots & \dots & \sigma_{pp} \end{pmatrix}$$

where $\rho_{ij} = \sigma_{ij}/\sigma_i\sigma_j$, $i, j = 1, \dots, p$.

Suppose a $(p+q)\times 1$ random vector V is partitioned into two subsets of random vectors, $p \times 1$ vector Y and $q \times 1$ vector X:

$$Y = \begin{pmatrix} Y \\ X \end{pmatrix} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_p \\ X_1 \\ \vdots \\ X_q \end{pmatrix}.$$

Then,

$$\mu = \text{E}(v) = \text{E}\begin{pmatrix} Y \\ X \end{pmatrix} = \begin{pmatrix} \text{E}(Y) \\ \text{E}(X) \end{pmatrix} = \text{E}\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix},$$

$$\Sigma = \text{Cov}(V) = \text{Cov}\begin{pmatrix} Y \\ X \end{pmatrix} = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix},$$

where $\Sigma_{xy} = \Sigma'_{yx}$.

Exercise: What are Σ_{yy}, Σ_{xy}, Σ_{yx} and Σ_{xx}?

Chapter 3

DISTRIBUTION THEORY

Definition 3.1 *The moment generating function (MGF) of a random variable X, denoted $\psi_X(t)$ is defined as*

$$\psi_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx}\, f(x)\, dx\ .$$

The integral is replaced by a sum if X is discrete.

Definition 3.2 *Normal distribution*
A random variable X is said to have a normal distribution with mean μ and variance σ^2, written $X \sim N(\mu, \sigma^2)$ if X has density

$$f(x) = (2\pi)^{-1/2}\sigma^{-1} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}.$$

If $\mu = 0$ and $\sigma = 1$, then $X \sim N(0,1)$ and we say that X has a standard normal distribution.

Theorem 3.1 *If $X \sim N(\mu, \sigma^2)$, then $\psi_X(t) = \exp\left\{t\mu + \frac{1}{2}t^2\sigma^2\right\}$.*

Proof: To prove this result we need to know how to complete the square. Recall that

$$\begin{aligned} ax^2 + bx &= a\left(x^2 + \frac{bx}{a}\right) \\ &= a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a}\ . \end{aligned}$$

Now

$$
\begin{aligned}
\psi_X(t) &= \int_{-\infty}^{\infty} \exp\{tx\}\,(2\pi)^{-1/2}\sigma^{-1}\exp\left\{\frac{-1}{2\sigma^2}(x-\mu)^2\right\}\,dx \\
&= \int_{-\infty}^{\infty} (2\pi)^{-1/2}\sigma^{-1}\exp\{tx\}\exp\left\{\frac{-1}{2\sigma^2}(x^2-2\mu x+\mu^2)\right\}\,dx \\
&= \exp\left\{\frac{-\mu^2}{2\sigma^2}\right\}\int_{-\infty}^{\infty}(2\pi)^{-1/2}\sigma^{-1}\exp\left\{\frac{-1}{2\sigma^2}(x^2-2x(\mu+\sigma^2 t))\right\}\,dx \\
&= \exp\left\{\frac{-\mu^2}{2\sigma^2}\right\}\exp\left\{\frac{1}{2\sigma^2}(\mu+\sigma^2 t)^2\right\} \\
&\times \int_{-\infty}^{\infty}(2\pi)^{-1/2}\sigma^{-1}\exp\left\{\frac{-1}{2\sigma^2}(x-(\mu+\sigma^2 t))^2\right\}\,dx \\
&= \exp\left\{\frac{-\mu^2}{2\sigma^2}\right\}\exp\left\{\frac{1}{2\sigma^2}(\mu+\sigma^2 t)^2\right\} \\
&= \exp\left\{\mu t+\frac{1}{2}\sigma^2 t^2\right\}.
\end{aligned}
$$

Theorem 3.2 *Suppose $Z_1, \ldots, Z_n$ are independently identically distributed $(i.i.d.)$ $N(0,1)$ random variables. Define*

$$X = \sum_{i=1}^{n} Z_i^2 \,.$$

Then $X \sim \chi^2(n)$.

Definition 3.3 Chi-square distribution
A random variable X is said to have a central chi-square distribution with n degrees of freedom, written $X \sim \chi^2(n)$, if X has density

$$f(x) = \frac{1}{\Gamma(n/2)}\left(\frac{1}{2}\right)^{n/2} x^{n/2-1}e^{-x/2}\,,$$

where $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy$.

Theorem 3.3 *If $X \sim \chi^2(n)$, then $\psi_X(t) = (1-2t)^{-n/2}$.*

The proof is left as an exercise.

Definition 3.4 Noncentral chi-square distribution *Let $Z_1, \cdots, Z_n$ be independent with $Z_i \sim N(\mu_i, 1)$. Then $W = \sum_{i=1}^n Z_i^2$ has a* noncentral chi-square distribution *with n degrees of freedom and noncentrality parameter $\gamma = \sum_{i=1}^n \mu_i^2/2$. We write $W \sim \chi^2(n, \gamma)$.*

Remark 3.1 *Noncentral chi-square distributions arise in hypothesis testing situations in linear models when one is interested in finding the distribution of the test statistic under the alternative hypothesis.*

Theorem 3.4 *Suppose $Y_1, \ldots, Y_n$ are independent and $Y_i \sim N(\mu_i, \sigma^2)$, $i = 1, \ldots, n$. Define*

$$X = \frac{1}{\sigma^2} \sum_{i=1}^n Y_i^2 \ .$$

Then $X \sim \chi^2(n, \gamma)$, where $\gamma = (2\sigma^2)^{-1} \sum_{i=1}^n \mu_i^2$.

Remark 3.2 *Properties of noncentral chi-square*

1) *If $X \sim \chi^2(n, \gamma)$, then*

$$\psi_X(t) = (1-2t)^{-n/2} \exp\left\{\frac{2\gamma t}{1-2t}\right\} \ .$$

This can be proved using the definition of the noncentral chi-square density above and interchanging the order of integration and summation.

2) *If $X \sim \chi^2(n, \gamma)$, then $E(X) = n + 2\gamma$ and Var$(X) = 2n + 8\gamma$. This can be proved using the MGF in 1).*

3) *If $X \sim \chi^2(n, \gamma)$ and $\gamma = 0$, then this corresponds to a central chi-square random variable with n degrees of freedom. That is, $X \sim \chi^2(n, 0) = \chi^2(n)$.*

Definition 3.5 *t distribution Suppose $X \sim N(0, 1)$, $Y \sim \chi^2(n)$, and X and Y are independent. Define the random variable*

$$T = \frac{X}{\sqrt{Y/n}} \ .$$

Then T is said to have a t distribution with n degrees of freedom. We write $T \sim t(n)$.

Definition 3.6 *Noncentral t distribution Suppose $X \sim N(\mu, 1)$ and $Y \sim \chi^2(n)$, and X and Y are independent. Define the random variable*

$$W = \frac{X}{\sqrt{Y/n}} \, .$$

Then W is said to have a noncentral t distribution with n degrees of freedom and noncentrality parameter μ. We write $W \sim t(n, \mu)$. If $\mu = 0$, then W reduces to a central t distribution with n degrees of freedom.

Definition 3.7 *F distribution Suppose $X_1 \sim \chi^2(n_1, \gamma_1)$ and $X_2 \sim \chi^2(n_2, \gamma_2)$, and X_1 and X_2 are independent. Define the random variable*

$$F = \frac{X_1/n_1}{X_2/n_2} \, .$$

Then F is said to have a doubly noncentral F distribution with (n_1, n_2) degrees of freedom and noncentrality parameters (γ_1, γ_2). We write $F \sim F(n_1, n_2, \gamma_1, \gamma_2)$.

a) *If $\gamma_2 = 0$, then F is said to have a noncentral F distribution. We represent this as $F \sim F(n_1, n_2, \gamma_1)$.*

b) *If $\gamma_1 = 0$ and $\gamma_2 = 0$, then F is said to have a central F distribution. We represent this as $F \sim F(n_1, n_2)$.*

Remark 3.3 *The central F distribution arises in hypothesis tests of nested linear models. In this setting, the distribution of the test statistic under the null hypothesis often has a central F distribution. Noncentral F distributions arise from distributions of the test statistic under the alternative hypothesis. The distribution of the test statistic under the alternative hypothesis is important for power calculations.*

Remark 3.4 *Properties of F distribution If $F \sim F(n_1, n_2, \gamma)$, then*

a)

$$E(F) = \frac{n_2(n_1 + 2\gamma)}{n_1(n_2 - 2)} \, , \quad n_2 > 2$$

b)

$$\mathit{Var}(F) = 2\left(\frac{n_2}{n_1}\right)^2 \frac{(n_1+2\gamma)^2 + (n_1+4\gamma)(n_2-2)}{(n_2-2)^2(n_2-4)}, \quad n_2 > 4\ .$$

The formulas for the mean and variance for a central F are obtained by setting $\gamma = 0$ in a) and b) above.

3.1 Multivariate Moment Generating Functions

Suppose $X = (X_1, \ldots, X_n)'$ is an $n \times 1$ random vector with n dimensional density $f(x_1, \ldots, x_n)$. The multivariate moment generating function of X is defined as

$$\begin{aligned} & \psi_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = E(e^{t_1X_1+\ldots+t_nX_n}) \\ = & \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} e^{t_1X_1+\ldots+t_nX_n} f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n \\ = & E(e^{t'X}), \text{ where } t = (t_1, \ldots, t_n)' \text{ and } X = (X_1, \ldots, X_n)'. \end{aligned}$$

Remark 3.5 *Properties of the multivariate MGF The properties of the multivariate MGF are similar to the univariate MGF. Again let $X = (X_1, \ldots, X_n)'$ and $t = (t_1, \ldots, t_n)'$.*

1) $\psi_X(0) = 1$.

2) *If $X_1, \ldots, X_n$ are independent, then*

$$\psi_X(t) = \prod_{i=1}^{n} \psi_{X_i}(t_i)$$

where $\psi_{X_i}(t_i)$ is the univariate MGF of X_i.

3) *Moments can be obtained by differentiating the multivariate MGF.*

$$\frac{\partial^{k_1+\ldots+k_n}}{\partial t_1^{k_1} \ldots \partial t_n^{k_n}} \psi_X(t_1, \ldots, t_n)\Big|_{t_1=\ldots=t_n=0} = E(X_1^{k_1} \ldots X_n^{k_n})\ .$$

For example suppose $n = 2$ so that $X = (X_1, X_2)'$, and $\psi_X(t_1, t_2) = E(e^{t_1X_1+t_2X_2})$. We have

$$\frac{\partial^5}{\partial t_1^2 \partial t_2^3}\, \psi_X(t_1, t_2) \left.\right|_{t_1=t_2=0} = E(X_1^2 X_2^3) \ .$$

4) *The MGF for any marginal distribution of X is obtained by setting equal to 0 those t_j's that correspond to the X_j's not in the marginal distribution. For example, suppose $n = 4$, so that $X = (X_1, X_2, X_3, X_4)'$ and $\psi_X(t_1, t_2, t_3, t_4)$ is the multivariate MGF of X. Then $\psi_{X_1,X_3}(t_1, t_3) = \psi_X(t_1, 0, t_3, 0)$, $\psi_{X_1}(t_1) = \psi_X(t_1, 0, 0, 0)$ and so on.*

Definition 3.8 *Multivariate Characteristic function The multivariate Characteristic function is defined as*

$$\phi_X(t) = E(e^{it'X})$$

where $i = \sqrt{-1}$. The characteristic function always exist for any random variable or vector, but the MGF may NOT exist for some random variables. Thus the characteristic function is a bit more useful than the MGF in proving certain results.

For example, consider the Cauchy distribution with median 0. The density for this Cauchy distribution is

$$f(x) = \frac{1}{\pi(1 + x^2)}, \quad -\infty < x < \infty \ .$$

For the Cauchy distribution, the MGF does NOT exist, but the characteristic function is given by

$$\phi_X(t) = \exp\{- \mid t \mid\} \ .$$

The relationship between the characteristic function and the MGF is given by

$$\phi_X(t_1, \ldots, t_n) = \psi_X(it_1, \ldots, it_n) \ .$$

3.2 Multivariate Normal Distribution

A lot of estimation and hypothesis testing results in linear models are derived assuming a multivariate normal distribution for ε.

Definition 3.9 *Suppose $Z_1, \ldots, Z_n$ are $i.i.d.$ $N(0,1)$ random variables. Let $Z = (Z_1, \ldots, Z_n)'$. We have $E(Z) = 0$ and Cov$(Z) = I$. We say that Y has an r dimensional multivariate normal distribution if Y has the same distribution as $AZ + b$ for some $r \times n$ matrix of constants A and an $r \times 1$ vector of constants b. We denote the distribution of Y by*

$$Y \sim N_r(b, AA') .$$

Remark 3.6 *1. $E(Y) = E(AZ + b) = AE(Z) + b = A0 + b = b$, and Cov$(Y)$ = Cov$(AZ + b) = A$Cov$(Z)A' = AIA' = AA'$. Thus b is the mean vector of Y and AA' is the covariance matrix of Y.*

2. *Thus, when we write $Y \sim N_n(\mu, \Sigma)$, this means that Y has an n dimensional multivariate normal distribution with mean vector μ and covariance matrix Σ. We will abbreviate multivariate normal by MVN.*

3. *Recall that Σ must be positive semidefinite since any covariance matrix must be positive semidefinite. If Σ is singular, then the MVN distribution is said to be singular normal. In these cases, the density does not exist. The density of a MVN distribution exists only when Σ is positive definite.*

Definition 3.10 *Suppose $X = (X_1, \ldots, X_n)'$. Then X is said to have an n dimensional multivariate normal distribution with mean μ and covariance matrix Σ if X has density*

$$f(x) = (2\pi)^{-n/2} \mid \Sigma \mid^{-1/2} \exp\left\{ -\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu) \right\} .$$

Remark 3.7 *This definition requires Σ to be positive definite. Thus, this definition is a bit less general than the definition of MVN given above.*

Definition 3.11 *Let Y be an $n \times 1$ random vector and let A be an $n \times n$ matrix. A quadratic form is a random variable defined by $Y'AY$ for some Y and A.*

Now we give two useful results for quadratic forms.

Remark 3.8 *Square completion in n dimensions Suppose $x = (x_1, \ldots, x_n)'$ is an $n \times 1$ vector, A is an $n \times n$ nonsingular matrix, and b is an $n \times 1$ vector. Then*

$$x'Ax + b'x = \left(x + \frac{A^{-1}b}{2}\right)' A \left(x + \frac{A^{-1}b}{2}\right) - \frac{b'A^{-1}b}{4}$$

This is the multivariate analog of the one dimensional square completion given earlier. This result is very useful in computing multivariate normal integrals as we will see shortly.

Remark 3.9 *Combining quadratic forms Suppose x is an $n \times 1$ vector, μ_1 and μ_2 are $n \times 1$ vectors, A_1 and A_2 are $n \times n$ matrices such that $A_1 + A_2$ is nonsingular. Then*

$$\begin{aligned} & (x-\mu_1)'A_1(x-\mu_1) + (x-\mu_2)'A_2(x-\mu_2) \\ = \; & (x-\mu^*)'(A_1+A_2)(x-\mu^*) + \mu_1'A_1\mu_1 + \mu_2'A_2\mu_2 \\ - \; & \mu^*(A_1+A_2)\mu^* \, , \end{aligned}$$

where

$$\mu^* = (A_1 + A_2)^{-1}(A_1\mu_1 + A_2\mu_2) \, .$$

We can generalize this result to combining m quadratic forms. We have

$$\begin{aligned} & (x-\mu_1)'A_1(x-\mu_1) + \ldots + (x-\mu_m)'A_m(x-\mu_m) \\ = \; & (x-\mu^*)'B(x-\mu^*) + \sum_{i=1}^{m} \mu_i'A_i\mu_i - \mu'^*B\mu^* \end{aligned}$$

where $B = \sum_{i=1}^m A_i$ and $\mu^* = B^{-1}\left(\sum_{i=1}^m A_i\mu_i\right)$.

