Linear Models and Experimental Designs

STATS 752

Winter 2023

Cameron Roopnarine[*] Shui Feng[†]

27th February 2023

Contents

1 **Lecture 1: Review of Linear Algebra** 2

2 **Lecture 2: Quadratic Forms and Distributions** 7

3 **Lecture 3: Some Basic Lemmas** 12

4 **Lecture 4: Quadratic Forms with Idempotency** 17

5 **Lecture 5: Criteria for Independence** 21

6 **Lecture 6: Cochran's Theorem** 24

7 **Lecture 7: Full Rank Regression** 27

8 **Lecture 8: Test of Overall Regression** 31

9 **Lecture 9: Lack of Fit** 36

10 **Lecture 10: Determination of Predictors and Generalized Inverse** 38

11 **Lecture 11: g-inverse** 41

12 **Lecture 12: Regression Without Full Rank** 45

13 **Lecture 13: Non-full Rank Regression** 48

---

[*] LaTeXer

[†] Instructor

1 Lecture 1: Review of Linear Algebra

DEFINITION 1.1: Vectors in $\mathbb{R}^n$

For any positive integer n, we define n-dimensional Euclidean space $\mathbb{R}^n$ by

$$\mathbb{R}^n = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \middle| \, x_1, \ldots, x_n \in \mathbb{R} \right\}.$$

If $\vec{x} \in \mathbb{R}^n$, then there exists $x_1, \ldots, x_n \in \mathbb{R}$ such that

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

DEFINITION 1.2: Matrix

An $n \times m$ **matrix $\mathbf{A}$** is a rectangular array with n rows and m columns. We denote the entry in the i^{th} row and j^{th} column by a_{ij} or $(\mathbf{A})_{ij}$. That is,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{im} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj} & \cdots & a_{nm} \end{pmatrix}.$$

The set of all $n \times m$ matrices with real entries is denoted by $\mathbb{R}^{n \times m}$.

DEFINITION 1.3: Transpose

Let $\mathbf{A} \in \mathbb{R}^{n \times m}$. We define the **transpose** of $\mathbf{A}$, denoted $\mathbf{A}'$, to be the $m \times n$ matrix whose ij^{th} entry is the ji^{th} entry of $\mathbf{A}$. That is,

$$(\mathbf{A}')_{ij} = (\mathbf{A})_{ji}.$$

DEFINITION 1.4: Square Matrix

An $n \times n$ matrix is called a **square matrix**.

DEFINITION 1.5: Symmetric Matrix

A matrix is called **symmetric** if $\mathbf{A}' = \mathbf{A}$.

DEFINITION 1.6: Diagonal Matrix

An $n \times n$ matrix $\mathbf{D}$ is said to be a **diagonal matrix** if $d_{ij} = 0$ for all $i \neq j$. We denote a diagonal matrix by

$$\mathbf{D} = \operatorname{diag}(d_1, d_2, \ldots, d_n).$$

DEFINITION 1.7: Identity Matrix

The $n \times n$ matrix $\mathbf{I}$ (or $\mathbf{I}_n$) such that $(\mathbf{I})_{ii}$ for $1 \leq i \leq n$, and $(\mathbf{I})_{ij} = 0$ whenever $i \neq j$ is called the **identity matrix**.

DEFINITION 1.8: Upper Triangular, Lower Triangular

An $n \times m$ matrix $\mathbf{U}$ is said to be **upper triangular** if $u_{ij} = 0$ whenever $i > j$. An $n \times m$ matrix $\mathbf{L}$ is said to be **lower triangular** if $\ell_{ij} = 0$ whenever $i < j$.

- Upper triangular:

$$\mathbf{U} = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1m} \\ 0 & u_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & u_{(n-1)m} \\ 0 & \cdots & 0 & u_{nm} \end{pmatrix}.$$

- Lower triangular:

$$\mathbf{L} = \begin{pmatrix} \ell_{11} & 0 & \cdots & 0 \\ \ell_{21} & \ell_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \ell_{n1} & \cdots & \ell_{n(m-1)} & \ell_{nm} \end{pmatrix}.$$

DEFINITION 1.9: Vector/Matrix of 1's and 0's

Vector of 1's:

$$\vec{j} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n.$$

Matrix of 1's:

$$\mathbf{J} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Zero vector:

$$\vec{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^n.$$

Zero Matrix:

$$\mathbf{O} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

DEFINITION 1.10: Matrix Mapping

If $\mathbf{A}$ is an $n \times m$ matrix, then we can define a function $T\colon \mathbb{R}^m \to \mathbb{R}^n$ by $T(\vec{x}) = \mathbf{A}\vec{x}$ called a **matrix mapping**. For this mapping, we define:

- **Kernel**

$$\operatorname{Ker}(T) = \{\vec{x} \in \mathbb{R}^m \mid \mathbf{A}\vec{x} = \vec{0}\}.$$

- **Image**

$$\text{Image}(T) = \{\mathbf{A}\vec{x} \in \mathbb{R}^n \mid \vec{x} \in \mathbb{R}^m\}.$$

- **Rank**

$$\text{rank}(T) = \dim(\text{Image}(T)).$$

- **Nullity**

$$\text{nullity}(T) = \dim(\text{Ker}(T)).$$

REMARK — Rank-Nullity Theorem

$$\text{rank}(T) + \text{nullity}(T) = m.$$

REMARK

Note that

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix} = \begin{pmatrix} \vec{a}_1 \\ \vdots \\ \vec{a}_n \end{pmatrix} = \begin{pmatrix} \vec{a}^1 & \cdots & \vec{a}^m \end{pmatrix}.$$

Clearly, the image of T is the space generated by $\vec{a}^1, \ldots, \vec{a}^m$. Therefore,

$$\text{rank}(T) = \text{column rank of } \mathbf{A}.$$

THEOREM 1.1

$\mathbf{A}$ and $\mathbf{A}'$ have the same column rank.

Proof: Let $\mathbf{A} \to T$, $\mathbf{A}'\mathbf{A} \to \tilde{T}$, and $\mathbf{A}' \to \hat{T}$.

(1) Let $\vec{x} \in \text{Ker}(T)$, so we have $\mathbf{A}\vec{x} = \vec{0} \implies \mathbf{A}'\mathbf{A}\vec{x} = \vec{0}$. Hence, $\vec{x} \in \text{Ker}(\tilde{T})$. So, $\text{Ker}(T) \subset \text{Ker}(\tilde{T})$.

(2) $\mathbf{A}'\mathbf{A}\vec{x} = \vec{0} \implies \vec{x}'\mathbf{A}'\mathbf{A}\vec{x} = 0 \implies (\mathbf{A}\vec{x})'\mathbf{A}\vec{x} = 0 \implies \mathbf{A}\vec{x} = 0$. So, $\text{Ker}(\tilde{T}) \subset \text{Ker}(T)$.