Remark 3.10 *Properties of MVN distributions*

1) *Suppose $X \sim N_n(\mu, \Sigma)$, then*

$$\psi_X(t) = \exp\{t'\mu + \frac{1}{2}t'\Sigma t\} \, .$$

Note: *the MGF does not require the inversion of Σ. Thus the MGF of MVN always exists and is given above even if Σ is singular. Thus the MGF of singular MVN distributions exists but the density does not.*

Proof: Assume Σ is positive definite. Then

$$
\begin{aligned}
&\psi_X(t) = E(e^{t'X}) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (2\pi)^{-n/2} \mid \Sigma \mid^{-1/2} \exp\{t'x\} \\
&\quad \times \exp\left\{\frac{-1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\} \, dx_1 \dots dx_n \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (2\pi)^{-n/2} \mid \Sigma \mid^{-1/2} \exp\{t'x\}
\end{aligned}
$$

$$
\begin{aligned}
&\quad \times \exp\left\{\frac{-1}{2}(x'\Sigma^{-1}x - 2x'\Sigma^{-1}\mu + \mu'\Sigma^{-1}\mu)\right\} \, dx_1 \dots dx_n \\
&= \exp\left\{\frac{-1}{2}\mu'\Sigma^{-1}\mu\right\} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (2\pi)^{-n/2} \mid \Sigma \mid^{-1/2} \\
&\quad \times \exp\left\{\frac{-1}{2}(x'\Sigma^{-1}x - 2x'(\Sigma^{-1}\mu + t))\right\} \, dx_1 \dots dx_n \\
&= \exp\left\{\frac{-1}{2}\mu'\Sigma^{-1}\mu\right\} \exp\left\{\frac{1}{2}(\mu + \Sigma t)'\Sigma^{-1}(\mu + \Sigma t)\right\} \\
&\quad \times \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (2\pi)^{-n/2} \mid \Sigma \mid^{-1/2} \\
&\quad \times \exp\left\{\frac{-1}{2}(x - (\mu + \Sigma t))'\Sigma^{-1}(x - (\mu + \Sigma t))\right\} \, dx_1 \dots dx_n \\
&= \exp\left\{\frac{-1}{2}\mu'\Sigma^{-1}\mu\right\} \exp\left\{\frac{1}{2}(\mu + \Sigma t)'\Sigma^{-1}(\mu + \Sigma t)\right\} \\
&= \exp\left\{t'\mu + \frac{1}{2}t'\Sigma t\right\}.
\end{aligned}
$$

This completes the proof. The characteristic function is given by

$$\phi_X(t) = \psi_X(it) = \exp\{it'\mu - \frac{1}{2}t'\Sigma t\} \, .$$

2) *A linear transformation of MVN's is MVN. Suppose $X \sim N_n(\mu, \Sigma)$, and define $Y = AX + b$, where A is an $r \times n$ matrix of constants and b is an $r \times 1$ vector of constants. Then*

$$Y \sim N_r(A\mu + b, A\Sigma A')$$

This result can easily be proved using the MGF. We have

$$\begin{aligned}\psi_Y(t) &= E(e^{t'Y}) \;=\; E(e^{t'(AX+b)}) \;=\; e^{t'b}E(e^{(A't)'X}) \\ &= \exp\{t'b\}\psi_X(A't) = \exp\{t'b\}\exp\{(A't)'\mu + \frac{1}{2}\left((A't)'\Sigma(A't)\right)\} \\ &= \exp\{t'(A\mu + b) + \frac{1}{2}t'A\Sigma A't\}.\end{aligned}$$

We can now recognize the MGF above as the MGF of a MVN distribution with mean $A\mu + b$ and covariance matrix $A\Sigma A'$.

3) *A linear combination of independent MVN's is MVN. Suppose $X_1, \ldots, X_k$ are independent and each $X_i \sim N_n(\mu_i, \Sigma_i)$, $i = 1, \ldots, k$. Suppose $a_1, \ldots, a_k$ are scalars and define*

$$Y = a_1X_1 + \ldots + a_kX_k \;.$$

Then

$$Y \sim N_n(\mu^*, \Sigma^*)$$

where $\mu^* = \sum_{i=1}^k a_i\mu_i$ and $\Sigma^* = \sum_{i=1}^k a_i^2\Sigma_i$. This again can be proved using MGF's.

4) *Marginal distributions of MVN are MVN. Suppose $X \sim N_n(\mu, \Sigma)$. Partition X into $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ where X_1 is $r \times 1$ and X_2 is $(n-r) \times 1$. Partition μ as $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ where μ_1 is $r \times 1$ and μ_2 is $(n-r) \times 1$. Similarly partition Σ as*

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

where Σ_{11} *is* $r \times r$*,* Σ_{12} *is* $r \times (n-r)$*,* $\Sigma_{21} = \Sigma_{12}'$ *is* $(n-r) \times r$*, and* Σ_{22} *is* $(n-r) \times (n-r)$*.*

The marginal distribution of X_1 *is given by*

$$X_1 \sim N_r(\mu_1, \Sigma_{11}) \; .$$

This can be proved using the MGF by putting in zeroes to the t_j*'s corresponding to* X_2*.*

5) *Conditional distributions of MVN are MVN. Suppose* $X \sim N_n(\mu, \Sigma)$*. Using the partition in 4), we have*

$$X_1 \mid X_2 = x_2 \sim N_r(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11.2})$$

where $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$*.*

Using MGF's to prove things about conditional distributions is usually hard since we do not have nice results for MGF's for conditional distributions. One usually has to rely on the density to prove results concerning conditional distributions.

Exercise

Suppose $X = (X_1, X_2, X_3)'$ is a 3×1 random vector and $X \sim N_3(\mu, \Sigma)$. Derive the conditional distribution of $(X_1, X_3) \mid X_2 = x_2$.

3.3 Connection of conditional distributions with linear regression

Consider the usual linear model

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \varepsilon \; .$$

Let $X = (X_1, \ldots, X_p)'$, and suppose

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim N_{p+1}(\mu, \Sigma) \ ,$$

where

$$\mu = \begin{pmatrix} E(Y) \\ E(X_1) \\ \vdots \\ E(X_p) \end{pmatrix} = \begin{pmatrix} \mu_y \\ \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}$$

and

$$\Sigma = \begin{pmatrix} \sigma_y^2 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where $\sigma_y^2 = \text{Var}(Y)$, $\Sigma_{22} = \text{Cov}(X)$, and Σ_{12} is a $1 \times p$ vector consisting of $\text{Cov}(Y, X)$.
Let $\mu_x = E(X) = (\mu_1, \ldots, \mu_p)'$. By property 5), we know that

$$Y \mid X = x \sim N_1(\mu_y + \Sigma_{12}\Sigma_{22}^{-1}(x - \mu_x), \sigma_y^2 - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \ .$$

Thus

$$\begin{aligned} E(Y \mid X = x) &= \mu_y + \Sigma_{12}\Sigma_{22}^{-1}(x - \mu_x) \\ &= (\mu_y - \Sigma_{12}\Sigma_{22}^{-1}\mu_x) + \Sigma_{12}\Sigma_{22}^{-1}x \end{aligned}$$

Now make the transformation

$$\beta_0 = \mu_y - \Sigma_{12}\Sigma_{22}^{-1}\mu_x \ ,$$
$$\beta' = \Sigma_{12}\Sigma_{22}^{-1} \ ,$$

and

$$\sigma^2 = \sigma_y^2 - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \ ,$$

where $\beta' = (\beta_1, \ldots, \beta_p)$. It can be shown that this transformation is one-to-one. The transformation implies

$$\begin{aligned} E(Y \mid X = x) &= \beta_0 + \beta' x \\ &= \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p \ . \end{aligned}$$

This result tells us that if we assume a multivariate normal distribution on the response Y and the (random) regressors X, then the regression function is the conditional expectation of $Y \mid X = x$.

Theorem 3.5 *If $X \sim N_n(\mu, \Sigma)$, then all marginals, conditionals, and linear combinations of the components of X are MVN.*

Remark 3.11 *The converse of the above theorem is NOT true. For example, if all of the marginals are MVN, this does NOT imply that the joint distribution is MVN.*

Example: Suppose (X_1, X_2) have joint density

$$\begin{aligned} f(x_1, x_2) &= \frac{1}{\pi\sqrt{2}} \exp\left\{-(x_1^2 + x_2^2)\right\} \\ &\times \left(\exp\left\{x_1^2/2\right\} + \exp\left\{x_2^2/2\right\} - \sqrt{2}\right) \end{aligned}$$

for $-\infty < x_1 < \infty$ and $-\infty < x_2 < \infty$.
A simple calculation shows that $X_1 \sim N(0,1)$ and $X_2 \sim N(0,1)$, but the joint distribution of (X_1, X_2) is not bivariate normal.

The multivariate normal distribution is completely characterized by its mean vector and covariance matrix. This means that once the mean vector and covariance matrix are specified, the density and MGF of the MVN are completely determined.

3.4 Independence of MVN

Definition 3.12 *General Definition of Independence*
Two random vectors are independent if their joint density $f(x, y)$ factors into

$$f(x, y) = f_1(x) f_2(y)$$

where $f_1(x)$ is the marginal density of X and $f_2(y)$ is the marginal density of Y.

Theorem 3.6 *If X and Y are independent random vectors then $G(X)$ and $H(Y)$ are independent where $G(.)$ and $H(.)$ are arbitrary functions. For example if X and Y are independent then X^2 and $Y^{10} + \exp\{Y\}$ are independent.*

Theorem 3.7 *Suppose $X \sim N_n(\mu, \Sigma)$. Define $Y_1 = AX$ and $Y_2 = BX$ where A is an $r \times n$ matrix of constants and B is an $s \times n$ matrix of constants. Then Y_1 and Y_2 are independent if and only if $A\Sigma B' = 0$. If $\Sigma = \sigma^2 I$, then Y_1 and Y_2 are independent if and only if $AB' = 0$.*

Theorem 3.8 *Suppose $X \sim N_n(\mu, \Sigma)$. Partition X into $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ where X_1 is $r \times 1$ and X_2 is $(n-r) \times 1$. Partition μ as $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ where μ_1 is $r \times 1$ and μ_2 is $(n-r) \times 1$. Similarly partition Σ as*

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

where Σ_{11} is $r \times r$, Σ_{12} is $r \times (n-r)$, $\Sigma_{21} = \Sigma_{12}'$ is $(n-r) \times r$, and Σ_{22} is $(n-r) \times (n-r)$. Then X_1 and X_2 are independent if and only if $\Sigma_{12} = 0$.

Theorem 3.9 *If $X \sim N_n(\mu_x, \Sigma_x)$ and $Y \sim N_m(\mu_y, \Sigma_y)$, and X and Y are independent, then*

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{n+m}(\mu, \Sigma)$$

where

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_x & 0 \\ 0' & \Sigma_y \end{pmatrix} .$$

Theorem 3.10 *If $X \sim N_n(\mu, \Sigma)$, then*

$$E(X) = \textit{mode}(X) = \textit{median}(X) = \mu .$$

Remark 3.12 *To show that the mode is μ, one needs to minimize $(x-\mu)'\Sigma^{-1}(x-\mu)$. To show that μ is the median, one can see that the density of X is symmetric about μ, that is, $f(\mu + x) = f(\mu - x)$.*

Theorem 3.11 *Suppose $X = (X_1, \dots X_n)'$ has density of the form*

$$f(x) = c \, \exp\{-Q/2\} \tag{3.1}$$

where $\exp\{-Q/2\} \propto \exp\left\{\frac{-1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}$ and c is the normalizing constant. Then $X \sim N_n(\mu, \Sigma)$.

Remark 3.13 *We note that if $f(x)$ has the form given in (1), we can find $\mu = E(X)$ by finding the mode of X. This reduces to minimizing Q. Thus, μ is the solution to the equations*

$$\frac{\partial Q}{\partial x_j} = 0 \, , \quad j = 1, \dots n.$$

These n equations will be <u>linear</u> in the x_j's.

Remark 3.14 *The elements of Σ^{-1} are contained in the quadratic term $x'\Sigma^{-1}x$. We note here that*

$$\begin{aligned} & (x-\mu)'\Sigma^{-1}(x-\mu) \\ = \; & x'\Sigma^{-1}x - 2x'\Sigma^{-1}\mu + \mu'\Sigma^{-1}\mu. \end{aligned}$$

<u>Example</u>

Suppose $X = (X_1, X_2)'$ has density of the form $f(x) = c \, \exp\{-Q/2\}$, where

$$Q = x_1^2 + 2x_1x_2 + 4x_2^2 + 2x_1 \; .$$

What is the distribution of X? We know that X must be multivariate normal by the theorem. To find $E(X)$, we have

$$\frac{\partial Q}{\partial x_1} = 2x_1 + 2x_2 + 2 = 0$$

and

$$\frac{\partial Q}{\partial x_2} = 2x_1 + 8x_2 = 0$$

Solving these equations for x_1 and x_2 leads to $x_1 = -4/3$ and $x_2 = 1/3$. Thus $\mu = (-4/3, 1/3)'$. To find the elements of Σ^{-1}, we look at the quadratic terms in Q. Let

$$\Sigma^{-1} = \begin{pmatrix} \sigma^{(11)} & \sigma^{(12)} \\ \sigma^{(12)} & \sigma^{(22)} \end{pmatrix}$$

Thus

$$\begin{aligned} x'\Sigma^{-1}x &= (x_1, x_2)\begin{pmatrix} \sigma^{(11)} & \sigma^{(12)} \\ \sigma^{(12)} & \sigma^{(22)} \end{pmatrix}(x_1, x_2)' \\ &= \sigma^{(11)}x_1^2 + 2\sigma^{(12)}x_1x_2 + \sigma^{(22)}x_2^2 \; . \end{aligned}$$

Thus for our problem, $\sigma^{(11)} = 1$, $2\sigma^{(12)} = 2$ and $\sigma^{(22)} = 4$. Therefore,

$$\Sigma^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix} \; .$$

Inverting this gives

$$\Sigma = \begin{pmatrix} 4/3 & -1/3 \\ -1/3 & 1/3 \end{pmatrix} \; .$$

Thus

$$X \sim N_2\left(\begin{pmatrix} -4/3 \\ 1/3 \end{pmatrix}, \begin{pmatrix} 4/3 & -1/3 \\ -1/3 & 1/3 \end{pmatrix}\right) \; .$$

Chapter 4

DISTRIBUTION OF QUADRATIC FORMS

Definition 4.1 *Suppose Y is an n dimensional random vector and let A be an $n \times n$ matrix of constants. A quadratic form is a random variable defined by $Y'AY$ for some Y and A. Since $Y'AY$ is real-valued, we have*

$$Y'AY = Y'A'Y = Y'\left(\frac{A+A'}{2}\right)Y\ .$$

Since $(A+A')/2$ is always symmetric for any A, we can without loss of generality restrict ourselves to quadratic forms where A is symmetric.

Remark 4.1 *Sums of squares Expressing a sum of squares encountered in regression or analysis of variance as a quadratic form $y'Ay$ where y is a random vector and A is a symmetric matrix of constants. We will eventually show that certain sums of squares have chi-square distributions and are independent.*

Remark 4.2 *Expressing some simple sums of squares as quadratic forms in y Let $y_1, y_2, \ldots, y_n$ be a random sample from a population with mean μ and variance σ^2. Then,*

$$\sum_{i=1}^{n} y_i^2 = \left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right) + n\bar{y} = \sum_{i=1}^{n}(y_i - \bar{y})^2 + n\bar{y}^2.$$

We now can express $\sum_{i=1}^{n} y_i^2$ as a quadratic form

$$\sum_{i=1}^{n} y_i^2 = y'y = y'Iy = y'\left(I - \frac{1}{n}J\right)y + y'\left(\frac{1}{n}J\right)y \tag{4.1}$$

where $y' = (y_1, y_2, \ldots, y_n)$, I is a $n \times n$ identity matrix, $j = (1, 1, \ldots, 1)'$ and $J = jj'$.

This holds since:
i) we can write $\bar{y}$ as

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n} j'y,$$

so

$$\begin{aligned} n\bar{y}^2 &= n\left(\frac{1}{n}j'y\right)^2 = n\left(\frac{1}{n}j'y\right) \\ &= n\left(\frac{1}{n}\right)^2 y'jj'y \\ &= n\left(\frac{1}{n}\right)^2 y'J'y \\ &= y'\left(\frac{1}{n}J\right)^2 y, \end{aligned}$$

ii) we can write $\sum_{i=1}^{n}(y_i - \bar{y})^2$ as

$$\begin{aligned} \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 &= y'Iy - y'\left(\frac{1}{n}J\right)y \\ &= y'\left(I - \frac{1}{n}J\right)y. \end{aligned}$$

The three quadratic forms in (4.1) have the following properties:

1. $I = \left(I - \frac{1}{n}J\right) = \frac{1}{n}J.$
2. $I, I - \frac{1}{n}J$, *and* $\frac{1}{n}J$ *are idempotent.*
3. $\left(I - \frac{1}{n}J\right)\left(\frac{1}{n}J\right) = O.$

Remark 4.3 *These properties will be used to show that $\sum(y_i - \bar{y})^2/\sigma^2$ and $n\bar{y}^2/\sigma^2$ have chi-square distributions and are independent.*

4.1 Mean, Variance and MGF of quadratic forms

Theorem 4.1 *If Y is a random vector with mean μ and covariance matrix Σ and if A is a symmetric matrix of constants, then*

$$E(Y'AY) = tr(A\Sigma) + \mu' A\mu.$$

Proof is left as a homework assignment.
Example: Mean of the sample variance s^2 where

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}.$$

The numerator of s^2 is

$$\sum_{i=1}^n (y_i - \bar{y})^2 = y' \left(I - \frac{1}{n}J\right) y,$$

where $y = (y_1, y_2, \ldots, y_n)'$.

Assume y_i's are i.i.d with mean μ and variance σ^2 then $\mathrm{E}(y) = \mu j$ and $\mathrm{Cov}(y) = \sigma^2 I$. Let $A = I - (1/n)J$, $\Sigma = \sigma^2 I$ and $\boldsymbol{\mu} = \mu j$. By the Theorem above,

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^n (y_i - \bar{y})^2\right] &= \mathrm{tr}\left[\left(I - \frac{1}{n}J\right)(\sigma^2 I)\right] + \mu j' \left(I - \frac{1}{n}J\right)\mu j \\ &= \cdots \\ &= \sigma^2(n-1).\end{aligned}$$

Therefore,

$$\mathrm{E}(s^2) = \frac{\mathrm{E}\left[\sum_{i=1}^n (y_i - \bar{y})^2\right]}{n-1} = \frac{(n-1)\sigma^2}{n-1} = \sigma^2.$$

Theorem 4.2 *If $y \sim N_p(\mu, \Sigma)$ then the MGF of $y'Ay$ is*

$$M_{y'Ay}(t) = |I - 2tA\Sigma|^{-1/2} \exp\left[-\mu'\{I - (I - 2tA\Sigma)\}^{-1}\Sigma^{-1}\mu/2\right].$$

Proof is left as a homework.

Theorem 4.3 *If $y \sim N_p(\mu, \Sigma)$ then*

$$\mathit{Var}(y'Ay) = 2\mathit{tr}[(A\Sigma)^2] + 4\mu' A\Sigma A\mu.$$

Theorem 4.4 *Suppose $Y \sim N_n(0, \sigma^2 I)$. Then*

$$\frac{1}{\sigma^2}(Y'MY) \sim \chi^2(r)$$

if and only if M is an orthogonal projection operator of rank r.

Theorem 4.5 *Suppose $Y \sim N_n(\mu, \sigma^2 I)$. Then*

$$\frac{1}{\sigma^2}(Y'MY) \sim \chi^2(r, \gamma)$$

if and only if M is an orthogonal projection operator of rank r and $\gamma = \mu' M\mu/(2\sigma^2)$.

Theorem 4.6 *Suppose $Y \sim N_n(\mu, \sigma^2 M)$ where M is an orthogonal projection operator of rank r and $\mu \in C(M)$. Then*

$$\frac{1}{\sigma^2}(Y'Y) \sim \chi^2(r, \gamma) \quad \textit{where } \gamma = \frac{\mu'\mu}{2\sigma^2}.$$

Theorem 4.7 *Suppose $Y \sim N_n(\mu, \Sigma)$ where Σ is positive definite. Then*

$$Y'AY \sim \chi^2(r, \gamma)$$

where $\gamma = (\mu' A\mu)/2$ if and only if <u>any</u> of the following conditions are satisfied:

i) $A\Sigma$ is a projection operator of rank r.

ii) ΣA is a projection operator of rank r.

iii) Σ is a generalized inverse of A and A has rank r.

Remark 4.4 Note: *The noncentrality parameter is always found by replacing Y with $E(Y)$ in the quadratic form. For example, for the theorem above, the noncentrality parameter is derived as*

$$\gamma = \frac{E(Y)'AE(Y)}{2} = \frac{\mu' A\mu}{2} \ .$$

Theorem 4.8 *Suppose $Y \sim N_n(\mu, \Sigma)$. Then $Y'AY \sim \chi^2(tr(A\Sigma), \gamma)$, $\gamma = \frac{\mu' A\mu}{2}$, if*

i) $\Sigma A \Sigma A \Sigma = \Sigma A \Sigma$ *and*

ii) $\mu' A \Sigma A \mu = \mu' A \mu$ *and*

iii) $\Sigma A \Sigma A \mu = \Sigma A \mu$.

Theorem 4.9 *Suppose $Y \sim N_n(\mu, \Sigma)$, where Σ is positive definite. Then $Y'AY$ has the same distribution as the random variable*

$$U = \sum_{i=1}^{n} d_{ii} U_i$$

where d_{ii} are the eigenvalues of $A\Sigma$ and $U_1., \ldots, U_n$ are independent non-central chi-square random variables with one degree of freedom.

4.2 Independence of Quadratic Forms

Theorem 4.10 *If $Y \sim N_n(\mu, \sigma^2 I)$, then*

i) $Y'AY$ and BY are independent if and only if $AB' = 0$ where A is a symmetric matrix.

ii) $Y'AY$ and $Y'BY$ are independent if and only if $AB = 0$, where A and B are symmetric.

Proof of i): Recall that AY and BY are independent if $\text{Cov}(AY, BY) = 0$. But $\text{Cov}(AY, BY) = A\text{Cov}(Y)B' = A(\sigma^2 I)B' = \sigma^2 AB'$. This quantity equals 0 if and only if $AB' = 0$. Since $Y'AY$ is a function of AY it follows that $Y'AY$ and BY are independent if and only if $AB' = 0$. Note that

$$\begin{aligned} AB' = 0 &\Leftrightarrow (AB')' = 0 \\ &\Leftrightarrow BA' = 0 \\ &\Leftrightarrow BA = 0 \quad \text{if } A \text{ is symmetric} \\ &\Leftrightarrow AB = 0 \quad \text{if } A \text{ and } B \text{ are symmetric.} \end{aligned}$$

A similar argument can be given for ii).

Theorem 4.11 *Suppose $Y \sim N_n(\mu, \Sigma)$, and suppose that A, B, and Σ are all positive semidefinite. Then $Y'AY$ and $Y'BY$ are independent if $\Sigma A \Sigma B \Sigma = 0$. If Σ is positive definite, then $Y'AY$ and $Y'BY$ are independent if $A\Sigma B = 0$.*

Proof: Since A, B, and Σ are positive semidefinite, we can write $A = RR'$, $B = SS'$ and $\Sigma = QQ'$. Then

$$Y'AY = Y'RR'Y = (R'Y)'(R'Y)$$

and

$$Y'BY = Y'SS'Y = (S'Y)'(S'Y)$$

Thus $Y'AY$ and $Y'BY$ are independent if $R'Y$ and $S'Y$ are independent. $R'Y$ and $S'Y$ are independent if and only if

$$\begin{aligned}\text{Cov}(R'Y, S'Y) &= 0 \\ &\Leftrightarrow R'\Sigma S = 0 \\ &\Leftrightarrow R'QQ'S = 0 \\ &\Leftrightarrow C(Q'S) \perp C(Q'R)\ .\end{aligned}$$

Since $C(AA') = C(A)$ for any matrix A, we have

$$\begin{aligned}C(Q'S) \perp C(Q'R) &\Leftrightarrow C(Q'SS'Q) \perp C(Q'RR'Q) \\ &\Leftrightarrow (Q'SS'Q)(Q'RR'Q) = 0 \\ &\Leftrightarrow Q'B\Sigma AQ = 0 \\ &\Leftrightarrow C(Q) \perp C(B\Sigma AQ) \\ &\Leftrightarrow C(QQ') \perp C(B\Sigma AQ) \\ &\Leftrightarrow QQ'B\Sigma AQ = 0 \\ &\Leftrightarrow \Sigma B\Sigma AQ = 0\end{aligned}$$

Since $C(Q) = C(QQ') = C(\Sigma)$, we have

$$\begin{aligned}\Sigma B\Sigma AQ = 0 &\Leftrightarrow \Sigma B\Sigma A\Sigma = 0 \\ &\Leftrightarrow \Sigma A\Sigma B\Sigma = 0 \quad \text{by taking transposes}\end{aligned}$$

This completes the proof for the first part. If Σ^{-1} exists, we have

$$\begin{aligned}\Sigma A\Sigma B\Sigma = 0 &\Leftrightarrow \Sigma^{-1}\Sigma A\Sigma B\Sigma\Sigma^{-1} = 0 \\ &\Leftrightarrow A\Sigma B = 0 \\ &\Leftrightarrow B\Sigma A = 0\end{aligned}$$

Theorem 4.12 *Suppose $Y \sim N_n(\mu, \Sigma)$ and suppose A and B are $n \times n$ symmetric matrices. If*

i) $\Sigma A\Sigma B\Sigma = 0$,

ii) $\Sigma A \Sigma B \mu = 0$,

iii) $\Sigma B \Sigma A \mu = 0$, *and*

iv) $\mu' A \Sigma B \mu = 0$.

Then $Y'AY$ and $Y'BY$ are independent.

Remark 4.5 *Here A and B are symmetric and not necessarily positive semidefinite.*

To prove this remark, we write $\Sigma = QQ'$ and $Y = \mu + QZ$, where $Z \sim N_n(0, I)$. Using this decomposition of Y, we multiply $Y'AY$ and $Y'BY$ out and check independence of the terms using Theorem 1.3.7 of Christensen and the argument contained in the proof of Theorem 1.3.8.

Remark 4.6 *If $Y \sim N_n(\mu, \Sigma)$, then AY and BY are independent if and only if $A\Sigma B' = 0$. If AY is independent of BY then $Y'AY$ and BY are independent, and $Y'AY$ and $Y'BY$ are independent.*

Remark 4.7 *If Σ is full rank (i.e., positive definite) and $Y'AY$ and BY are independent, then AY and BY are independent. Also, if $Y'AY$ and $Y'BY$ are independent, then AY and BY are independent.*

Remark 4.8 *However if Σ is less than full rank then, if $Y'AY$ and BY are independent, then this does NOT imply that AY and BY are independent. Also, if $Y'BY$ and AY are independent, then this does NOT imply that AY and BY are independent.*

Chapter 5

ESTIMATION

We now consider the problem of estimation in the linear model.

Consider the linear model

$$Y = X\beta + \varepsilon \tag{5.1}$$

where

$$E(\varepsilon) = 0 \ , \ \text{Cov}(\varepsilon) = \sigma^2 I \ , \tag{5.2}$$

and Y is an $n \times 1$ random vector, X is an $n \times p$ fixed matrix of rank $r \leq p$, $\beta = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of regression coefficients, and ε is an $n \times 1$ vector of random errors.

Our goal is to estimate β or more generally, a linear function of β, say $\lambda'\beta$, where λ is a $p \times 1$ vector of constants.

For example, if $\lambda = (1, -1, 0, \ldots, 0)'$, then $\lambda'\beta = \beta_1 - \beta_2$.

We need to develop the notion of an estimable function.

Definition 5.1 *Estimability $\lambda'\beta$ is estimable if there exists an $n \times 1$ vector of constants ρ, such that*

$$E(\rho' Y) = \lambda'\beta$$

for any β.

Definition 5.2 *An estimate $f(Y)$ of $\lambda'\beta$ is said to be unbiased for $\lambda'\beta$ if*

$$E(f(Y)) = \lambda'\beta \ .$$

Definition 5.3 *$f(Y)$ is a linear estimate of $\lambda'\beta$ if*

$$f(Y) = a_0 + a'Y$$

for some vectors of constants a_0 and a.

Thus, $\lambda'\beta$ is estimable if there exists a linear unbiased estimate of it. This leads to the following theorem.

Theorem 5.1 *$a_0 + a'Y$ is unbiased for $\lambda'\beta$ if and only if $a_0 = 0$ and $a'X = \lambda'$.*

Proof:

1) Necessity ($\leftarrow$) : If $a_0 = 0$ and $a'X = \lambda'$, then $E(a_0 + a'Y) = 0 + a'X\beta = \lambda'\beta$.

2) Sufficiency ($\rightarrow$): If $a_0 + a'Y$ is unbiased for $\lambda'\beta$, then $\lambda'\beta = E(a_0 + a'Y) = a_0 + a'X\beta$ for any β. Subtracting $a'X\beta$ from both sides gives

$$(\lambda' - a'X)\beta = a_0 \text{ for any } \beta \ .$$

This can only be true if $a_0 = 0$ and $\lambda' = a'X$.

Corollary 5.1.1 *$\lambda'\beta$ is estimable if and only if there exists an $n \times 1$ vector ρ such that*

$$\rho' X = \lambda' \ .$$

Remark 5.1 **(1)** *The statement above implies $\lambda = X'\rho$. Thus $\lambda'\beta$ is estimable if $\lambda \in C(X')$.*

(2) *We note that the concept of estimability is based entirely on the assumption that $E(Y) = X\beta$. Estimability does not depend on Cov(Y).*

Definition 5.4 *Suppose Λ is a $p \times s$ matrix of constants. Then the $s \times 1$ vector of linear functions $\Lambda'\beta$ is estimable if and only if there exists an $n \times s$ matrix of constants P such that*

$$P'X = \Lambda' .$$

Remark 5.2 **1)** *$\Lambda'\beta$ is estimable if each of its components is estimable.*

2) *P above is not unique. However MP is unique, where $M = X(X'X)^{-}X'$. To see that MP is unique, let P_1, P_2 be such that $P_1'X = \Lambda'$ and $P_2'X = \Lambda'$. Then*

$$\begin{aligned} MP_1 &= X(X'X)^{-}X'P_1 = X(X'X)^{-}\Lambda \\ &= X(X'X)^{-}X'P_2 = MP_2 . \end{aligned}$$

3) *The components of β need not be estimable, but linear combinations of the components of β may be estimable.*

4) *If X is of full rank, then β is estimable and every linear combination of the components is estimable.*

5) *If X is of full rank, then we can pick $P' = (X'X)^{-1}X'$ and thus $P'X\beta = \beta$. Note here that $\Lambda' = P'X = (X'X)^{-1}(X'X) = I_{p\times p}$.*

5.1 Least Squares Estimation

Notation: The squared length of a vector will be denoted by $||.||^2$. Thus

$$||Y||^2 = Y'Y .$$

If A is an $s \times n$ matrix, then

$$||AY||^2 = Y'A'AY .$$

If A is an orthogonal projection operator, then

$$\|AY\|^2 = Y'AY \ .$$

Definition 5.5 *The least squares estimate of β, denoted $\hat{\beta}$, satisfies*

$$(Y - X\hat{\beta})'(Y - X\hat{\beta}) = \min_{\beta}(Y - X\beta)'(Y - X\beta) \ .$$

Thus, the least squares estimate of β minimizes the squared Euclidean distance between Y and its mean $\mu = X\beta$. If $Y \notin C(X)$, then we know that a solution of $Y = X\beta$ does not exist in general. If $Y \in C(X)$, a solution exists. Thus the least squares solution will be the closest vector to Y in $C(X)$. We know that this vector is MY. We are led to the following theorem.

Theorem 5.2 *$\hat{\beta}$ is a least squares solution to β if and only if*

$$X\hat{\beta} = MY$$

where $M = X(X'X)^{-}X'$. We note here that $\hat{\beta}$ is not necessarily unique. Uniqueness will depend on estimability.

Proof: Let $\tilde{\beta}$ be an arbitrary estimate of β. We can write

$$\begin{aligned} & (Y - X\tilde{\beta})'(Y - X\tilde{\beta}) \\ = \ & (Y - MY + MY - X\tilde{\beta})'(Y - MY + MY - X\tilde{\beta}) \\ = \ & (Y - MY)'(Y - MY) + (Y - MY)'(MY - X\tilde{\beta}) \\ + \ & (MY - X\tilde{\beta})'(Y - MY) + (MY - X\tilde{\beta})'(MY - X\tilde{\beta}) \end{aligned}$$

We note here that $(Y - MY)'(MY - X\tilde{\beta}) = Y'(I - M)MY - Y'(I - M)X\tilde{\beta} = 0 - 0 = 0$.

Substitution now gives

$$\begin{aligned} & (Y - X\tilde{\beta})'(Y - X\tilde{\beta}) \\ = \ & (Y - MY)'(Y - MY) + (MY - X\tilde{\beta})'(MY - X\tilde{\beta}) \end{aligned}$$

Both terms on the right hand side are nonnegative and the first term does not depend on $\tilde{\beta}$. Thus $(Y - X\tilde{\beta})'(Y - X\tilde{\beta})$ will be minimized by minimizing $(MY - X\tilde{\beta})'(MY - X\tilde{\beta})$, which is the squared distance between MY and $X\tilde{\beta}$. This distance is 0 if and only if $MY = X\tilde{\beta}$.