Therefore, $\text{rank}(T) = \text{rank}(\tilde{T})$. By Rank-Nullity theorem, $\mathbf{A}$ and $\mathbf{A}'$ have the same column rank, noting that $\text{Image}(\tilde{T}) \subset \text{Image}(\hat{T})$. Hence, the column rank of $\mathbf{A}$ is less than or equal to the column rank of $\mathbf{A}'$. By symmetry, the column rank of $\mathbf{A}'$ is less than or equal to the column rank of $(\mathbf{A}')' = \mathbf{A}$. Therefore, $\mathbf{A}$ and $\mathbf{A}'$ have the same column rank.

REMARK

The column rank of $\mathbf{A}'$ is the row rank of $\mathbf{A}$. Hence,

$$\begin{aligned} \text{rank}(\mathbf{A}) &= \text{maximum number of linearly independent rows of } \mathbf{A} \\ &= \text{maximum number of linearly independent columns of } \mathbf{A}. \end{aligned}$$

DEFINITION 1.11: Full Rank

Let $\mathbf{A} \in \mathbb{R}^{n \times m}$. We say $\mathbf{A}$ has **full rank** if $\text{rank}(\mathbf{A}) = \min\{n, m\}$

THEOREM 1.2

Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ and $\mathbf{B} \in \mathbb{R}^{m \times p}$.

(1) $(\mathbf{A}\mathbf{B})' = \mathbf{B}'\mathbf{A}' \in \mathbb{R}^{p \times n}$.

(2) $\vec{j}'\vec{j} = n$ *and* $\vec{j}\vec{j}' = \mathbf{J}$.

(3) $\mathbf{J}\mathbf{J} = n\mathbf{J}$.

THEOREM 1.3

(1) $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}')$.

(2) $\operatorname{rank}(\mathbf{A}'\mathbf{A}) = \operatorname{rank}(\mathbf{A}\mathbf{A}') = \operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}')$.

(3) $\operatorname{rank}(\mathbf{A}\mathbf{B}) \leq \min\{\operatorname{rank}(\mathbf{A}), \operatorname{rank}(\mathbf{B})\}$.

Proof: We have already shown (1) and (2). For (3), we have

$$\operatorname{rank}(\mathbf{A}\mathbf{B}) \leq \operatorname{rank}(\mathbf{A}).$$

On the other hand,

$$\operatorname{rank}(\mathbf{A}\mathbf{B}) = \operatorname{rank}(\mathbf{B}'\mathbf{A}') \leq \operatorname{rank}(\mathbf{B}') = \operatorname{rank}(\mathbf{B}).$$

REMARK — Invertible Matrix Theorem

$\mathbf{A} \in \mathbb{R}^{n\times n}$ is invertible (non-singular) if and only if $\operatorname{rank}(\mathbf{A}) = n$, and we denote the inverse of $\mathbf{A}$ by $\mathbf{A}^{-1}$.

REMARK — Properties of Invertible Matrices

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

(2) $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$.

(3) $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.

(4) $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

DEFINITION 1.12: Positive Definite, Positive Semidefinite

- $\mathbf{A}$ is **positive definite** when $\vec{x}'\mathbf{A}\vec{x} > 0 \iff \vec{x} \neq \vec{0}$.
- $\mathbf{A}$ is **positive semidefinite** when $\vec{x}'\mathbf{A}\vec{x} \geq 0$ for all $\vec{x}$ and there exists $\vec{x} \neq \vec{0}$ such that $\vec{x}'\mathbf{A}\vec{x} = 0$.

DEFINITION 1.13: Orthogonal

$\mathbf{A} \in \mathbb{R}^{n\times n}$ is **orthogonal** if $\mathbf{A}' = \mathbf{A}^{-1}$.

DEFINITION 1.14: Eigenvalue, Eigenvector, Spectrum

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$. If there exists a vector $\vec{x} \neq \vec{0}$ such that $\mathbf{A}\vec{x} = \lambda\vec{x}$, then λ is called an **eigenvalue** of $\mathbf{A}$ and $\vec{v}$ is called an **eigenvector** of $\mathbf{A}$ corresponding to λ. The set of all eigenvalues of $\mathbf{A}$ is called the **spectrum** for $\mathbf{A}$.

EXAMPLE 1.1

$$\underbrace{\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} 1 \\ 1 \end{pmatrix}}_{\vec{v}} = \underbrace{3}_{\lambda} \underbrace{\begin{pmatrix} 1 \\ 1 \end{pmatrix}}_{\vec{v}}.$$

THEOREM 1.4: Spectral Decomposition

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ *with eigenvalues* $\lambda_1, \ldots, \lambda_n$. $\mathbf{A}$ *is symmetric if and only if*

$$\mathbf{A} = \mathbf{Q}' \operatorname{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{Q},$$

where $\mathbf{Q}$ *is an orthogonal matrix, that is,* $\mathbf{Q}\mathbf{Q}' = \mathbf{I}$.

THEOREM 1.5

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ *be a symmetric matrix.*

(i) $\mathbf{A}$ *is positive definite (semidefinite) if and only if all eigenvalues are positive (non-negative).*

(ii) $\mathbf{A}$ *is positive definite if and only if there exists a unique lower triangular matrix* $\mathbf{L}$ *with positive diagonal elements such that* $\mathbf{A} = \mathbf{L}\mathbf{L}'$ *(Cholesky decomposition).*

DEFINITION 1.15: Idempotent, Trace

- $\mathbf{A} \in \mathbb{R}^{n\times n}$ is **idempotent** if $\mathbf{A} = \mathbf{A}^2$.
- Let $\operatorname{tr}\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ be defined by

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$$

(called the **trace** of a matrix).

THEOREM 1.6

Let $\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathbf{B} \in \mathbb{R}^{m\times p}$, $\mathbf{C} \in \mathbb{R}^{p\times n}$, *and* $a, b \in \mathbb{R}$.

(1) tr *is linear:* $\operatorname{tr}(a\mathbf{A} + b\mathbf{B}) = a\operatorname{tr}(\mathbf{A}) + b\operatorname{tr}(\mathbf{B})$.

(2) Cyclic property: $\operatorname{tr}(\mathbf{ABC}) = \operatorname{tr}(\mathbf{CAB}) = \operatorname{tr}(\mathbf{BCA})$.

Proof:

(1) By definition.

(2) Note that

$$\begin{aligned}
\operatorname{tr}(\mathbf{ABC}) &= \sum_{i=1}^{n}\sum_{k=1}^{m}\sum_{j=1}^{p} a_{ik}b_{kj}c_{ji} \\
&= \sum_{j=1}^{p}\sum_{i=1}^{n}\sum_{k=1}^{m} c_{ji}a_{ik}b_{kj} \\
&= \operatorname{tr}(\mathbf{CAB}) \\
&= \sum_{k=1}^{m}\sum_{j=1}^{p}\sum_{i=1}^{n} b_{kj}c_{ji}a_{ik} \\
&= \operatorname{tr}(\mathbf{BCA}).
\end{aligned}$$

REMARK — Properties of Idempotent Matrices

(1) Eigenvalues of idempotent matrices are 1 or 0 since

$$\begin{aligned} & \mathbf{A}\vec{x} = \lambda\vec{x} \\ \implies & \mathbf{A}\mathbf{A}\vec{x} = \lambda\mathbf{A}\vec{x} = \lambda^2\vec{x} \\ \implies & \lambda\vec{x} = \lambda^2\vec{x} \\ \implies & \lambda = \lambda^2 \\ \implies & \lambda = 0 \text{ or } 1. \end{aligned}$$

(2) Idempotent matrices are diagonalizable, that is, there exists an invertible matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{P}^{-1}\mathbf{D}\mathbf{P}$, where

$$\mathbf{D} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n),\ \forall\lambda_i = 0 \text{ or } 1.$$

(3) If $\mathbf{A}$ is idempotent, then $\operatorname{tr}(\mathbf{A}) = \operatorname{rank}(\mathbf{A})$.

$$\begin{aligned} \operatorname{tr}(\mathbf{A}) &= \operatorname{tr}(\mathbf{P}^{-1}\mathbf{D}\mathbf{P}) \\ &= \operatorname{tr}(\mathbf{D}\mathbf{P}\mathbf{P}^{-1}) \\ &= \operatorname{tr}(\mathbf{D}) \\ &= \lambda_1 + \cdots + \lambda_n \\ &= \#\text{ of non-zero } \lambda_i\text{'s} \\ &= \operatorname{rank}(\mathbf{A}). \end{aligned}$$

LECTURE 2
12th January

2 Lecture 2: Quadratic Forms and Distributions

DEFINITION 2.1: Quadratic Form

The **quadratic form** associated $\mathbf{A} \in \mathbb{R}^{n\times n}$ is defined as

$$\vec{x}'\mathbf{A}\vec{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j.$$

If $\tilde{\mathbf{A}} = \dfrac{\mathbf{A}+\mathbf{A}'}{2}$, then note that $\tilde{\mathbf{A}}$ is symmetric and

$$\begin{aligned} \vec{x}'\tilde{\mathbf{A}}\vec{x} &= \sum_{i=1}^{n}\sum_{j=1}^{n} x_j \frac{a_{ij}+a_{ji}}{2} x_j \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j \\ &= \vec{x}'\mathbf{A}\vec{x}. \end{aligned}$$

Therefore, there is a one-to-one correspondence between quadratic forms and symmetric matrices.

EXAMPLE 2.1

Let S^2 denote the sample variance of a random sample $X_1, \ldots, X_n$. Set

$$\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \qquad \vec{\mu} = \mathbb{E}[\vec{X}] = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}.$$

Write $(n-1)S^2$ as a quadratic form and identify the matrix $\mathbf{A}$.

Solution:

$$\begin{aligned}
(n-1)S^2 &= \sum_{i=1}^{n}(X_i - \bar{X})^2 \\
&= \sum_{i=1}^{n} X_i^2 - 2\left(\sum_{i=1}^{n} X_i\right)\bar{X} + n\bar{X}^2 \\
&= \sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2 \\
&= \vec{X}'\vec{X} - \tfrac{1}{n}(\vec{j}'\vec{X})^2 \\
&= \vec{X}'\vec{X} - \tfrac{1}{n}\vec{X}'\vec{j}\vec{j}'\vec{X} \\
&= \vec{X}'(\mathbf{I} - \tfrac{1}{n}\vec{j}\vec{j}')\vec{X} \\
&= \vec{X}'(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\vec{X}.
\end{aligned}$$

Hence, $\mathbf{A} = \mathbf{I} - \frac{1}{n}\mathbf{J}$ is symmetric.

THEOREM 2.1

If $\vec{X}$ is a random vector with mean $\vec{\mu}$, covariance matrix $\mathbf{\Sigma}$, and $\mathbf{A}$ is a symmetric matrix of constants, then

$$\mathbb{E}[\vec{X}'\mathbf{A}\vec{X}] = \mathrm{tr}(\mathbf{A\Sigma}) + \vec{\mu}'\mathbf{A}\vec{\mu}.$$

Proof:

$$\begin{aligned}
\mathbb{E}[\vec{X}'\mathbf{A}\vec{X}] &= \mathbb{E}[\mathrm{tr}(\vec{X}'\mathbf{A}\vec{X})] \\
&= \mathbb{E}[\mathrm{tr}(\mathbf{A}\vec{X}\vec{X}')] \\
&= \mathrm{tr}(\mathbb{E}[\mathbf{A}\vec{X}\vec{X}']) \\
&= \mathrm{tr}(\mathbf{A}\,\mathbb{E}[\vec{X}\vec{X}']) \\
&= \mathrm{tr}\Big(\mathbf{A}\,\mathbb{E}\big[(\vec{X} - \vec{\mu} + \vec{\mu})(\vec{X}' - \vec{\mu}' + \vec{\mu}')\big]\Big) \\
&= \mathrm{tr}\big(\mathbf{A}[\mathbf{\Sigma} + \vec{\mu}\vec{\mu}']\big) \\
&= \mathrm{tr}(\mathbf{A\Sigma}) + \mathrm{tr}(\mathbf{A}\vec{\mu}\vec{\mu}') \\
&= \mathrm{tr}(\mathbf{A\Sigma}) + \mathrm{tr}(\vec{\mu}'\mathbf{A}\vec{\mu}) \\
&= \mathrm{tr}(\mathbf{A\Sigma}) + \vec{\mu}'\mathbf{A}\vec{\mu}.
\end{aligned}$$

EXAMPLE 2.2

Assume that $\mu_1 = \cdots = \mu_n = \mu$ and $\mathbf{\Sigma} = \sigma^2\mathbf{I}$. Find $\mathbb{E}[S^2]$.