Corollary 5.2.1 *$\tilde{\beta} = (X'X)^- X'Y$ is a least squares estimate of β (plug in, $X\tilde{\beta} = X(X'X)^- X'Y = MY$). The set of all least squares estimates of β are of the form*

$$\hat{\beta} = X^+ Y + (I - X'(XX')^- X)z \; , \;\; z \in R^p \; .$$

It turns out that least squares estimates of $\lambda'\beta$ are unique if and only if $\lambda'\beta$ is estimable. We are led to the following theorem.

Theorem 5.3 *$\lambda' = \rho' X$ if and only if $\lambda'\hat{\beta}_1 = \lambda'\hat{\beta}_2$ for any $\hat{\beta}_1$, $\hat{\beta}_2$ satisfying*

$$X\hat{\beta}_1 = MY, \;\; X\hat{\beta}_2 = MY \; .$$

Proof:

1) $\Rightarrow$: If $\lambda' = \rho' X$, then $\lambda'\hat{\beta}_1 = \rho' X\hat{\beta}_1 = \rho' MY = \rho' X\hat{\beta}_2 = \lambda'\hat{\beta}_2$.

2) $\Leftarrow$: Decompose λ into vectors in $C(X')$ and $C(X')^\perp$. Moreover, let $N = X'(XX')^- X$. N is the orthogonal projection operator onto $C(X')$, and $I - N$ is the orthogonal projection operator onto $C(X')^\perp$. Thus $\lambda = X'\rho_1 + (I - N)\rho_2$, where $\rho_1 \in R^n$ and $\rho_2 \in R^p$. We have $\lambda' = \rho_1' X + \rho_2'(I - N)$, $X'\rho_1 \in C(X')$, and $(I - N)\rho_2 \in C(X')^\perp$. Thus

$$\begin{aligned}
&\lambda'(\hat{\beta}_1 - \hat{\beta}_2) = 0 \\
&\Leftrightarrow (\rho_1' X + \rho_2'(I - N))(\hat{\beta}_1 - \hat{\beta}_2) = 0 \\
&\Leftrightarrow \rho_1'(X\hat{\beta}_1 - X\hat{\beta}_2) + \rho_2'(I - N)(\hat{\beta}_1 - \hat{\beta}_2) = 0 \\
&\Leftrightarrow \rho_1'(MY - MY) + \rho_2'(I - N)(\hat{\beta}_1 - \hat{\beta}_2) = 0 \\
&\Leftrightarrow \rho_2'(I - N)(\hat{\beta}_1 - \hat{\beta}_2) = 0 \; .
\end{aligned}$$

Thus

$$\rho_2'(I - N)(\hat{\beta}_1 - \hat{\beta}_2) = 0 \tag{5.3}$$

for any $\hat{\beta}_1 - \hat{\beta}_2$. Let $t = \hat{\beta}_1 - \hat{\beta}_2$, and decompose $t = t_1 + t_2$, where $t_1 \in C(X')$ and $t_2 \in C(X')^\perp$. Thus (5.3) implies

$$\begin{aligned}
&\rho_2'(I - N)(t_1 + t_2) = 0 \Rightarrow \\
&\rho_2'(I - N)t_1 + \rho_2'(I - N)t_2 = 0
\end{aligned}$$

Now $\rho_2'(I-N)t_1 = 0$ since $t_1 \in C(X')$ by definition. This implies $\rho_2'(I-N)t_2 = 0$ for any $t_2 \in C(X')^\perp$, and thus for any $t \in R^p$, $\rho_2'(I-N)t = 0$. This implies that $\rho_2'(I-N) = 0$, which implies $(I-N)\rho_2 = 0$. Thus $\lambda = X'\rho_1 + (I-N)\rho_2 = X'\rho_1 + 0 = X'\rho_1$.

Therefore $\lambda \in C(X')$. This completes the proof.

Corollary 5.3.1 *The unique least squares estimate of $\rho' X\beta$ is $\rho' MY$.*

Corollary 5.3.2 *The unique least squares estimate of $P' X\beta$ is $P' MY$.*

Corollary 5.3.3 *The unique least squares estimator of $\mu = X\beta$ is MY.*

If $\lambda'\beta$ is estimable, then its unique least squares estimate is unbiased. This leads to the following theorem.

Theorem 5.4 *If $\lambda' = \rho' X$, then $E(\rho' MY) = \lambda'\beta$.*

Proof: $E(\rho' MY) = \rho' ME(Y) = \rho' MX\beta = \rho' X\beta = \lambda'\beta$.

The squared length of MY is the regression sums of squares. That is

$$\|MY\|^2 = Y'MY \ .$$

Estimation of σ^2

We have the decomposition

$$Y = MY + (I-M)Y$$

Now MY is the least squares estimate of $X\beta$. We note that

$$MY = MX\beta + M\varepsilon = X\beta + M\varepsilon$$

so that $MY = X\beta + M\varepsilon$ where $E(M\varepsilon) = ME(\varepsilon) = M0 = 0$. Similarly, we have

$$(I-M)Y = (I-M)X\beta + (I-M)\varepsilon = (I-M)\varepsilon$$

so that $(I-M)Y$ depends only on ε. Since $(I-M)Y$ depends only on ε, it is reasonable to use some function of $(I-M)Y$ to estimate σ^2. The function we use is the squared length of $(I-M)Y$.

Theorem 5.5 *Suppose $r(X) = r$. Then*

$$\frac{\|(I-M)Y\|^2}{n-r} = \frac{Y'(I-M)Y}{n-r}$$

is an unbiased estimate of σ^2.

Proof: We have

$$\begin{aligned} E(Y'(I-M)Y) &= (X\beta)'(I-M)X\beta + tr(\sigma^2 I(I-M)) \\ &= \beta' X'(I-M)X\beta + \sigma^2 tr(I-M) \\ &= 0 + \sigma^2(n-r) \end{aligned}$$

Thus $E\left(Y'(I-M)Y/(n-r)\right) = \sigma^2$.

$(I-M)Y$ is the residual vector and its squared length is the error sum of squares. Thus $\|(I-M)Y\|^2 = Y'(I-M)Y$ is the error sum of squares (SSE). $\|(I-M)Y\|^2/(n-r)$ is the mean square error (MSE).

Remark 5.3 *Properties of Estimators It is desirable to have estimators which satisfy certain properties such as*

1) *Unbiasedness*

2) *Minimum variance*

3) *Efficiency*

4) *Asymptotic normality*

Suppose our goal is to estimate $\lambda'\beta$ and we want to find the "best" linear unbiased estimate of it. The word "best" is in the sense of minimum variance. Thus, we seek linear estimators in Y, say $a'Y$, such that

$$E(a'Y) = \lambda'\beta$$

and

$$\text{Var}(a'Y) \leq \text{Var}(b'Y) \text{ for any } b \in R^n \, .$$

Can we find such an estimator? The answer is *YES*, and this leads us to the Gauss-Markov theorem.

Theorem 5.6 *Gauss-Markov Theorem Consider the linear model*

$$Y = X\beta + \varepsilon$$

where $E(\varepsilon) = 0$ and Cov$(\varepsilon) = \sigma^2 I$, $\sigma^2 > 0$. If $\lambda'\beta$ is estimable, then the (unique) least squares estimate of $\lambda'\beta$ is the unique best linear unbiased estimator (BLUE) of $\lambda'\beta$.

Proof: Let $M = X(X'X)^{-}X'$. Since $\lambda'\beta$ is estimable, let $\lambda' = \rho'X$ for some ρ. We need to show that if $a'Y$ is an unbiased estimate of $\lambda'\beta$, then

$$\text{Var}(a'Y) \geq \text{Var}(\rho'MY) \text{ for any } a \in R^n \, .$$

Since $a'Y$ is unbiased for $\lambda'\beta$, $\lambda'\beta = E(a'Y) = a'E(Y) = a'X\beta$ for any β. Therefore, $\rho'X = \lambda' = a'X$. Now write

$$\begin{aligned} \text{Var}(a'Y) &= \text{Var}(a'Y - \rho'MY + \rho'MY) \\ &= \text{Var}(a'Y - \rho'MY) + \text{Var}(\rho'MY) \\ &+ 2\text{Cov}(a'Y - \rho'MY, \rho'MY) \end{aligned}$$

Note thatr Var$(a'Y - \rho'MY) \geq 0$.

$$\begin{aligned} &\text{Cov}(a'Y - \rho'MY, \rho'MY) = \text{Cov}((a' - \rho'M)Y, \rho'MY) \\ &= (a' - \rho'M)\text{Cov}(Y)(\rho'M)' = (a' - \rho'M)(\sigma^2 I)M\rho \\ &= \sigma^2(a' - \rho'M)M\rho = \sigma^2(a'M - \rho'M)\rho \end{aligned}$$

As shown above, $a'X = \rho'X$. This implies $a'X(X'X)^{-}X' = \rho'X(X'X)^{-}X'$, and therefore $a'M = \rho'M$. Thus it follows that $\sigma^2(a'M - \rho'M)\rho = 0$. This establishes $\text{Cov}(a'Y - \rho'MY, \rho'MY) = 0$.

Now we want to show that $\rho'MY$ is unique. We have just shown that for any linear unbiased estimate $a'Y$ of $\lambda'\beta$,

$$\text{Var}(a'Y) = \text{Var}(\rho'MY) + \text{Var}(a'Y - \rho'MY).$$

Thus if $a'Y$ is BLUE of $\lambda'\beta$, then it must be true that $\text{Var}(a'Y - \rho'MY) = 0$. It is clear that

$$\begin{aligned} 0 &= \text{Var}(a'Y - \rho'MY) \ = \ \text{Var}((a' - \rho'M)Y) \\ &= (a' - \rho'M)(\sigma^2 I)(a' - \rho'M)' \ = \ \sigma^2(a - M\rho)'(a - M\rho) \ = \ \sigma^2\|a - M\rho\|^2. \end{aligned}$$

Thus $\sigma^2\|a - M\rho\|^2 = 0$ if and only if $a - M\rho = 0$, which implies $a = M\rho$. Thus $\rho'MY$ is the unique BLUE of $\lambda'\beta$.

Remark 5.4 *$C(X)$ is sometimes called the estimation space and $C(X)^{\perp}$ is sometimes called the error space.*

5.2 Generalized Least Squares(GLS)

Consider the linear model

$$Y = X\beta + \varepsilon \tag{5.4}$$

where

$$\text{E}(\varepsilon) = 0 \ \text{ and } \ \text{Cov}(\varepsilon) = \sigma^2 V$$

and V is a known positive definite matrix.

Remark 5.5 *Applications with covariance matrices equal to $\sigma^2 V$*

1) *Repeated measures. V is typically unknown.*

2) *Split-plot design models. V is typically unknown.*

3) *When the observations consist of averages of independent random variables, say $\bar{y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$, $i = 1, \ldots, n$, then the covariance matrix $\sigma^2 V$ of the response vector is a diagonal matrix and the ith diagonal element of V is $1/m_i$. Here the elements in V are all known, and this problem fits into the generalized least squares framework.*

We want to characterize the least squares estimates of (β, σ^2). Since V is positive definite, we can write V as $V = QQ'$ for some nonsingular matrix Q ($Q = P\Lambda^{\frac{1}{2}}$, where P is an orthogonal matrix of orthonormal eigenvectors corresponding to eigenvalues of V). Now instead of working with (5.4) we can equivalently work with

$$Q^{-1}Y = Q^{-1}X\beta + Q^{-1}\varepsilon \; . \tag{5.5}$$

From (5.5), we notice that $E(Q^{-1}\varepsilon) = Q^{-1}E(\varepsilon) = Q^{-1}0 = 0$, and

$$\text{Cov}(Q^{-1}\varepsilon) = Q^{-1}(\sigma^2 V)Q'^{-1} = \sigma^2 Q^{-1}VQ'^{-1} = \sigma^2 I.$$

This transformation leads us back to the usual linear model. Thus a least squares estimate of β is a minimizer of

$$\begin{aligned} & (Q^{-1}Y - Q^{-1}X\beta)'(Q^{-1}Y - Q^{-1}X\beta) \\ = \;& (Y - X\beta)'(Q'^{-1}Q^{-1})(Y - X\beta) \\ = \;& (Y - X\beta)'V^{-1}(Y - X\beta). \end{aligned}$$

We now present a theorem characterizing estimability of $\lambda'\beta$.

Theorem 5.7 *a) $\lambda'\beta$ is estimable in (5.4) if and only if $\lambda'\beta$ is estimable in (5.5).*

b) *$\hat{\beta}$ is a generalized least squares estimate of β if and only if*

$$X(X'V^{-1}X)^{-}X'V^{-1}Y = X\hat{\beta}\ .$$

c) *For any estimable function $\lambda'\beta$, $\lambda' = \rho' X$, the unique generalized least squares estimate of $\lambda'\beta$ is $\rho' AY$, where $A = X(X'V^{-1}X)^{-}X'V^{-1}$. The unique generalized least squares estimate of $\mu = X\beta$ is AY.*

d) *For any estimable function $\lambda'\beta$, $\lambda' = \rho' X$, $\rho' AY$ is the BLUE of $\lambda'\beta$, where A is the matrix in part c).*

5.3 Properties of Generalized Least Squares Estimates

Theorem 5.8 *Let $A = X(X'V^{-1}X)^{-}X'V^{-1}$. Then*

a) *A is invariant with respect to the choice of generalized inverse.*

b) *A is a projection operator onto $C(X)$ along $\mathcal{N}(A)$.*

Proof:

a) Let $B = V^{-1/2}X$. Then we can write

$$\begin{aligned} A &= X(X'V^{-1}X)^{-}X'V^{-1} \\ &= V^{1/2}(V^{-1/2}X)(X'V^{-1/2}V^{-1/2}X)^{-}(X'V^{-1/2})V^{-1/2} \\ &= V^{1/2}B(B'B)^{-}B'V^{-1/2}\ . \end{aligned}$$

We know from previous results that $B(B'B)^{-}B'$ is the orthogonal projection operator onto $C(B)$ and is invariant with respect to the choice of generalized inverse. Since V is nonsingular, it follows that A is invariant with respect to the choice of generalized inverse.

b) We have $V = QQ'$, where Q is nonsingular. Consider the orthogonal projection operator onto $C(Q^{-1}X)$, denoted by P, which is given by

$$\begin{aligned} P &= (Q^{-1}X)\left((Q^{-1}X)'(Q^{-1}X)\right)^{-}(Q^{-1}X)' \\ &= (Q^{-1}X)(X'(QQ')^{-1}X)^{-}X'Q'^{-1} \\ &= Q^{-1}X(X'V^{-1}X)^{-}X'Q'^{-1}\ . \end{aligned}$$

By the definition of a projection operator, we have

$$\begin{aligned} & PQ^{-1}X = Q^{-1}X \\ \Leftrightarrow\ & Q^{-1}X(X'V^{-1}X)^{-}X'(QQ')^{-1}X = Q^{-1}X \\ \Leftrightarrow\ & Q^{-1}X(X'V^{-1}X)^{-}X'V^{-1}X = Q^{-1}X \\ \Leftrightarrow\ & Q^{-1}AX = Q^{-1}X \\ \Leftrightarrow\ & AX = X\ . \end{aligned} \tag{5.6}$$

It is now obvious by the definition of a projection operator that A is a projection operator onto $C(X)$. To formally finish the proof, let $v \in C(X)$, and write $X = (x_1, \ldots, x_n)$, where x_j is the jth column of X. Then $v = \alpha_1 x_1 + \ldots + \alpha_n x_n$, where the α_j's are scalars. Then

$$\begin{aligned} Av &= A(\alpha_1 x_1 + \ldots + \alpha_n x_n) \\ &= \alpha_1 Ax_1 + \ldots + \alpha_n Ax_n \\ &= \alpha_1 x_1 + \ldots + \alpha_n x_n \qquad \text{from 5.6) above} \\ &= v\ . \end{aligned}$$

Thus for any $v \in C(X)$, $Av = v$, and so A is a projection operator onto $C(X)$.

Remark 5.6 *1) We can see that A is not an orthogonal projection operator with respect to the inner product $x'y$. This is easily seen by noting that A is NOT symmetric.*

2) If one defines the inner product between two vectors x and y as $x'V^{-1}y$, then A is an orthogonal projection operator with respect to this inner product. In this inner product, x and y are orthogonal if $x'V^{-1}y = 0$, where $x \in C(X)$ and $y \in C(X)^{\perp}$.