Solution:

$$\begin{aligned}
\mathbb{E}[S^2] &= \tfrac{1}{n-1}\,\mathbb{E}\big[(n-1)S^2\big] \\
&= \tfrac{1}{n-1}\,\mathbb{E}\Big[\vec{X}'(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\vec{X}\Big] \\
&= \tfrac{1}{n-1}\,\mathbb{E}\big[\mathrm{tr}\big((\mathbf{I} - \tfrac{1}{n}\mathbf{J})\mathbf{\Sigma}\big) + \vec{\mu}'(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\vec{\mu}\big] \\
&= \tfrac{1}{n-1}\,\mathbb{E}\Big[\mathrm{tr}\big((\mathbf{I} - \tfrac{1}{n}\mathbf{J})\sigma^2\mathbf{I}\big) + \mu^2\vec{j}'(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\vec{j}\Big] \\
&= \tfrac{1}{n-1}\,\mathbb{E}\Big[\sigma^2(n-1) + \mu^2(\vec{j}'\vec{j} - \tfrac{1}{n}\vec{j}'\mathbf{J}\vec{j})\Big] \\
&= \tfrac{1}{n-1}\,\mathbb{E}[\sigma^2(n-1) + 0] && \text{since } \vec{j}'\mathbf{J}\vec{j} = n\vec{j}'\vec{j} \\
&= \sigma^2.
\end{aligned}$$

REMARK — Multivariate Normal Distribution

Let $\vec{X} = (X_1, X_2, \ldots, X_n)$ be a $1 \times n$ random vector with $\mathbb{E}[X_i] = \mu_i$ and $\mathrm{Cov}(X_i, X_j) = \sigma_{ij}$, for $i, j = 1, 2, \ldots, n$. Let $\vec{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ be the mean vector and $\mathbf{\Sigma}$ be the $n \times n$ symmetric covariance matrix whose (i, j) entry is σ_{ij}. Suppose that also the inverse matrix of $\mathbf{\Sigma}$, $\mathbf{\Sigma}^{-1}$ exists. If the joint probability density function of $(X_1, \ldots, X_n)$ is given by

$$f(\vec{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu})'\mathbf{\Sigma}^{-1}(\vec{x} - \vec{\mu})\right\} \text{ for } \vec{x} \in \mathbb{R}^n$$

where $\vec{x} = (x_1, x_2, \ldots, x_n)$, then $\vec{X}$ is said to have a **multivariate normal distribution**. We write $\vec{X} \sim \mathrm{MN}(\vec{\mu}, \mathbf{\Sigma})$.

REMARK — Aitken's Integral

For any positive definite matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, we have

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\vec{x}'\mathbf{A}\vec{x}\right\} \mathrm{d}\vec{x} = (2\pi)^{n/2}|A|^{-1/2}.$$

THEOREM 2.2

If $\vec{X} \sim MN(\vec{\mu}, \mathbf{\Sigma})$, its moment generating function is given by

$$M_{\vec{X}}(\vec{t}) = \mathbb{E}[e^{\vec{t}'\vec{X}}] = \exp\left\{\vec{t}'\vec{\mu} + \frac{\vec{t}'\mathbf{\Sigma}\vec{t}}{2}\right\}.$$

Proof:

$$\begin{aligned}
M_{\vec{X}}(\vec{t}) &= (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left\{\vec{t}'\vec{x}-\frac{1}{2}(\vec{x}-\vec{\mu})'\mathbf{\Sigma}^{-1}(\vec{x}-\vec{\mu})\right\}\mathrm{d}\vec{x}\\
&= (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\left[(\vec{x}-\vec{\mu})'\mathbf{\Sigma}^{-1}(\vec{x}-\vec{\mu})-2\vec{t}'\vec{x}\right]\right\}\mathrm{d}\vec{x}\\
&= (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\\
&\exp\left\{-\frac{1}{2}\left[(\vec{x}'-\vec{\mu}'-\vec{t}'\mathbf{\Sigma}'+\vec{t}'\mathbf{\Sigma}')\mathbf{\Sigma}^{-1}(\vec{x}-\vec{\mu}-\mathbf{\Sigma}\vec{t}+\mathbf{\Sigma}\vec{t})-2\vec{t}'\vec{x}\right]\right\}\mathrm{d}\vec{x}\\
&= (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\\
&\exp\left\{-\frac{1}{2}(\vec{x}'-\vec{\mu}'-\vec{t}'\mathbf{\Sigma}'+\vec{t}'\mathbf{\Sigma}')\mathbf{\Sigma}^{-1}(\vec{x}-\vec{\mu}-\mathbf{\Sigma}\vec{t})\right\}\\
&\exp\left\{-\frac{1}{2}[\vec{x}'\vec{t}-\vec{\mu}'\vec{t}-\vec{t}'\mathbf{\Sigma}'\vec{t}+\vec{t}'\vec{x}-\vec{t}'\vec{\mu}-\vec{t}'\mathbf{\Sigma}\vec{t}+\vec{t}'\mathbf{\Sigma}'\vec{t}-2\vec{t}'\vec{x}]\right\}\mathrm{d}\vec{x}\\
&= \underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu}-\mathbf{\Sigma}\vec{t})'\mathbf{\Sigma}^{-1}(\vec{x}-\vec{\mu}-\mathbf{\Sigma}\vec{t})\right\}\mathrm{d}\vec{x}}_{=1}\\
&\exp\left\{\vec{t}'\vec{\mu}+\frac{\vec{t}'\mathbf{\Sigma}\vec{t}}{2}\right\}\\
&= \exp\left\{\vec{t}'\vec{\mu}+\frac{\vec{t}'\mathbf{\Sigma}\vec{t}}{2}\right\}
\end{aligned}$$

REMARK — Gamma Distribution

Y is said to have a **Gamma distribution** with shape α and scale β when

$$f_Y(y) = \frac{y^{\alpha-1}e^{-y/\beta}}{\Gamma(\alpha)\beta^\alpha},\ \text{for } y>0,\ \alpha>0,\ \beta>0,$$

and 0 otherwise. We write $Y \sim \text{GAM}(\alpha,\beta)$.

(1) $\mathbb{E}[Y] = \alpha\beta$, $\text{Var}(Y) = \alpha\beta^2$.

(2) $M_Y(t) = (1-\beta t)^{-\alpha}$ for $t < 1/\beta$.

REMARK — Chi-Squared Distribution

Q is said to have a **Chi-squared distribution** with $n \in \mathbb{Z}^+$ degrees of freedom when $Q \sim \text{GAM}(n/2, 2)$. We write $Q \sim \chi^2(n)$.

(1) $\mathbb{E}[Q] = k$, $\text{Var}(Q) = 2k$.

(2) $M_Q(t) = (1-2t)^{-n/2}$ for $t < 1/2$.

THEOREM 2.3

Let $\vec{X} \sim MN(\vec{0}, \mathbf{\Sigma})$. Then $\vec{X}'\mathbf{\Sigma}^{-1}\vec{X} \sim \chi^2(n)$.

Proof: Let $Y = \vec{X}'\mathbf{\Sigma}^{-1}\vec{X}$. Then,

$$\begin{aligned}
M_Y(t) &= \mathbb{E}[e^{tY}] \\
&= \mathbb{E}[e^{\vec{X}'(t\mathbf{\Sigma}^{-1})\vec{X}}] \\
&= (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{\vec{x}'(t\mathbf{\Sigma}^{-1})\vec{x} - \frac{1}{2}\vec{x}'\mathbf{\Sigma}^{-1}\vec{x}\right\} \mathrm{d}\vec{x} \\
&= (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\vec{x}'(1-2t)\mathbf{\Sigma}^{-1}\vec{x}\right\} \mathrm{d}\vec{x} \\
&= |\mathbf{\Sigma}|^{-1/2}\left|\left((1-2t)\mathbf{\Sigma}^{-1}\right)^{-1}\right|^{1/2} \\
&= (1-2t)^{-n/2}
\end{aligned}$$

REMARK — Non-Central Chi-Squared Distribution

Let $X_1, \ldots, X_n$ be independent and $X_i \sim \mathcal{N}(\mu_i, 1)$. Set $\lambda = \frac{1}{2}\vec{\mu}'\vec{\mu} = \frac{1}{2}\sum_{i=1}^{n}\mu_i^2$ and $W = \sum_{i=1}^{n} X_i^2$. Then, W has a **non-central chi-squared distribution** with degrees of freedom n and non-centrality parameter λ. We write $W \sim \chi^2(n, \lambda)$. The usual chi-square corresponds to $\lambda = 0$.
Note: The factor $\frac{1}{2}$ is used for this course.

Not covered in notes:

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

REMARK — F-Distribution

If $X \sim \chi^2(n)$ independently of $Y \sim \chi^2(m)$ for $n, m > 0$, then we say $U = \frac{X/n}{Y/m}$ has a (central) F-**distribution**. We write $U \sim F(n, m)$.

If $X \sim \chi^2(n, \lambda)$ independently of $Y \sim \chi^2(m)$ for $n, m, \lambda > 0$, then we say $U = \frac{X/n}{Y/m}$ has a **non-central** F**-distribution**. We write $U \sim F(n, m, \lambda)$. If $\lambda = 0$, then $U \sim F(n, m)$.

Transformation of Multivariate Normal

If $\mathbf{\Sigma}$ is symmetric, then by the spectral theorem, there exists an orthogonal matrix $\mathbf{\Gamma}$ such that

$$\mathbf{\Sigma} = \mathbf{\Gamma}' \operatorname{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{\Gamma},$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\mathbf{\Sigma}$. Note that $\lambda_i > 0$ for all $i \in [1, n]$ since $\mathbf{\Sigma}$ is positive definite. Furthermore, if we set

$$\mathbf{\Sigma}^{1/2} = \mathbf{\Gamma}' \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})\mathbf{\Gamma},$$

we see that

$$\begin{aligned}
\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2} &= \mathbf{\Gamma}' \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}) \underbrace{\mathbf{\Gamma}\mathbf{\Gamma}'}_{\mathbf{I}} \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})\mathbf{\Gamma} \\
&= \mathbf{\Gamma}' \operatorname{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{\Gamma} \\
&= \mathbf{\Sigma}.
\end{aligned}$$

Therefore, $\mathbf{\Sigma}^{1/2}$ is well-defined and is called the **square root** of $\mathbf{\Sigma}$.

REMARK

If $\vec{X} \sim \text{MN}(\vec{\mu}, \mathbf{\Sigma})$, and $\vec{Z} = \mathbf{\Sigma}^{-1/2}(\vec{X} - \vec{\mu})$, where $\mathbf{\Sigma}^{-1/2} = (\mathbf{\Sigma}^{1/2})^{-1}$, then

$$\vec{Z} \sim \text{MN}(\vec{0}, \mathbf{I}).$$

Proof:

$$\begin{aligned}
M_{\vec{Z}}(\vec{t}) &= \mathbb{E}[e^{\vec{t}'\vec{Z}}] \\
&= \mathbb{E}\Big[\exp\{\vec{t}'\mathbf{\Sigma}^{-1/2}(\vec{X} - \vec{\mu})\}\Big] \\
&= \exp\{-\vec{t}'\mathbf{\Sigma}^{-1/2}\vec{\mu}\}\,\mathbb{E}\Big[\exp\{\vec{t}'\mathbf{\Sigma}^{-1/2}\vec{X}\}\Big] \\
&= \exp\{-\vec{t}'\mathbf{\Sigma}^{-1/2}\vec{\mu}\}\exp\bigg\{\vec{t}'\mathbf{\Sigma}^{-1/2}\vec{\mu} + \frac{\vec{t}'\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}\mathbf{\Sigma}^{-1/2}\vec{t}}{2}\bigg\} \\
&= \exp\bigg\{\frac{\vec{t}'\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{-1/2}\vec{t}}{2}\bigg\} \\
&= \exp\bigg\{\frac{\vec{t}'\vec{t}}{2}\bigg\} \\
&= \exp\bigg\{\frac{1}{2}\sum_{i=1}^{n} t_i^2\bigg\}.
\end{aligned}$$

Lecture 3
16th January

3 Lecture 3: Some Basic Lemmas

LEMMA 3.1

Let $\vec{b}$ be a vector and $\mathbf{W}$ be a positive definite symmetric matrix. Then,

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\bigg\{-\frac{1}{2}\vec{x}'\mathbf{W}^{-1}\vec{x} + \vec{b}'\vec{x}\bigg\}\,\mathrm{d}\vec{x} = (2\pi)^{n/2}|\mathbf{W}|^{1/2}\exp\bigg\{\frac{\vec{b}'\mathbf{W}\vec{b}}{2}\bigg\}.$$

LEMMA 3.2

Let $\mathbf{A}$ be a symmetric matrix and $\vec{X} \sim MN(\vec{\mu}, \mathbf{\Sigma})$. Then,

$$M_{\vec{X}'\mathbf{A}\vec{X}}(t) = |\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma}|^{-1/2}\exp\bigg\{-\frac{1}{2}\vec{\mu}'\big[\mathbf{I} - (\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma})^{-1}\big]\mathbf{\Sigma}^{-1}\vec{\mu}\bigg\}$$

for small enough t.