To see that A is an orthogonal projection operator with respect to the $x'V^{-1}y$ inner product, we note that for any $x \in C(X)$, and $y \in C(X)^{\perp}$,

$$Ay = X(X'V^{-1}X)^{-}X'V^{-1}y = X(X'V^{-1}X)^{-}0 = 0$$

3) *The generalized least squares estimator equals the ordinary least squares estimator under certain conditions. This is given in the next two theorems.*

Theorem 5.9 *Suppose X is $n \times p$ of rank r, and V is a positive definite matrix. Then $C(V^{-1}X) = C(X)$ if and only if $C(VX) = C(X)$.*

The proof is left as an exercise.

Theorem 5.10 *Consider the linear model*

$$Y = X\beta + \varepsilon \ ,$$

where $E(\varepsilon) = 0$ and Cov$(\varepsilon) = \sigma^2 V$, where V is a known positive definite matrix and X has rank r. Consider estimating $\rho' X\beta$. Then

$$\rho' AY = \rho' MY$$

if and only if $C(VX) = C(X)$, where $M = X(X'X)^{-}X'$ and $A = X(X'V^{-1}X)^{-}X'V^{-1}$.

The proof is left as an exercise.

This theorem says that the generalized least squares estimate equals the ordinary least squares estimate if and only if $C(VX) = C(X)$.

Corollary 5.10.1 *Suppose X has full rank p. Then the generalized least squares estimate of β is given by $\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$, and the ordinary least squares estimate of β is given by $\hat{\beta} = (X'X)^{-1}X'Y$. Then $\tilde{\beta} = \hat{\beta}$ if and only if $C(VX) = C(X)$.*

<u>Proof:</u>

1) ($\Rightarrow$) Suppose $\tilde{\beta} = \hat{\beta}$. Then

$$\begin{aligned} & (X'V^{-1}X)^{-1}X'V^{-1}Y = (X'X)^{-1}X'Y \\ \Leftrightarrow\ & (X'V^{-1}X)^{-1}X'V^{-1} = (X'X)^{-1}X' \\ \Leftrightarrow\ & V^{-1}X(X'V^{-1}X)^{-1} = X(X'X)^{-1} \end{aligned}$$

Now let $T_1 = (X'V^{-1}X)^{-1}$, and $T_2 = (X'X)^{-1}$. The columns of T_1 and T_2 are bases for R^p. Now we have $V^{-1}XT_1 = XT_2$ which implies $V^{-1}X = XT_2T_1^{-1}$. Let $T_3 = T_2T_1^{-1}$. Then the columns of T_3 are a basis for R^p and $V^{-1}X = XT_3$, which implies $X = VXT_3$. Since T_3 is nonsingular, $C(VXT_3) = C(VX)$, and thus $C(X) = C(VX)$.

2) The proof of ($\Leftarrow$) is left as an exercise.

Estimation of σ^2

For the model in (5.4), the residual vector is given by $(I - A)Y$. Thus the estimate of σ^2 is some function of $(I - A)Y$. We search for an unbiased estimate. If we use the transformed model in (5.5), it is easy to find such an estimate. Recall that the transformed model is given by $Q^{-1}Y = Q^{-1}X\beta + Q^{-1}\varepsilon$. Let M^* denote the orthogonal projection operator onto $C(Q^{-1}X)$. Thus $M^* = (Q^{-1}X)(X'V^{-1}X)^{-}X'Q'^{-1}$. An unbiased estimator of σ^2 is given by

$$\hat{\sigma}^2 = \frac{\|(I - M^*)Q^{-1}Y\|^2}{n - r} \,.$$

We note that

$$\begin{aligned} & (I - M^*)Q^{-1} = Q^{-1} - Q^{-1}X(X'V^{-1}X)^{-}X'Q'^{-1}Q^{-1} \\ =\ & Q^{-1} - Q^{-1}X(X'V^{-1}X)^{-}X'V^{-1} \ =\ Q^{-1} - Q^{-1}A \ =\ Q^{-1}(I - A)\,. \end{aligned}$$

Thus $\|(I - M^*)Q^{-1}Y\|^2 = \|Q^{-1}(I - A)Y\|^2$, and therefore

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\|Q^{-1}(I - A)Y\|^2}{n - r} = \frac{Y'(I - A)'(QQ')^{-1}(I - A)Y}{n - r} \\ &= \frac{Y'(I - A)'V^{-1}(I - A)Y}{n - r}\,. \end{aligned}$$

Theorem 5.11

$$V^{-1}(I-A) = (I-A)'V^{-1}(I-A)\ .$$

<u>Proof:</u> We have

$$(I-A)'V^{-1}(I-A) = V^{-1}(I-A) - A'V^{-1}(I-A).$$

The proof will be completed if we can show $A'V^{-1}(I-A) = 0$ which is equivalent to show $A'V^{-1} = A'V^{-1}A$. Now

$$\begin{aligned} A'V^{-1}A &= \left[V^{-1}X(X'V^{-1}X)^{-}X'\right]V^{-1}\left[X(X'V^{-1}X)^{-}X'V^{-1}\right] \\ &= V^{-1}X(X'V^{-1}X)^{-}(X'V^{-1}X)(X'V^{-1}X)^{-}X'V^{-1} \\ &= V^{-1}X(X'V^{-1}X)^{-}X'V^{-1} \ = \ A'V^{-1}\ . \end{aligned}$$

Theorem 5.12

$$AVA = AV = VA'\ .$$

The proof is similar to the one given above.

Remark 5.7 *Covariance Matrices of Estimates*

a) *The BLUE of $\rho'X\beta$ in the usual linear model (i.e., when Cov$(\varepsilon) = \sigma^2 I$) is $\rho'MY$.*

$$\begin{aligned} \textit{Cov}(\rho'MY) &= (\rho'M)\textit{Cov}(Y)(\rho'M)' \\ &= (\rho'M)(\sigma^2 I)(\rho'M)' = \sigma^2\rho'M\rho. \end{aligned}$$

b) *The covariance matrix of the residual vector for the usual linear model is*

$$\begin{aligned} \textit{Cov}((I-M)Y) &= (I-M)(\sigma^2 I)(I-M) \\ &= \sigma^2(I-M)\ . \end{aligned}$$

Note that the covariance matrix of the residuals is singular. It is an $n \times n$ positive semidefinite matrix of rank $n-r$.

c) *For generalized least squares,*

$$\begin{aligned} \textit{Cov}(\rho'AY) &= (\rho'A)(\sigma^2 V)(\rho'A)' \\ &= \sigma^2\rho'AVA'\rho. \end{aligned}$$

Also,

$$Cov((I-A)Y) = \sigma^2(I-A)V(I-A)' \; .$$

d) *For the usual linear model with an intercept, the residuals sum to 0. To see this, let* $J = (1, \ldots, 1)'$ *denote the* $n \times 1$ *vector of ones. Then*

$$\begin{aligned} J'(I-M)Y &= J'Y - J'MY \\ &= J'Y - J'Y = 0 \; . \end{aligned}$$

We note that since $J \in C(X)$, $J \in C(M)$ *since* $C(X) = C(M)$, *so* $MJ = J$, *which implies* $(MJ)' = J'M = J'$.

We have a similar result for generalized least squares. That is

$$\begin{aligned} J'(I-A)Y &= J'Y - J'AY \\ &= J'Y - J'Y = 0 \; . \end{aligned}$$

Remark 5.8 *We have NOT made ANY distributional assumptions on* ε *to obtain least squares estimates, generalized least squares estimates, or BLUE's.*
We need distributional assumptions to construct tests of hypotheses, confidence regions, and prediction regions.

5.4 Maximum Likelihood Estimation

We want to examine maximum likelihood estimation of β and σ^2 in the linear model.

Consider the usual linear model

$$Y = X\beta + \varepsilon \; ,$$

where $\varepsilon \sim N_n(0, \sigma^2 I)$ and X has rank r. Using properties of MVN, this implies that

$$Y \sim N_n(X\beta, \sigma^2 I) \ .$$

The joint density of $Y = (Y_1, \ldots, Y_n)'$ is given by

$$f(Y) = (2\pi)^{-n/2}\sigma^{-n} \exp\left\{\frac{-1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\} \ .$$

We note here that $\mid \sigma^2 I \mid^{-1/2} = \sigma^{-n}$. The likelihood function of the parameters, denoted $L(\beta, \sigma)$, is any function proportional to the density function $f(Y)$. That is

$$L(\beta, \sigma) \propto f(Y) \ .$$

Dropping the $(2\pi)^{-n/2}$ term, we have

$$L(\beta, \sigma) = \sigma^{-n} \exp\left\{\frac{-1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\} \ .$$

The maximizers of L are called the maximum likelihood estimates (MLE's).

Maximizing $L(\beta, \sigma)$ is equivalent to maximizing $\ell(\beta, \sigma) = \log(L(\beta, \sigma))$. We have

$$\ell(\beta, \sigma) = -n\log(\sigma) - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta) \ .$$

Now maximizing $\ell(\beta, \sigma)$ with respect to β is equivalent to minimizing $g(\beta) = (Y - X\beta)'(Y - X\beta)$ with respect to β. This is just the least squares criterion. Thus the MLE of β reduces to the least squares criterion, and thus the MLE of β for the usual linear model, denoted $\hat{\beta}_{ml}$ satisfies

$$X\hat{\beta}_{ml} = MY$$

where $M = X(X'X)^{-}X'$ is the orthogonal projection operator onto $C(X)$.

To find the MLE of σ, we substitute the MLE of β into $\ell(\beta, \sigma)$. Thus

$$\begin{aligned}
\ell(\hat{\beta}_{ml}, \sigma) &= -n\log(\sigma) - \frac{1}{2\sigma^2}(Y - X\hat{\beta}_{ml})'(Y - X\hat{\beta}_{ml}) \\
&= -n\log(\sigma) - \frac{1}{2\sigma^2}(Y - MY)'(Y - MY) \\
&= -n\log(\sigma) - \frac{1}{2\sigma^2}Y'(I - M)Y\ .
\end{aligned}$$

Now we take derivatives with respect to σ and set equal to 0. Thus

$$\begin{aligned}
\frac{\partial\ell(\hat{\beta}_{ml}, \sigma)}{\partial\sigma} &= \frac{-n}{\sigma} + \frac{1}{\sigma^3}(Y'(I - M)Y) = 0 \\
&\Leftrightarrow -n\sigma^2 + Y'(I - M)Y = 0 \\
&\Leftrightarrow \sigma^2 = \frac{Y'(I - M)Y}{n}
\end{aligned}$$

Thus the maximum likelihood estimate of σ^2 is

$$\hat{\sigma}^2_{ml} = \frac{Y'(I - M)Y}{n}\ .$$

We see that $\hat{\sigma}^2_{ml}$ is biased for estimating σ^2. In particular,

$$E(\hat{\sigma}^2_{ml}) = \frac{n - r}{n}\sigma^2,$$

which converges to σ^2 as $n \to \infty$. Thus $\hat{\sigma}^2_{ml}$ is asymptotically unbiased.

5.5 Minimum Variance Unbiased Estimation

We saw by the Gauss-Markov theorem that the least squares estimator of $\rho' X\beta$ is the unique minimum variance unbiased linear estimator. Now we will show that if $\varepsilon \sim N_n(0, \sigma^2 I)$, then the least squares estimator is the uniform minimum variance unbiased estimator (UMVUE). This is a stronger result since the UMVUE is not restricted to linear estimators. It covers the class of ALL unbiased estimators.

To show this, we need to establish the notion of completeness.

Definition 5.6 *Suppose $T(Y)$ is a vector valued statistic in Y. Then $T(Y)$ is said to be a complete sufficient statistic for the family of distributions indexed by $\theta \in \Theta$, if $T(Y)$ is sufficient and*

$$E\left[f(T(Y))\right] = 0 \;\;\Rightarrow\;\; f(T(Y)) = 0$$

with probability 1 for all $\theta \in \Theta$, where Θ denotes the parameter space.

Theorem 5.13 *If $T(Y)$ is a complete sufficient statistic, then $f(T(Y))$ is the unique UMVUE of $E(f(T(Y)))$.*

5.6 Complete Sufficient Statistics for Exponential Families

Theorem 5.14 *Let $\theta = (\theta_1, \ldots, \theta_p)'$ and let Y be a random vector with density*

$$f(Y) = h(Y)c(\theta)\exp\left\{\sum_{i=1}^{p} \theta_i T_i(Y)\right\} ,$$

then $T(Y) = (T_1(Y), \ldots, T_p(Y))'$ is a complete sufficient statistic for θ, if neither θ nor $T(Y)$ satisfy any linear constraints.

In linear models when the X matrix is less than full rank, the components of β have constraints. That is, the model is overparameterized and some components of β are redundant.

To overcome this, we consider a reparametrization to remove the redundant components of β. Suppose $r(X) = r$ and let Z be a matrix whose columns form a basis for $C(X)$. Thus for some matrix A, $X = ZA$ where Z is $n \times r$ of rank r and A is $r \times p$.

Let $\lambda'\beta$ be an estimable function. Then $\lambda' = \rho' X$ for some $\rho \in R^n$. Thus

$$\lambda'\beta = \rho' X\beta = \rho' ZA\beta \; .$$

Now let $\gamma = A\beta$. Thus γ is $r \times 1$, Now consider the reparametrized linear model

$$Y = Z\gamma + \varepsilon$$

where $\varepsilon \sim N_n(0, \sigma^2 I)$. Z is $n \times r$ of full rank r. The least squares estimate of $\lambda'\beta = \rho' Z\gamma$ is $\rho' MY$ regardless of the rank of the model. That is, regardless of the reparameterized model or the originally parameterized model.

We want to now show that $\rho' MY$ is the unique minimum variance unbiased estimate of $\rho' X\beta$. The density of Y is given by

$$\begin{aligned} f(Y) &= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{\frac{-1}{2\sigma^2}(Y - Z\gamma)'(Y - Z\gamma)\right\} \\ &= (2\pi)^{-n/2}\sigma^{-n} \exp\left\{\frac{-1}{2\sigma^2}(\gamma' Z' Z\gamma)\right\} \\ &\times \; \exp\left\{\frac{-1}{2\sigma^2}(Y'Y) + \frac{\gamma'}{\sigma^2}(Z'Y)\right\} \\ &= h(Y)c(\gamma, \sigma^2) \exp\left\{\frac{-1}{2\sigma^2}(Y'Y) + \frac{\gamma'}{\sigma^2}(Z'Y)\right\} \; , \end{aligned}$$

$$\text{where} \quad h(Y) = (2\pi)^{-n/2} \quad \text{and} \quad c(\gamma, \sigma^2) = \sigma^{-n} \exp\left\{\frac{-1}{2\sigma^2}(\gamma' Z' Z\gamma)\right\} \; .$$

This density is of the form in the Theorem, and there are NO restrictions on γ. It follows that $(Y'Y, Z'Y)$ are complete sufficient statistics for θ.

An unbiased estimate of $\lambda'\beta = \rho' X\beta$ is $\rho' MY = \rho' Z(Z'Z)^{-1}Z'Y$. Thus $\rho' MY$ is a function of the complete sufficient statistic so it is the unique minimum variance unbiased estimator of $\rho' X\beta$.

Moreover, $Y'(I - M)Y/(n - r)$ is an unbiased estimate of σ^2, and $Y'(I - M)Y = Y'Y - (Y'Z)(Z'Z)^{-1}Z'Y$ is a function of the complete sufficient statistic $(Y'Y, Y'Z)$. Therefore it is the unique minimum variance unbiased estimator of σ^2. We now have the following result.

Theorem 5.15 *Suppose*

$$Y = X\beta + \varepsilon$$

where $\varepsilon \sim N_n(0, \sigma^2 I)$ *and* $r(X) = r$. *Then*

$$\frac{Y'(I - M)Y}{n - r} \text{ is the unique UMVUE of } \sigma^2 \text{ and}$$

$$\rho' MY \text{ is the unique UMVUE of } \rho' X\beta.$$

5.7 Sampling Distributions of Estimates

Consider the usual linear model

$$Y = X\beta + \varepsilon$$

where $\varepsilon \sim N_n(0, \sigma^2 I)$, which implies $Y \sim N_n(X\beta, \sigma^2 I)$.