Proof: By definition,

$$\begin{aligned}
&M_{\vec{X}'\mathbf{A}\vec{X}}(t)\\
&= \mathbb{E}\big[\exp\{t\vec{x}'\mathbf{A}\vec{x}\}\big]\\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}}\exp\{t\vec{x}'\mathbf{A}\vec{x}\}\exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})'\mathbf{\Sigma}^{-1}(\vec{x}-\vec{\mu})\right\}\mathrm{d}\vec{x}\\
&= (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left\{\vec{x}'t\mathbf{A}\vec{x}-\frac{1}{2}(\vec{x}'\mathbf{\Sigma}^{-1}\vec{x}-2\vec{x}'\mathbf{\Sigma}^{-1}\vec{\mu}+\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu})\right\}\mathrm{d}\vec{x}\\
&= (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}\right\}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\vec{x}'\underbrace{(\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma})\mathbf{\Sigma}^{-1}}_{\mathbf{W}}\vec{x}+\underbrace{\vec{\mu}'\mathbf{\Sigma}^{-1}}_{\vec{b}}\vec{x}\right\}\mathrm{d}\vec{x}\\
&= (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}\right\}(2\pi)^{n/2}\left|\mathbf{W}^{-1}\right|^{1/2}\exp\left\{\frac{\vec{b}'\mathbf{W}^{-1}\vec{b}}{2}\right\}\text{ by Lemma 3.1}\\
&= |\mathbf{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}\right\}\left|\mathbf{W}^{-1}\right|^{1/2}\exp\left\{\frac{\vec{b}'\mathbf{W}^{-1}\vec{b}}{2}\right\}.
\end{aligned}$$

Note that

$$\mathbf{W}^{-1} = \big[(\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma})\mathbf{\Sigma}^{-1}\big]^{-1} = \mathbf{\Sigma}(\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma})^{-1},$$

and

$$\begin{aligned}
\vec{b}\mathbf{W}^{-1}\vec{b} &= \vec{\mu}'\mathbf{\Sigma}^{-1}\mathbf{\Sigma}(\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma})^{-1}\mathbf{\Sigma}^{-1}\vec{\mu}\\
&= \vec{\mu}'(\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma})^{-1}\mathbf{\Sigma}^{-1}\vec{\mu},
\end{aligned}$$

and

$$|\mathbf{W}^{-1}|^{1/2} = \left|\big[(\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma})\mathbf{\Sigma}^{-1}\big]^{-1}\right|^{1/2} = |\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma}|^{-1/2}|\mathbf{\Sigma}|^{1/2}.$$

Continuing,

$$\begin{aligned}
M_{\vec{X}'\mathbf{A}\vec{X}}(t) &= |\mathbf{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}\right\}\left|\mathbf{W}^{-1}\right|^{1/2}\exp\left\{\frac{\vec{b}'\mathbf{W}^{-1}\vec{b}}{2}\right\}\\
&= \exp\left\{-\frac{1}{2}\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}\right\}\exp\left\{\frac{1}{2}\vec{\mu}'(\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma})\mathbf{\Sigma}^{-1}\vec{\mu}\right\}|\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma}|^{-1/2}\\
&= \exp\left\{-\frac{1}{2}\vec{\mu}'\mathbf{\Sigma}^{-1}\vec{\mu}+\frac{1}{2}\vec{\mu}'(\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma})\mathbf{\Sigma}^{-1}\vec{\mu}\right\}|\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma}|^{-1/2}\\
&= \frac{1}{\sqrt{|\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma}|}}\exp\left\{-\frac{1}{2}\vec{\mu}'\big[\mathbf{I}-(\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma})^{-1}\big]\mathbf{\Sigma}^{-1}\vec{\mu}\right\}.
\end{aligned}$$

LEMMA 3.3

Let $\lambda_1,\ldots,\lambda_n$ be the eigenvalues of $\mathbf{A}\mathbf{\Sigma}$. Then,

$$|\mathbf{I}-2t\mathbf{A}\mathbf{\Sigma}| = (1-2t\lambda_1)\cdots(1-2t\lambda_n).$$

Proof: By the spectral theorem,

$$\mathbf{A}\mathbf{\Sigma} = \mathbf{Q}'\operatorname{diag}(\lambda_1,\ldots,\lambda_n)\mathbf{Q}.$$

Then,

$$\begin{aligned}\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma} &= \mathbf{I} - 2t\mathbf{Q}' \operatorname{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{Q} \\ &= \mathbf{I} - \mathbf{Q}' \operatorname{diag}(2t\lambda_1, \ldots, 2t\lambda_n)\mathbf{Q} \\ &= \mathbf{Q}' \operatorname{diag}(1 - 2t\lambda_1, \ldots, 1 - 2t\lambda_n)\mathbf{Q}.\end{aligned}$$

Therefore,

$$\begin{aligned}|\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma}| &= |\mathbf{Q}'||\operatorname{diag}(1 - 2t\lambda_1, \ldots, 1 - 2t\lambda_n)||\mathbf{Q}| \\ &= (1 - 2t\lambda_1)\cdots(1 - 2t\lambda_n).\end{aligned}$$

LEMMA 3.4

For t small enough, we have

$$\mathbf{I} - (\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma})^{-1} = -\sum_{r=1}^{\infty}(2t)^r(\mathbf{A}\mathbf{\Sigma})^r.$$

Proof: If t is small enough, then $\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma}$ is invertible. Thus,

$$(\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma})\left[\mathbf{I} - (\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma})^{-1}\right] = \mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma} - \mathbf{I} = -2t\mathbf{A}\mathbf{\Sigma}.$$

On the other hand,

$$\begin{aligned}(\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma})\left(-\sum_{r=1}^{\infty}(2t)^r(\mathbf{A}\mathbf{\Sigma})^r\right) &= -\sum_{r=1}^{\infty}(2t)^r(\mathbf{A}\mathbf{\Sigma})^r + \underbrace{\sum_{r=2}^{\infty}(2t)^r(\mathbf{A}\mathbf{\Sigma})^r}_{\sum_{r=1}^{\infty}(2t)^{r+1}(\mathbf{A}\mathbf{\Sigma})^{r+1}} \\ &= -(2t)\mathbf{A}\mathbf{\Sigma}.\end{aligned}$$

Therefore,

$$(\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma})\left[\mathbf{I} - (\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma})^{-1}\right] = (\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma})\left(-\sum_{r=1}^{\infty}(2t)^r(\mathbf{A}\mathbf{\Sigma})^r\right)$$

For small enough t, the inverse of $\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma}$ exists, so

$$\mathbf{I} - (\mathbf{I} - 2t\mathbf{A}\mathbf{\Sigma})^{-1} = -\sum_{r=1}^{\infty}(2t)^r(\mathbf{A}\mathbf{\Sigma})^r.$$

DEFINITION 3.1: Cumulant Generating Function

Let $M_X(t)$ be the moment generating function of X. Then,

$$K_X(t) = \log(M_X(t))$$

is called the **cumulant generating function**. By Taylor's expansion,

$$K_X(t) = \sum_{n=1}^{\infty}\kappa_n\frac{t^n}{n!}.$$

$\kappa_n = K^{(n)}(0)$ is the n-th cumulant.

EXAMPLE 3.1

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then,

$$M_X(t) = \exp\left\{\mu t + \frac{\sigma^2}{2}t^2\right\} \implies K_X(t) = \mu t + \frac{\sigma^2}{2}t^2 \implies \kappa_1 = \mu, \quad \kappa_2 = \sigma^2, \quad \kappa_i = 0,\ i \geq 3.$$

LEMMA 3.5

For any X with $K_X(t)$ well-defined,

$$\kappa_1 = \mathbb{E}[X], \quad \kappa_2 = \mathrm{Var}(X).$$

Proof:

$$\begin{aligned}
\left.\frac{\mathrm{d}K_X(t)}{\mathrm{d}t}\right|_{t=0} &= \left.\frac{\mathrm{d}\log(M_X(t))}{\mathrm{d}t}\right|_{t=0} \\
&= \left.\frac{M_X'(t)}{M_X(t)}\right|_{t=0} \\
&= M_X'(0) \\
&= \mathbb{E}[X] \\
&= \kappa_1.
\end{aligned}$$

$$\begin{aligned}
\left.\frac{\mathrm{d}K_X'(t)}{\mathrm{d}t}\right|_{t=0} &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\frac{M_X'(t)}{M_X(t)}\right|_{t=0} \\
&= \left.\frac{M_X''(t)M_X(t) - (M_X'(t))^2}{M_X^2(t)}\right|_{t=0} \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \mathrm{Var}(X) \\
&= \kappa_2.
\end{aligned}$$

THEOREM 3.1

Let $\vec{X} \sim MN(\vec{\mu}, \mathbf{\Sigma})$. For any symmetric matrix $\mathbf{A}$ and $\lambda_1, \ldots, \lambda_n$ are eigenvalues of $\mathbf{A\Sigma}$

$$K_{\vec{X}'\mathbf{A}\vec{X}}(t) = -\frac{1}{2}\sum_{i=1}^{n}\log(1 - 2t\lambda_i) + \frac{1}{2}\vec{\mu}'\sum_{r=1}^{\infty}(2t)^r(\mathbf{A\Sigma})^r\mathbf{\Sigma}^{-1}\vec{\mu}.$$

$$\kappa_r = 2^{r-1}(r-1)!\big[\mathrm{tr}((\mathbf{A\Sigma})^r) + r\vec{\mu}'\mathbf{A}(\mathbf{\Sigma A})^{r-1}\vec{\mu}\big].$$

In particular,

$$\begin{aligned}
\kappa_1 &= \mathbb{E}[\vec{X}'\mathbf{A}\vec{X}] = \mathrm{tr}(\mathbf{A\Sigma}) + \vec{\mu}'\mathbf{A}\vec{\mu}. \\
\kappa_2 &= \mathrm{Var}(\vec{X}'\mathbf{A}\vec{X}) \\
&= 2\big[\mathrm{tr}\big((\mathbf{A\Sigma})^2\big) + 2\vec{\mu}'\mathbf{A\Sigma A}\vec{\mu}\big] \\
&= 2\,\mathrm{tr}\big((\mathbf{A\Sigma})^2\big) + 4\vec{\mu}'\mathbf{A\Sigma A}\vec{\mu}.
\end{aligned}$$

Proof:

Step 1: Since $\vec{X} \sim \text{MN}(\vec{\mu}, \mathbf{\Sigma})$, by Lemma 3.2,

$$M_{\vec{X}'\mathbf{A}\vec{X}}(t) = |\mathbf{I} - 2t\mathbf{A\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}\vec{\mu}'\left[\mathbf{I} - (\mathbf{I} - 2t\mathbf{A\Sigma})^{-1}\right]\mathbf{\Sigma}^{-1}\vec{\mu}\right\}.$$

Therefore,

$$\begin{aligned}
K_{\vec{X}'\mathbf{A}\vec{X}}(t) &= \log(M_{\vec{X}'\mathbf{A}\vec{X}}(t)) \\
&= \log\left(|\mathbf{I} - 2t\mathbf{A\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}\vec{\mu}'\left[\mathbf{I} - (\mathbf{I} - 2t\mathbf{A\Sigma})^{-1}\right]\mathbf{\Sigma}^{-1}\vec{\mu}\right\}\right) \\
&= -\frac{1}{2}\log(|\mathbf{I} - 2t\mathbf{A\Sigma}|) - \frac{1}{2}\vec{\mu}'\left[\mathbf{I} - (\mathbf{I} - 2t\mathbf{A\Sigma})^{-1}\right]\mathbf{\Sigma}^{-1}\vec{\mu} \\
&= -\frac{1}{2}\log((1-2t\lambda_1)\cdots(1-2t\lambda_n)) - \frac{1}{2}\vec{\mu}'\left(-\sum_{r=1}^{\infty}(2t)^r(\mathbf{A\Sigma})^r\right)\mathbf{\Sigma}^{-1}\vec{\mu} \quad \text{Lemma 3.3, 3.4} \\
&= -\frac{1}{2}\sum_{i=1}^{n}\log(1-2t\lambda_i) + \frac{1}{2}\vec{\mu}'\sum_{r=1}^{\infty}(2t)^r(\mathbf{A\Sigma})^r\mathbf{\Sigma}^{-1}\vec{\mu}.
\end{aligned}$$

Step 2: The Taylor expansion for $\log(1-x)$ is

$$-x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots = -\sum_{r=1}^{\infty}\frac{x^r}{r}.$$

Therefore, using the Taylor expansion of $\log(1-2t\lambda_i)$, we may rewrite the first term of $K_{\vec{X}'\mathbf{A}\vec{X}}(t)$ as

$$\begin{aligned}
-\frac{1}{2}\sum_{i=1}^{n}\log(1-2t\lambda_i) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{r=1}^{\infty}\frac{(2t\lambda_i)^r}{r} \\
&= \frac{1}{2}\sum_{r=1}^{\infty}\frac{(2t)^r\sum_{i=1}^{n}\lambda_i^r}{r} \\
&= \frac{1}{2}\sum_{r=1}^{\infty}\frac{(r-1)!(2t)^r\,\text{tr}((\mathbf{A\Sigma})^r)}{r!} \qquad \text{since } \frac{(r-1)!}{r!} = \frac{1}{r}.
\end{aligned}$$