We want to obtain the sampling distributions of the least squares and maximum likelihood estimates.
Suppose $\Lambda'\beta$ is an estimable vector of linear functions of β, where Λ is a $p \times s$ matrix. We have $\Lambda' = P'X$ for some $n \times s$ matrix P.

i) $E(P'MY) = P'ME(Y) = P'MX\beta = P'X\beta$

ii) $\text{Cov}(P'MY) = (P'M)\text{Cov}(Y)(P'M)' = P'M(\sigma^2 I)MP = \sigma^2 P'MP.$

Thus

$$P'MY \sim N_s(P'X\beta, \sigma^2 P'MP) \; .$$

Note that we can write $P'MP = P'X(X'X)^{-}X'P = \Lambda'(X'X)^{-}\Lambda$, and therefore an alternative representation of the sampling distribution of $P'MY$ is

iii) $P'MY \sim N_s(\Lambda'\beta, \sigma^2\Lambda'(X'X)^{-}\Lambda)$.

Remark 5.9 *Some special cases of interest*

a) *Let $P = I_{n\times n}$. Then the least squares estimate of $E(Y) = \mu = X\beta$ is MY. Thus*

$$MY \sim N_n(X\beta, \sigma^2 M)\ .$$

b) *Suppose X has full rank* p. *Then β is estimable and the unique least squares of β is $\hat{\beta} = (X'X)^{-1}X'Y$. We have*

$$\begin{aligned} E(\hat{\beta}) &= E\left((X'X)^{-1}X'Y\right) \\ &= (X'X)^{-1}X'E(Y) = (X'X)^{-1}X'(X\beta) \\ &= (X'X)^{-1}(X'X)\beta \\ &= \beta\ . \end{aligned}$$

$$\begin{aligned} \mathit{Cov}(\hat{\beta}) &= \left((X'X)^{-1}X'\right)\mathit{Cov}(Y)\left((X'X)^{-1}X'\right)' \\ &= (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}(X'X)(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}\ . \end{aligned}$$

Thus

$$\hat{\beta} \sim N_p(\beta, \sigma^2(X'X)^{-1})\ .$$

The sampling distribution for the estimator of σ^2 is obtained as follows.

Since $Y \sim N_n(X\beta, \sigma^2 I)$, and $I - M$ is an orthogonal projection operator of rank $n - r$, it follows by an earlier theorem that

$$\frac{1}{\sigma^2}\left(Y'(I-M)Y\right) \sim \chi^2(n-r,\gamma)\ ,$$

where

$$\gamma = \frac{(X\beta)'(I-M)(X\beta)}{2\sigma^2} = \frac{\beta' X'(I-M)X\beta}{2\sigma^2} = 0\ ,$$

since $(I - M)X = 0$.

Thus

$$\frac{1}{\sigma^2}\left(Y'(I-M)Y\right) \sim \chi^2(n-r)\ .$$

We can also write

$$Y'(I-M)Y \sim \sigma^2\chi^2(n-r)\ .$$

Now consider the linear model

$$Y = X\beta + \varepsilon$$

where $\varepsilon \sim N_n(0, \sigma^2 V)$ where V is a known positive definite matrix. We have the following results.

i) $\rho' AY$ is the unique UMVUE of $\rho' X\beta$, where $A = X(X'V^{-1}X)^{-}X'V^{-1}$.

ii) $\rho' AY \sim N_1(\rho' X\beta, \sigma^2\rho' AVA'\rho)$.

iii) $P'AY$ is the unique UMVUE of $P'X\beta$, where P is an $n \times s$ matrix of constants. Also,

$$P'AY \sim N_s(P'X\beta, \sigma^2 P'AVA'P)\ .$$

iv) $Y'(I - A)'V^{-1}(I - A)Y/(n - r)$ is the unique UMVUE of σ^2.

v) Any generalized least squares estimate of β is an MLE of β.

vi) If X has full rank p, then the UMVUE of β is

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y \ ,$$

and

$$\hat{\beta} \sim N_p(\beta, \sigma^2(X'V^{-1}X)^{-1}) \ .$$

Chapter 6

HYPOTHESES TESTING

We now consider the problem of hypothesis testing in the linear model.

Consider the linear model

$$Y = X\beta + \varepsilon \tag{6.1}$$

where

$$E(\varepsilon) = 0 \ , \ \text{Cov}(\varepsilon) = \sigma^2 I \ , \tag{6.2}$$

and Y is an $n \times 1$ random vector, X is an $n \times p$ fixed matrix of rank $r \leq p$, $\beta = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of regression coefficients, and ε is an $n \times 1$ vector of random errors.

In hypothesis testing terminology, we call $C(X)$ the estimation space and $C(X)^\perp$ the error space.

Since $E(Y) = X\beta$, model (6.1) specifies that $E(Y) \in C(X)$ and Cov$(Y) = \sigma^2 I$.

Goal: testing a linear model against a reduced linear model;

That is, we are interested in testing nested linear models. This is the only hypothesis testing situation we will consider. We will develop the null and alternative hypotheses in terms of vector spaces.

All hypothesis testing in linear models reduces to specifying a constraint on the estimation space $C(X)$. We usually start with the "full" model in (6.1) and we wish to reduce this model somehow. That is, we want to know if a simpler, more parsimonious model is acceptable. Thus, we consider the reduced model

$$Y = X_0\gamma_0 + \varepsilon \tag{6.3}$$

where $\varepsilon \sim N_n(0, \sigma^2 I)$ and $C(X_0) \subset C(X)$.

Since $C(X_0) \subset C(X)$, we say that the model in (6.3) is *nested* in the model in (6.1). Model (6.3) specifies that

$$E(Y) \in C(X_0).$$

Examples of nested models

1) One-way ANOVA

The "full" model for one-way ANOVA is

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij},$$

where μ denotes the grand mean and α_i is the ith treatment effect, $j = 1, \ldots, n_i, i = 1, \ldots, t$. We often wish to test the hypothesis of no treatment effect, that is, $H_0 : \alpha_1 = \ldots = \alpha_t$. Thus the reduced model becomes

$$Y_{ij} = \mu + \varepsilon_{ij}.$$

2) Multiple linear regression

Suppose the full model is

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

where X_1 is $n \times r$, X_2 is $n \times (p - r)$, β_1 is $r \times 1$ and β_2 is $(p - r) \times 1$. Often we are interested in testing the hypothesis $H_0 : \beta_1 = 0$, so that the reduced model becomes

$$Y = X_2\beta_2 + \varepsilon.$$

In this example, the X matrix for the full model is $X = (X_1, X_2)$ and the X matrix for the reduced model is $X_0 = X_2$.

Null and alternative hypotheses

We can now express the null and alternative hypotheses for testing model (6.3) against model (6.1) in terms of $E(Y)$. The null hypothesis specifies $E(Y) \in C(X_0)$. To make the null and alternative disjoint hypotheses, the alternative specifies that $E(Y) \in C(X)$ and $E(Y) \notin C(X_0)$. But this is equivalent to $H_a : E(Y) \in C(X) \cap C(X_0)^c$. In summary, we have

$$\begin{aligned} &H_0 : E(Y) \in C(X_0) \\ &H_a : E(Y) \in C(X) \cap C(X_0)^c. \end{aligned}$$

F-test

We now give a heuristic argument for the F test for testing model (6.3) against model (6.1).

Let $M = X(X'X)^- X'$ denote the orthogonal projection operator onto $C(X)$ and $M_0 = X_0(X_0'X_0)^- X_0'$ denotes the orthogonal projection operator onto $C(X_0)$. Since $C(M_0) \subset C(M)$,

1. $M - M_0$ is also an orthogonal projection operator.
2. $C(M - M_0)$ is precisely the subspace within M that is orthogonal to M_0.

Under the full model in (6.1), the unique UMVUE of $\mu = E(Y)$ is MY, and under model (6.3), the unique UMVUE of $\mu = E(Y)$ is M_0Y.

If model (6.3) is correct, then MY and M_0Y are estimates of the same quantity $E(Y)$, since the validity of model (6.3) implies the validity of model (6.1). Thus, if model (6.3) is true, $MY - M_0Y = (M - M_0)Y$ should be small in some sense.

On the other hand, a large difference between MY and M_0Y suggests that MY and M_0Y are NOT estimating the same quantity. If M_0Y is not estimating $E(Y)$, then model (6.3) cannot be correct because model (6.3) implies that M_0Y is an estimate of $E(Y)$.

The decision about whether model (6.3) is correct depends on whether the vector $(M - M_0)Y$ is large. An obvious measure of size is the squared length of this vector given by

$$\|(M - M_0)Y\|^2 = Y'(M - M_0)Y.$$

If we adjust for the relative sizes of the subspaces $C(M)$ and $C(M_0)$, we divide $\|(M - M_0)Y\|$ by $r(M - M_0)$. Since Y is random, our measure of size is

$$\mathrm{E}\left(\frac{\|(M - M_0)Y\|^2}{r(M - M_0)}\right) = \mathrm{E}\left(\frac{Y'(M - M_0)Y}{r(M - M_0)}\right). \tag{6.4}$$

So, how large this measure will be when model (6.3) is correct and when it is not correct? To answer this, we need to compute the expectation of (6.4) above under both of these scenarios. We have

1. If model (6.3) is NOT true

$$\begin{aligned}
&\mathrm{E}\left(\frac{Y'(M - M_0)Y}{r(M - M_0)}\right)\\
=&\mathrm{tr}\left(\frac{\sigma^2(M - M_0)}{r(M - M_0)}\right) + \frac{(X\beta)'(M - M_0)(X\beta)}{r(M - M_0)}\\
=&\sigma^2\frac{r(M - M_0)}{r(M - M_0)} + \frac{\|(I - M_0)X\beta\|^2}{r(M - M_0)}\\
=&\sigma^2 + \frac{\|(I - M_0)X\beta\|^2}{r(M - M_0)}.
\end{aligned}$$

2. If model (6.3) is TRUE then $X\beta$ is replaced by $X_0\gamma_0$ in $i)$ and we have

$$\frac{\|(I - M_0)X\beta\|^2}{r(M - M_0)} = 0$$

and the formula reduces to

$$\mathrm{E}\left(\frac{Y'(M-M_0)Y}{r(M-M_0)}\right)=\sigma^2. \tag{6.5}$$

If model (6.3) were correct, the quadratic form should be close to σ^2, and thus the ratio

$$\left(\frac{Y'(M-M_0)Y}{r(M-M_0)\sigma^2}\right)\sim 1.$$

If the ratio above is much larger than 1, this would indicate that model (6.3) is not correct. Typically, σ^2 is not known in practice, and thus needs to be estimated. Since we always assume the full model to be "true", we estimate σ^2 from the full model. The estimate is thus the UMVUE of σ^2 based on the full model. We have

$$\hat{\sigma}^2 = MSE = \frac{\|(I-M)Y\|^2}{r(I-M)} = \frac{Y'(I-M)Y}{r(I-M)}.$$

Thus, the test statistic F for the null hypothesis $H_0 : \mathrm{E}(Y) \in C(X_0)$ is

$$F=\left(\frac{Y'(M-M_0)Y}{MSE\cdot r(M-M_0)}\right).$$

The term

$$\|(I-M_0)X\beta\|^2$$

is crucial in evaluating the behavior of the test statistic when model (6.3) is not correct. This term is a part of the noncentrality parameter. Specifically, the noncentrality parameter for the test above is

$$\gamma=\frac{\|(I-M_0)X\beta\|^2}{2\sigma^2}=\frac{\|(M-M_0)X\beta\|^2}{2\sigma^2}.$$

Note: The size of the noncentrality parameter plays a major role in deciding on the validity of model (6.3). If γ is large, this implies that model (6.3) is not correct and if $\gamma = 0$, this implies that model (6.3) is correct. We are now led to the following theorem.

Theorem 6.1 *Consider the usual linear model*

$$Y = X\beta + \varepsilon,$$

where $\varepsilon \sim N_n(0, \sigma^2 I)$. Consider the reduced model

$$Y = X_0\gamma_0 + \varepsilon,$$

where $C(X_0) \subset C(X)$. We wish to test the hypothesis

$$\begin{aligned} &H_0 : E(Y) \in C(X_0) \\ &H_a : E(Y) \in C(X) \cap C(X_0)^c. \end{aligned}$$

Let $M_0 = X_0(X_0'X_0)^- X_0$ be the orthogonal projection operator onto $C(X_0)$ and $M = X(X'X)^- X'$ is the orthogonal projection operator onto $C(X)$. Further, let X be $n \times p$ with $r = r(X)$ and $r_0 = r(X_0), r_0 < r$. Then, under H_a,

$$F = \frac{\|(M0 - M0)Y\|^2/(r - r_0)}{\|(I - M)Y\|^2/(n - r)} \sim F(r - r_0, n - r, \gamma)$$

where $\gamma = \|(I - M_0)X\beta\|^2/(2\sigma^2) = \|(M - M_0)X\beta\|^2/(2\sigma^2)$.

If model (6.3) is assumed correct, then $\gamma = 0$ and

$$F = \frac{\|(M - M0)Y\|^2/(r - r_0)}{\|(I - M)Y\|^2/(n - r)} \sim F(r - r_0, n - r).$$

Thus an α level test of the hypothesis

$$\begin{aligned} H_0 &: \boldsymbol{E}(Y) \in C(X_0) \\ H_a &: \boldsymbol{E}(Y) \in C(X) \cap C(X_0)^c \end{aligned}$$

rejects H_0 if

$$F = \frac{\|(M - M_0)Y\|^2/(r - r_0)}{\|(I - M)Y\|^2/(n - r)} > F(1 - \alpha, r - r_0, n - r),$$

where $F(1-\alpha, r-r_0, n-r)$ is the $(1-\alpha) \times 100\%$ percentile of the F distribution with $(r - r_0, n - r)$ degrees of freedom.

Proof: What we need to show: i) quadratic form in the numerator is noncentral chi-square, ii) the denominator quadratic form is central chi-square, and iii) the two quadratic forms are independent.

1. Recall the theorem on quadratic forms states that if $Y \sim N_n(\mu, \sigma^2 I)$, then $(Y'MY)/\sigma^2 \sim \chi^2(r, \gamma)$ if and only if M is an orthogonal projection operator of rank r, and $\gamma = \|M\mu\|^2/(2\sigma^2) = (\mu' M\mu)/(2\sigma^2)$.

   By the theorem, we have $Y'(M - M_0)Y \sim \sigma^2\chi^2(r - r_0, \gamma)$, where $\gamma = \|(M - M_0)X\beta||^2/(2\sigma^2)$. For the denominator quadratic form, we have $Y'(I - M)Y \sim \sigma^2\chi^2(n - r)$. The denominator quadratic form is a central chi-square since (WHY?)

2. To prove independence of the quadratic forms, we use our theorems on independence of quadratic forms. It suffices to show that $(M - M_0)(I - M) = 0$. We see that

$$\begin{aligned} (M - M_0)(I - M) &= M - M^2 - M_0 + M_0 M \\ &= M - M - M_0 + M_0 = 0. \end{aligned}$$

Note: The F test given here is very general in that it applies in any hypothesis testing situation in which we have nested models.

6.1 Testing Linear Parametric Functions

Consider tests of hypotheses with linear constraints on the parameter vector β. Consider the usual linear model

$$Y = X\beta + \varepsilon,$$

where $\varepsilon \sim N_n(0, \sigma^2 I)$.

Suppose we want to test a hypothesis concerning an estimable function $\Lambda'\beta$, where $\Lambda' = P'X$, and P' is an $s \times n$ matrix of constants. That is, we want to test $H_0 : P'X\beta = 0$. More formally, the hypotheses are

$$\begin{aligned} H_0 &: \mathrm{E}(Y) \in C(X) \text{ and } P'X\beta = 0 \\ H_a &: \mathrm{E}(Y) \in C(X) \text{ and } P'X\beta \neq 0. \end{aligned}$$

Examples

a) Suppose $\beta_1 - \beta_2$ is estimable and we wish to test $H_0 : \beta_1 - \beta_2 = 0$. Here, $s = 1$ and $\lambda' = (1, -1, 0, \ldots, 0)$.

b) Suppose $\begin{pmatrix} \beta_1 + \beta_3 \\ \beta_2 \end{pmatrix}$ is estimable and we wanted to to test the hypothesis

$$H_0 : \beta_1 + \beta_3 = 0 \text{ and } \beta_2 = 0.$$

What is Λ?

Note: The condition $P'X\beta = 0$ imposes a linear constraint on β.

We need to find the reduced model X_0 that corresponds to this linear constraint. Since $\mathrm{E}(Y) = X\beta$, $P'X\beta = 0$ imposes a restriction on $\mathrm{E}(Y)$. This restriction is $\mathrm{E}(Y) \in \mathcal{N}(P') = C(P)^{\perp}$. Thus the null hypothesis may now be stated as

$$H_0 : \mathrm{E}(Y) \in C(X) \cap C(P)^{\perp} \tag{6.6}$$

We need to find a matrix X_0 so that $C(X_0) = C(X) \cap C(P)^{\perp}$, then we can use the F test derived earlier for testing nested linear models.

Note: The choice of X_0 corresponding to (6.6) is NOT unique since P is not unique.

Let M denote the orthogonal projection operator onto $C(X)$. We can decompose P into $P = MP + (I - M)P$, and thus

$$P'X\beta = P'MX\beta + P'(I - M)X\beta = P'MX\beta,$$

since $(I - M)X = 0$.

Thus $P'X\beta = 0$ if and only if $P'MX\beta = 0$ and therefore $E(Y) \perp C(P)$ if and only if $E(Y) \perp C(MP)$. The null hypothesis then can be stated as

$$H_0 : E(Y) \in C(X) \cap C(MP)^{\perp}$$

Note: $C(MP)$ is a subspace in $C(M)$ obtained by projecting P onto $C(M)$. Some choices for X_0 are stated in the following two theorems.

Let M_{MP} denote the orthogonal projection operator onto $C(MP)$. By definition, M_{MP} is given by

$$M_{MP} = MP(P'MP)^{-}P'M$$

We now have the following theorem

Theorem 6.2

$$C((I - M_{MP})X) = C(X) \cap C(MP)^{\perp}$$

Proof:
1) $\Rightarrow$
We want to show $C((I - M_{MP})X) \subset C(X) \cap C(MP)^{\perp}$.

Let $\nu \in C((I - M_{MP})X)$. Then $\mathbf{u} = (I - M_{MP})X_z$ for some $z \in R^p$. But $(I - M_{MP})X_z = X_z - M_{MP}X_z$. Now $X_z \in C(X)$ and $M_{MP}X_z \in C(X)$ by definition of column space and since $M_{MP} = MP(P'MP)^{-}P'M$ and $M = X(X'X)^{-}X'$.

By definition of subspace, the difference of two vectors in a subspace, results in a vector in the same subspace, and thus $\nu \in C(X)$.

To show that $\nu \in C(MP)^{\perp}$, notice that $I - M_{MP}$ is the orthogonal projection operator onto $C(MP)^{\perp}$, and thus by definition $(I - M_{MP})X_z$ is a vector in $C(MP)^{\perp}$ for any $z \in R^p$. Thus $\nu \in C(MP)^{\perp}$, and thus $\nu \in C(X) \cap C(MP)^{\perp}$.

2) $\Leftarrow$
We need to show $C(X) \cap C(MP)^{\perp} \subset C((I - M_{MP})X)$.

Let $\nu \in C(X) \cap C(M_{MP})^{\perp}$. Then $\nu = Xw$ for some w and $M_{MP}\nu = 0$. Thus, $\nu = (I - M_{MP})\nu = (I - M_{MP})Xw$, so $\nu \in C((I - M_{MP})X)$.

Remark 6.1 *1. This theorem implies that one choice of X_0 is $X_0 = (I - M_{MP})X$.*

2. Since $C(X) = C(M)$, we can replace $C(X)$ with $C(M)$ in the theorem above to obtain

$$C(X) \cap C(MP)^{\perp} = C(M) \cap C(MP)^{\perp} = C(M - M_{MP}).$$

Thus another choice for X_0 is $X_0 = M - M_{MP}$. This choice of X_0 is the most common one and will be the one we use.

Since $C(X) = C(M)$, $C(X)$ can be replaced by $C(M)$ in the theorem above to obtain

$$C(X) \cap C(MP)^{\perp} = C(M) \cap C(MP)^{\perp} = C(M - M_{MP}).$$

Therefore another choice for X_0 is $X_0 = M - M_{MP}$. This choice of X_0 is the most common one and will be the one we use.

Remark 6.2 *X_0 is not unique and need not have the same dimension as other X_0's.*

Example

$X_0 = (I - M_{MP})X$ is an $n \times p$ matrix and $X_0 = M - M_{MP}$ is an $n \times n$ matrix.

A test of the reduced model for the hypothesis

$$H_0 : \mathrm{E}(Y) \in C(X) \cap C(MP)^{\perp}$$

has numerator sum of squares given by

$$\begin{aligned} Y'(M - M_0)Y &= Y'(M - (M - M_{MP}))Y' \\ &= Y'M_{MP}Y \\ &= \|M_{MP}Y\|^2. \end{aligned}$$

Thus, the F test is given by

$$F = \frac{\|M_{MP}Y\|^2/r(M_{MP})}{\|(I - M)Y\|^2/r(I - M)} \sim F(r(M_{MP}), r(I - M), \gamma),$$

where $\gamma = \|M_{MP}X\beta\|^2/2\sigma^2$. Under the null hypothesis $\gamma = 0$, and the test statistic has a central F distribution.

Recall that the original hypothesis was $H_0 : \Lambda\beta = 0$, where $\Lambda' = P'X$. We can write the F statistic in terms of Λ and $\hat{\beta}$, where $\hat{\beta}$ is an MLE of β. We have

$$\begin{aligned} Y'M_{MP}Y &= Y'MP(P'MP)^{-}P'MY \\ &= \hat{\beta}'\Lambda(P'X(X'X)^{-}X'P)^{-}\Lambda^{-}\hat{\beta} \\ &= \hat{\beta}'\Lambda(\Lambda'(X'X)^{-}\Lambda)^{-}\Lambda'\hat{\beta}. \end{aligned}$$

Remark 6.3 $r(M_{MP}) = r(\Lambda)$*, and thus the* F *statistic becomes*

$$F = \frac{\hat{\beta}'\Lambda(\Lambda'(X'X)^{-}\Lambda)^{-}\Lambda'\hat{\beta}/r(\Lambda)}{MSE}$$

where $MSE = \|(I - M)Y\|^2/r(I - M)$*. The noncentrality parameter can be written as*

$$\gamma = \beta'\Lambda(\Lambda'(X'X)^{-}\Lambda)^{-}\Lambda'\beta.$$

We note here that

$$\text{Cov}(\Lambda'\hat{\beta}) = \sigma^2\Lambda'(X'X)^{-}\Lambda.$$

We can also write the test statistic as

$$F = \frac{(\Lambda'\hat{\beta})'(\hat{\text{Cov}}(\Lambda'\hat{\beta})^{-})(\Lambda'\hat{\beta})}{r(\Lambda)}.$$

Thus,

$$\frac{(\Lambda'\hat{\beta})'(\hat{\text{Cov}}(\Lambda'\hat{\beta})^{-})(\Lambda'\hat{\beta})}{r(\Lambda)} \sim F(r(\Lambda), r(I - M), \gamma)$$

where

$$\gamma = \beta'\Lambda(\Lambda'(X'X)^{-}\Lambda)^{-}\Lambda'\beta.$$

The above form of the F test tells us that in order to obtain the F test of the hypothesis $H_0 : \Lambda'\beta = 0$, we need

1. The UMVUE of $\Lambda'\beta$, denoted $\Lambda'\hat{\beta}$.
2. The $\text{Cov}(\Lambda'\hat{\beta})$.
3. $r(\Lambda)$.
4. $MSE = \|(I - M)Y\|^2/r(I - M)$.

Remark 6.4 *A special case of the general F test is testing the hypothesis*

$$H_0 : \lambda'\beta = 0.$$

where $\lambda' = \rho' X, \rho \in R^n$. The F test for this hypothesis is

$$F = \frac{(\lambda'\hat{\beta})^2}{MSE\lambda'(X'X)^{-}\lambda} \sim F(1, r(I - M), \gamma)$$

where

$$\gamma = \frac{(\lambda'\beta)^2}{2\sigma^2\lambda'(X'X)^{-}\lambda}.$$

6.2 Generalized Hypothesis Test Procedure

Suppose we wish to test

$$H_0 : \Lambda'\beta = d \tag{6.7}$$

where $\Lambda' = P'X$, where P' is $s \times n$, X is $n \times p$, and d is a known $s \times 1$ vector. We want to derive the general F test for this hypothesis.

Remark 6.5 *d may be non-zero. In such cases, a set described in (6.7) is not a subspace. We have to translate it back so that we can write the null hypothesis in terms of subspaces.*

Let b be any solution to the equation $\Lambda'\beta = d$ ($p \times 1$ vector). Then,

$$P'Xb = d,$$

and the null hypothesis in (6.7) can be written as

$$H_0 : P'X\beta = P'Xb,$$

or

$$H_0 : P'(X\beta - Xb) = 0. \tag{6.8}$$

From the formulation in (6.8), we can write the reduced model as

$$Y = X\beta + \varepsilon \text{ and } P'(X\beta - Xb) = 0. \tag{6.9}$$

Letting $\beta = \beta - b$, we can rewrite (6.9) as

$$Y - Xb = X\beta^* + \varepsilon \text{ and } P'X\beta^* = 0. \tag{6.10}$$

We can write (6.10) in terms of subspaces.

$$H_0 : \mathrm{E}(Y - Xb) \in C(X) \text{ and } \mathrm{E}(Y - Xb) \perp C(P),$$

or, equivalently,

$$H_0 : \mathrm{E}(Y - Xb) \in C(X) \ \text{ and } \ \mathrm{E}(Y - Xb) \cap C(MP)^{\perp}. \tag{6.11}$$

Thus, for the hypothesis in (6.11), the reduced model can be written as

$$Y - Xb = X_0\gamma_0 + \varepsilon,$$

where

$$X_0 = M - M_{MP}.$$

The F statistic for testing (6.11) can be written as

$$\begin{aligned} F &= \frac{\|M_{MP}(Y - Xb)\|^2 / r(M_{MP})}{\|(I - M)(Y - Xb)\|^2 / r(I - M)} \\ &= \frac{(Y - Xb)' M_{MP}(Y - Xb) / r(M_{MP})}{(Y - Xb)'(I - M)(Y - Xb) / r(I - M)}. \end{aligned}$$

Remark 6.6 *Both the numerator and denominator of the F test do NOT depend on b: denominator $= Y'(I - M)Y$ and numerator $= (\Lambda'\hat{\beta} - d)'(\Lambda'(X'X)^{-}\Lambda)^{-}(\Lambda'\hat{\beta} - d)$ (Exercise!). Thus, the F test is invariant with respect to the choice of b.*

The F test can be written as

$$\begin{aligned} F = &= \frac{\|M_{MP}(Y - Xb)\|^2 / r(M_{MP})}{\|(I - M)Y\|^2 / r(I - M)} \\ &= \frac{(Y - Xb)' M_{MP}(Y - Xb) / r(M_{MP})}{Y'(I - M)Y / r(I - M)} \\ &\sim F(r(M_{MP}), r(I - M), \gamma) \end{aligned}$$

where

$$\begin{aligned}\gamma &= \frac{\|M_{MP}(X\beta - Xb)\|^2}{2\sigma^2} \\ &= \frac{(X\beta - Xb)' M_{MP}(X\beta - Xb)}{2\sigma^2}.\end{aligned}$$

The F test can also be written in terms of Λ and $\hat{\beta}$.

$$\begin{aligned}F &= \frac{(\Lambda'\hat{\beta} - d)'(\Lambda'(X'X)^{-}\Lambda)^{-}(\Lambda'\hat{\beta} - d)}{MSE} \\ &\sim F(r(\Lambda), r(I - M), \gamma)\end{aligned}$$

where

$$MSE = \frac{Y'(I - M)Y}{r(I - M)} \text{ and } \gamma = \frac{(\Lambda'\beta - d)'(\Lambda'(X'X)^{-}\Lambda)^{-}(\Lambda'\beta - d)}{2\sigma^2}.$$

Example

Consider the linear model

$$Y = X\beta + \varepsilon$$

where $Y = (Y_1, Y_2, Y_3)'$, $\beta = (\beta_1, \beta_2)'$,

$$X = \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{pmatrix}$$

and $\varepsilon \sim N_3(0, \sigma^2 I)$. Suppose we wish to test

$$H_0 : \beta_1 + 2\beta_2 = 1.$$

This H_0 can be expressed as

$$H_0 : \rho'(X\beta - Xb) = 0$$

where $\rho' = (1/3, 1/3, 1/3)$, and b is any solution to $\rho' X\beta = d (= 1)$.

Solving this equation for b, we have

$$\rho' Xb = (1/3, 1/3, 1/3) \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = 1.$$

This equation implies that $b_1 + 2b_2 = 3$, so $b_1 = 1$ and $b_2 = 1$ is a solution.

Thus $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Thus we can write the null hypothesis as

$$H_0 : \rho'(X\beta - Xb) = 0$$

where $\rho' = (1/3, 1/3, 1/3)$ and $b = (1, 1)'$.

Note: ρ was chosen so that $\rho \in C(X)$ and thus $M\rho = \rho = (1/3, 1/3, 1/3)'$. Thus,

$$\begin{aligned} M_{MP} &= (M\rho)(\rho' M\rho)^-(\rho' M) \\ &= \rho(\rho'\rho)^-\rho' = \frac{\rho\rho'}{\rho'\rho} \\ &= \frac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \end{aligned}$$

Note: $r(M_{MP} = r(\rho) = 1$ and

$$Xb = \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 3J_3$$

where $J_3 = (1, 1, 1)'$. Thus, the numerator of the F test can be written as

$$\begin{aligned}&\frac{(Y-3J_3)'\rho\rho'(Y-3J_3)/r(\rho)}{\rho'\rho}\\ =&\frac{(\rho'Y-3\rho'J_3)(\rho'Y-3\rho'J_3)/1}{1/9}\\ =&9(\rho'Y-3\rho'J_3)^2=9(\bar{Y}-3)^2\end{aligned}$$

where $\bar{Y}=(\sum_{i=1}^{3}Y_i)/3$.

Note: Since $C(X)=\mathcal{S}\left\{\begin{pmatrix}1\\1\\1\end{pmatrix}\right\}$, we have $M=\frac{1}{3}\begin{pmatrix}1&1&1\\1&1&1\\1&1&1\end{pmatrix}$, and $r(M)=1$. Thus,

$$MSE=\frac{Y'(I-M)Y}{r(I-M)}=\frac{1}{2}\sum_{i=1}^{3}(Y_i-\bar{Y})^2.$$

Thus, the F test can be written as

$$\begin{aligned}F=&\frac{9(\bar{Y}-3)^2}{\sum_{i=1}^{3}(Y_i-\bar{Y})^2/2}\\ \sim&F(1,2,\gamma)\end{aligned}$$

where

$$\gamma=\frac{9(\mathrm{E}(\bar{Y}-3)^2}{2\sigma^2}=\frac{9(\beta_1+2\beta_2-3)^2}{2\sigma^2}.$$

Note:

$$\mathrm{E}(\bar{Y})=\frac{1}{3}\sum_{i=1}^{3}\mathrm{E}(Y_i)=\frac{1}{3}\sum_{i=1}^{3}(\beta_1+2\beta_2)=\beta_1+2\beta_2.$$

6.3 Breaking a sum of squares into independent (orthogonal) components

Goal: We wish to decompose a quadratic form into a sum of independent quadratic forms, where each quadratic form has one degree of freedom.

To motivate the idea of breaking up sums of squares, we consider the two way ANOVA model without interaction. This model is given by

$$Y_{ijk} = \mu + \alpha_i + \eta_j + \varepsilon_{ijk},$$

$i = 1, \ldots, a, j = 1, \ldots, b$, and $k = 1, \ldots, N$. Let $n = abN$.

The ANOVA table is given by

Source	DF	SS	MS
Mean	1	$Y' \left(J_n^n/n\right) Y$	$Y' \left(J_n^n/n\right) Y$
Treatments (α)	$a-1$	$Y'M_\alpha Y$	$Y'M_\alpha Y/(a-1)$
Treatments (η)	$b-1$	$Y'M_\eta Y$	$Y'M_\eta Y/(b-1)$
Error	$n-a-b+1$	$Y'(I-M)Y$	$Y'(I-M)Y/(n-a-b+1)$
Total	n	$Y'Y$	

where $J_n^n = JJ'$. Here,

- M_α: orthogonal projection operator onto $C(X_\alpha)$,

- X_α: design matrix corresponding to the following model

$$Y_{ijk} = \mu + \alpha_i + \varepsilon_{ijk},$$

In the ANOVA setting, it is often an interest to test single degree of freedom contrasts such as

$$\sum_{i=1}^{a} \lambda_i \alpha_i = 0.$$

To test such a contrast, we need to break up the α treatment sums of squares into $a-1$ separate components, each having 1 degree of freedom. That is, the quadratic form $Y'M_\alpha Y$ must be decomposed into

$$Y'M_\alpha Y = \sum_{i=1}^{a-1} Y'M_i Y$$

where each M_i has rank 1 and $M_i M_j = 0$ for $i \neq j$.

In terms of subspaces, $C(M_\alpha)$ is decomposed into a sum of $a-1$ orthogonal subspaces each of dimension 1. Thus,

$$C(M_\alpha) = C(M_1) + C(M_2) + \ldots + C(M_{a-1}).$$

6.4 Decomposing a subspace into a sum of 1 dimensional orthogonal subspaces

Consider the linear model

$$Y = X\beta + \varepsilon$$

where $\varepsilon \sim N_n(0, \sigma^2 I)$. Let $M = X(X'X)^- X'$ denote the orthogonal projection onto $C(X)$. Let M_T be any orthogonal projection operator with the property $C(M_T) \subset C(M)$.

Then M_T defines a statistic

$$\begin{aligned} F =& \frac{\|M_T Y\|^2 / r(M_T)}{\|(I-M)Y\|^2 / r(I-M)} \\ =& \frac{(Y'M_T Y)/r(M_T)}{(Y'(I-M)Y)/r(I-M)} \end{aligned}$$

for testing the reduced model

$$Y = (M - M_T)\gamma_0 + \varepsilon.$$

- $C(M - M_T)$: estimation space under H_0;
- $C(M_T)$: “test” space;
- $C(I - (M - M_T))$: error space.

Suppose $r(M_T) = r$, and we want to decompose $C(M_T)$ into a sum of r orthogonal subspaces

$$C(M_T) = C(M_1) + \ldots + C(M_r)$$

where each M_i is an orthogonal projection operator of rank 1 and $M_i M_j = 0$ for $i \neq j$.

Suppose $R = (R_1, \ldots, R_r)$ is an orthonormal basis for $C(M_T)$, where R_i is an $n \times 1$ vector. Then, we know by a previous theorem that the orthogonal projection operator onto $C(M_T)$ is

$$\begin{aligned} M_T &= RR' \\ &= (R_1, \ldots, R_r) \begin{pmatrix} R_1' \\ \vdots \\ R_r' \end{pmatrix} \\ &= \sum_{i=1}^{r} R_i R_i'. \end{aligned}$$

Let $M_i = R_i R_i'$ then by definition M_i is an orthogonal projection operator and $M_i M_j = 0$ for $i \neq j$. Thus, $Y'M_iY$ and $Y'M_jY$ are independent for each $i \neq j$. Then,

$$Y'M_TY = Y' \left(\sum_{i=1}^{r} M_i \right) Y = \sum_{i=1}^{r} Y'M_iY.$$

Thus, we have

$$\begin{aligned}F =& \frac{\|M_iY\|^2/r(M_T)}{\|(I-M)Y\|^2/r(I-M)}\\ =& \frac{(Y'M_iY)/r(M_T)}{(Y'(I-M)Y)/r(I-M)} \sim F(1, r(I-M), \gamma)\end{aligned}$$

where $\gamma = \|M_iX\beta\|^2/(2\sigma^2)$.

Remark 6.7 *In one-way ANOVA, $Y'M_TY$ corresponds to the treatment sums of squares, while the $Y'M_iY$'s correspond to the sums of squares for a set of orthogonal contrasts.*

Question: What is the correspondence between the hypothesis tested using $Y'M_TY$ and that using the $Y'M_iY$'s?

Since M_T and the M_i's are positive semidefinite, we have

$$0 = \|M_TX\beta\|^2 = \sum_{i=1}^{r} \|M_iX\beta\|^2$$

if and only if $\|M_iX\beta\|^2 = 0$ for all $i = 1, \ldots, r$.

Remark 6.8 *H_0 corresponds to M_T is true if and only if H_0 corresponding to ALL of the M_i's is true.*

Equivalently, if H_0 corresponding to M_T is NOT true, we have

$$0 < \|M_TX\beta\|^2 = \sum_{i=1}^{r} \|M_iX\beta\|^2.$$

This occurs if and only if at least one of the $\|M_iX\beta\|^2 > 0$ ($\because$ M_T and M_i's are positive semidefinite).

Remark 6.9 *H_0 corresponds to M_T is NOT true if and only if AT LEAST ONE of H_0's corresponding to the M_i's is NOT true.*

Remark 6.10 *In terms of one-way ANOVA, these results corresponding to stating that*

1. *The hypothesis of no treatment effect is true if and only if all the contrasts in a set of orthogonal contrasts are 0,*
2. *Or, equivalently, the hypothesis of no treatment effects is NOT true if and only if AT LEAST ONE contrast in a set of orthogonal contrasts is NOT 0.*

6.5 Confidence Regions

Consider the problem of finding a confidence region for an estimable vector $\Lambda'\beta$ where $\Lambda' = P'X$.

Examples:

a) Suppose $\beta_1 - \beta_2$ is estimable and we wish to construct a 95% confidence interval for it. In this case, $\lambda' = (1, -1, 0, \ldots, 0)$.

b) Suppose $\begin{pmatrix} \beta_1 + \beta_3 \\ \beta_2 \end{pmatrix}$ is estimable and we wish to find a 95% joint confidence region for these parameters. What is Λ?

We have

$$\frac{(Y - X\beta)'M_{MP}(Y - X\beta)/r(M_{MP})}{(Y - X\beta)'(I - M)(Y - X\beta)/r(I - M)} \sim F(r(M_{MP}), r(I - M)).$$

Remark 6.11 *The noncentrality parameter is 0 (Why?)*

Moreover,

$$(Y - X\beta)'(I - M)(Y - X\beta) = Y'(I - M)Y$$

so that the denominator equals MSE. Also, the numerator can be written as

$$(Y - X\beta)' M_{MP} (Y - X\beta) = (\Lambda'\hat{\beta} - \Lambda'\beta)'(\Lambda'(X'X)^{-}\Lambda)^{-}(\Lambda'\hat{\beta} - \Lambda'\beta)$$

Thus, a $(1-\alpha) \times 100\%$ confidence region for $\Lambda'\beta$ is

$$\left\{ \beta : \frac{(\Lambda'\hat{\beta} - \Lambda'\beta)'(\Lambda'(X'X)^{-}\Lambda)^{-}(\Lambda'\hat{\beta} - \Lambda'\beta)/r(\Lambda)}{MSE} \leq c_\alpha \right\}$$

where $c_\alpha = F(1-\alpha, r(\Lambda), r(I-M)$ is the upper $(1-\alpha) \times 100\%$ point of a central F distribution with degrees of freedom $(r(\Lambda), r(I-M))$.

Remark 6.12 *The confidence region is an s dimensional ellipsoid, where Λ' is $s \times p$.*

Special case:

Suppose X is of full rank p so that β is estimable. In this case, $\Lambda = I_{p \times p}$. A $(1-\alpha) \times 100\%$ confidence region for β is

$$\left\{ \beta : \frac{(\hat{\beta} - \beta)'(X'X)(\hat{\beta} - \beta)/p}{MSE} \leq F(1-\alpha, p, n-p) \right\}$$

6.6 Hypothesis Tests for Generalized Least Squares

Consider the model

$$Y = X\beta + \varepsilon \tag{6.12}$$

where $\varepsilon \sim N_n(0, \sigma^2 V)$, where V is a known positive definite matrix. The transformed model is given by

$$Q^{-1}Y = Q^{-1}X\beta + Q^{-1}\varepsilon \tag{6.13}$$

where $V = QQ'$ and Q is nonsingular.

Now, consider testing

$$Y = X_0\gamma_0 + \varepsilon \tag{6.14}$$

where $\varepsilon \sim N_n(0, \sigma^2 V)$ and $C(X_0) \subset C(X)$.

Consider the transformed model

$$Q^{-1}Y = Q^{-1}X_0\gamma_0 + Q^{-1}\varepsilon. \tag{6.15}$$

Claim: To test (6.12) against (6.14) by testing (6.13) against (6.15).

To test (6.13) against (6.15), we must show that

$$C(Q^{-1}X_0) \subset C(Q^{-1}X).$$

Theorem 6.3 *If* $C(X_0) \subset C(X)$ *and* Q *is nonsingular, then*

$$C(Q^{-1}X_0) \subset C(Q^{-1}X).$$

Proof: Suppose X_0 is $n \times q$, and $C(X_0) \subset C(X)$. Then there exists a G so that G is $p \times q$ and

$$X_0 = XG.$$

If $\nu \in C(Q^{-1}X_0)$ then $\nu = Q^{-1}X_0 d$ for some d. Substituting for X_0 gives $\nu = Q^{-1}XGd$, so ν is a linear combination of the columns of $Q^{-1}X$ so that $C(Q^{-1}X_0) \subset C(Q^{-1}X)$.

For model (6.13), recall that

$$MSE = \frac{\|Q^{-1}(I-A)Y\|^2}{n-r(X)} = \frac{Y'(I-A)'V^{-1}(I-A)Y}{n-r(X)}.$$

Define

$$A_0 = X_0(X_0'V^{-1}X_0)^- X_0'V^{-1}.$$

Then A_0 is a projection operator onto $C(X_0)$. We now have the following result.

Theorem 6.4 *To test (6.13) against (6.15), the test statistic is*

$$\begin{aligned} F &= \frac{Y'(A-A_0)'V^{-1}(A-A_0)Y/(r(X)-r(X_0)}{MSE} \\ &\sim F(r(X)-r(X_0), n-r(X), \gamma) \end{aligned}$$

where

$$\gamma = \frac{\beta'X'(A-A_0)'V^{-1}(A-A_0)X\beta}{2\sigma^2}.$$

Remark 6.13 $\gamma = 0$ *if and only if* $E(Y) \in C(X_0)$. *That is, if* H_0 *is true.*

For the generalized least squares model, suppose we with to test

$$H_0 : \Delta'\beta = 0$$

where $\Delta'\beta = P'X\beta$ is an estimable vector. The F statistic for this hypothesis is given by

$$F = \frac{\hat{\beta}'\Delta(\Delta'(X'V^{-1}X)^-\Delta)^-\Delta'\hat{\beta}/r(\Delta)}{MSE} \sim F(r(\Delta), n-r(X), \gamma)$$

where

$$\gamma = \frac{\beta'\Delta(\Delta'(X'V^{-1}X)^{-}\Delta)^{-}\Delta'\beta/r(\Delta)}{2\sigma^2}$$

$$MSE = \frac{Y'(I-A)'V^{-1}(I-A)Y}{n-r(X)},$$

and $\Delta\hat{\beta}$ is the unique UMVUE of $\Delta\beta$.

By writing $\Delta'\beta = P'X\beta$, we can rewrite the F test above as

$$\begin{aligned} F &= \frac{Y'A'P(P'X(X'V^{-1}X)^{-}X'P)^{-}P'AY/r(A'P}{MSE} \\ &\sim F(r(A'P), n-r(X), \gamma) \end{aligned}$$

where

$$\gamma = \frac{\beta'X'P(P'X(X'V^{-1}X)^{-}X'P)^{-}P'X\beta}{2\sigma^2}.$$

6.7 Likelihood Ratio Test

The F test for testing nested linear models is equivalent to the likelihood ratio test (LRT). We now give the general definition of LRT.

Definition 6.1 *Let Θ denote the parameter space, and let $\Theta_0 \subset \Theta$. Let θ be a vector in Θ, and let y denote the data. The LRT for testing*

$$\begin{aligned} H_0 :& \theta \in \Theta_0 \\ H_a :& \theta \in \Theta_0^c \end{aligned}$$

is given by

$$\lambda(y) = \frac{\sup_{\Theta_0} L(\theta|y)}{\sup_{\Theta} L(\theta|y)}.$$

The LRT is any test that has a rejection region of the form $\{y : \lambda(y) \leq c\}$, where c is any number satisfying $0 \leq c \leq 1$.

Example:

Consider the linear model

$$Y = X\beta + \varepsilon$$

where $\varepsilon \sim N_n(0, \sigma^2 I)$. Let $C(X_0) \subset C(X)$ and suppose we wish to test

$$\begin{aligned} H_0 &: \mathrm{E}(Y) \in C(X_0) \\ H_a &: \mathrm{E}(Y) \in C(X) \cap C(X_0)^c. \end{aligned}$$

The mode under H_0 is $Y = X_0\gamma_0 + \varepsilon$.

Claim: To derive the LRT for this hypothesis.

The likelihood function under H_0 is given by

$$L(\gamma_0, \sigma|Y) = \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}(Y - X_0\gamma_0)'(Y - X_0\gamma_0)\right\}.$$

To get the numerator of the LRT, we need to maximize this likelihood with respect to (γ_0, σ). From the previous results, we know that the maximizer of γ_0 satisfies

$$X_0\hat{\gamma}_0 = M_0Y$$

where $M_0 = X_0(X_0'X_0)^- X_0'$. We also know from the previous results that

$$\hat{\sigma}_0^2 = \frac{Y'(I - M_0)Y}{n}.$$

To compute the denominator of the LRT, we compute the supremum of the likelihood over the entire parameter space, i.e., under the full model $Y = X\beta + \varepsilon$. Thus, the likelihood function under H_a is

$$L(\beta, \sigma|Y) = \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\}.$$

for which $X\hat{\beta} = MY$ and

$$\hat{\sigma}^2 = \frac{Y'(I - M)Y}{n}.$$

Thus,

$$
\begin{aligned}
\lambda(y) &= \frac{\sup_{\gamma_0,\sigma} L(\gamma_0,\sigma|y)}{\sup_{\beta,\sigma} L(\beta,\sigma|y)} \\
&= \frac{(\hat{\sigma}_0^2)^{-n/2}\exp\left\{-\frac{1}{2(\hat{\sigma}_0^2)}(Y - X_0\hat{\gamma}_0)'(Y - X_0\hat{\gamma}_0)\right\}}{(\hat{\sigma}^2)^{-n/2}\exp\left\{-\frac{1}{2(\hat{\sigma}^2)}(Y - X_0\hat{\beta}_0)'(Y - X_0\hat{\beta}_0)\right\}} \\
&= \left(\frac{Y'(I-M)Y}{Y'(I-M_0)Y}\right)^2 \frac{\exp\left\{-\frac{1}{2(\hat{\sigma}_0^2)}(Y - X_0\hat{\gamma}_0)'(Y - X_0\hat{\gamma}_0)\right\}}{\exp\left\{-\frac{1}{2(\hat{\sigma}^2)}(Y - X_0\hat{\beta}_0)'(Y - X_0\hat{\beta}_0)\right\}} \\
&= \left(\frac{Y'(I-M)Y}{Y'(I-M_0)Y}\right)^{n/2} \\
&\times \frac{\exp\left\{-\frac{n}{2(Y'(I-M_0)Y)}(Y - X_0\hat{\gamma}_0)'(Y - X_0\hat{\gamma}_0)\right\}}{\exp\left\{-\frac{n}{2(Y'(I-M)Y)}(Y - X_0\hat{\beta}_0)'(Y - X_0\hat{\beta}_0)\right\}} \\
&= \left(\frac{Y'(I-M)Y}{Y'(I-M_0)Y}\right)^{n/2} \frac{\exp(-n/2)}{\exp(-n/2)} \\
&= \left(\frac{Y'(I-M)Y}{Y'(I-M_0)Y}\right)^{n/2}.
\end{aligned}
$$

Thus, we reject H_0 if

$$
\begin{aligned}
\left(\frac{Y'(I-M)Y}{Y'(I-M_0)Y}\right)^{n/2} &\leq c \\
\Longleftrightarrow \quad \frac{Y'(I-M)Y}{Y'(I-M_0)Y} &\leq c_1
\end{aligned}
\tag{6.16}
$$

Write $I - M_0 = (I - M) + (M - M_0)$ so that

$$
Y'(I-M_0)Y = Y'(I-M)Y + Y'(M-M_0)Y.
$$

Thus (6.16) becomes

$$\begin{aligned} \frac{Y'(I-M)Y}{Y'(I-M)Y+Y'(M-M_0)Y} &\leq c_1 \\ \iff \frac{Y'(I-M)Y+Y'(M-M_0)Y}{Y'(I-M)Y} &\geq c_2 \end{aligned} \tag{6.17}$$

where $c_2 = c_1^{-1}$. Now, (6.17) can be written as

$$1 + \frac{Y'(M-M_0)Y}{Y'(I-M)Y} \geq c_2$$

$$\iff \frac{Y'(M-M_0)Y}{Y'(I-M)Y} \geq c_3$$

where $c_3 = c_2 - 1$. Finally, this expression is equivalent to

$$\frac{Y'(M-M_0)Y/r(M-M_0)}{Y'(I-M)Y/r(I-M)} \geq c_4 \tag{6.18}$$

The LRT in (6.18) is equivalent to the F test.