Step 3: Rewrite the second term of $K_{\vec{X}'\mathbf{A}\vec{X}}(t)$ as

$$\frac{1}{2}\vec{\mu}'\sum_{r=1}^{\infty}(2t)^r(\mathbf{A\Sigma})^r\mathbf{\Sigma}^{-1}\vec{\mu} = \frac{1}{2}\sum_{r=1}^{\infty}(2t)^r\left(\vec{\mu}'(\mathbf{A\Sigma})^r\mathbf{\Sigma}^{-1}\vec{\mu}\right) = \frac{1}{2}\sum_{r=1}^{\infty}\frac{r!2^rt^r\vec{\mu}'(\mathbf{A\Sigma})^r\mathbf{\Sigma}^{-1}\vec{\mu}}{r!}.$$

Step 4: Combining steps 1 to 3, we get

$$\begin{aligned}
K_{\vec{X}'\mathbf{A}\vec{X}}(t) &= \frac{1}{2}\sum_{r=1}^{\infty}\frac{(r-1)!(2t)^r\,\text{tr}((\mathbf{A\Sigma})^r)}{r!} + \frac{r!2^rt^r\vec{\mu}'(\mathbf{A\Sigma})^r\mathbf{\Sigma}^{-1}\vec{\mu}}{r!} \\
&= \sum_{r=1}^{\infty}2^{r-1}(r-1)!\frac{[\text{tr}((\mathbf{A\Sigma})^r) + r\vec{\mu}'(\mathbf{A\Sigma})^r\mathbf{\Sigma}^{-1}\vec{\mu}]t^r}{r!} \\
&= \sum_{r=1}^{\infty}\underbrace{2^{r-1}(r-1)!\left[\text{tr}((\mathbf{A\Sigma})^r) + r\vec{\mu}'(\mathbf{A\Sigma})^r\mathbf{\Sigma}^{-1}\vec{\mu}\right]}_{\kappa_r}\frac{t^r}{r!},
\end{aligned}$$

noting that $(\mathbf{A\Sigma})^r\mathbf{\Sigma}^{-1} = (\mathbf{\Sigma A})^{r-1}$ to get the desired result.

4 Lecture 4: Quadratic Forms with Idempotency

LEMMA 4.1

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be symmetric and $\mathbf{B} \in \mathbb{R}^{n\times n}$ be positive definite. If the eigenvalues of $\mathbf{AB}$ are 0's or 1's, then $\mathbf{AB}$ is idempotent.

Proof: By Cholesky decomposition, there exists an invertible lower triangular matrix $\mathbf{L}$ such that

$$\mathbf{B} = \mathbf{LL}'.$$

If the eigenvalues of $\mathbf{AB}$ are 0's or 1's, then the equation $|\mathbf{AB} - \lambda\mathbf{I}| = 0$ has roots 0 or 1.

$$\begin{aligned} |\mathbf{AB} - \lambda\mathbf{I}| &= \left|\mathbf{L}'(\mathbf{AB} - \lambda\mathbf{I})(\mathbf{L}')^{-1}\right| \\ &= \left|\mathbf{L}'\mathbf{AB}(\mathbf{L}')^{-1} - \lambda\mathbf{I}\right| \\ &= \left|\mathbf{L}'\mathbf{ALL}'(\mathbf{L}')^{-1} - \lambda\mathbf{I}\right| \\ &= |\mathbf{L}'\mathbf{AL} - \lambda\mathbf{I}| \\ &= 0 \end{aligned}$$

has roots 0 or 1. Since $\mathbf{L}'\mathbf{AL}$ is symmetric, and thus diagonalizable, it follows that

$$\mathbf{L}'\mathbf{AL}$$

is idempotent since

$$\begin{aligned} (\mathbf{L}'\mathbf{AL})(\mathbf{L}'\mathbf{AL}) &= \mathbf{Q}'\operatorname{diag}(\lambda_1, \dots, \lambda_n)\underbrace{\mathbf{QQ}'}_{\mathbf{I}}\operatorname{diag}(\lambda_1, \dots, \lambda_n)\mathbf{Q} \\ &= \mathbf{Q}'\operatorname{diag}(\lambda_1^2, \dots, \lambda_n^2)\mathbf{Q} \\ &= \mathbf{Q}'\operatorname{diag}(\lambda_1, \dots, \lambda_n)\mathbf{Q} \\ &= \mathbf{L}'\mathbf{AL}. \end{aligned}$$

Therefore,

$$\begin{aligned} & \mathbf{L}'\mathbf{AL} = \mathbf{L}'\mathbf{ALL}'\mathbf{AL} \\ \implies & \mathbf{AL} = \mathbf{ALL}'\mathbf{AL} \\ \implies & \mathbf{ALL}' = \mathbf{AB} = \mathbf{ALL}'\mathbf{ALL}' = \mathbf{ABAB} \\ \implies & \mathbf{AB} \text{ is idempotent.} \end{aligned}$$

THEOREM 4.1

Let $\vec{X} \sim MN(\vec{\mu}, \mathbf{\Sigma})$ and $\mathbf{A}$ be a symmetric matrix with rank r. Then,

$$\vec{X}'\mathbf{A}\vec{X} \sim \chi^2(r, \lambda), \ \forall\vec{\mu}$$

with $\lambda = \frac{1}{2}\vec{\mu}'\mathbf{A}\vec{\mu}$ if and only if $\mathbf{A\Sigma}$ is idempotent.

Proof:
($\Longleftarrow$) Assume that $\mathbf{A\Sigma}$ is idempotent, then all eigenvalues of $\mathbf{A\Sigma}$ are 1 or 0 (which we denote as λ_i).

Since $\mathbf{\Sigma}$ has full rank,

$$\operatorname{rank}(\mathbf{A\Sigma}) = \operatorname{rank}(\mathbf{A}) = r.$$

Therefore, r eigenvalues are 1 and $n-r$ are 0. By Theorem 3.1, we have

$$\begin{aligned}
M_{\vec{X}'\mathbf{A}\vec{X}}(t) &= \prod_{i=1}^{n}(1-2t\lambda_i)^{-1/2}\exp\left\{\frac{1}{2}\vec{\mu}'\sum_{j=1}^{\infty}(2t)^j(\mathbf{A\Sigma})^j\mathbf{\Sigma}^{-1}\vec{\mu}\right\} \\
&= (1-2t)^{-r/2}\exp\left\{\frac{1}{2}\vec{\mu}'\sum_{j=1}^{\infty}(2t)^j\mathbf{A\Sigma\Sigma}^{-1}\vec{\mu}\right\} \\
&= (1-2t)^{-r/2}\exp\left\{\frac{1}{2}\vec{\mu}'\mathbf{A}\vec{\mu}\sum_{j=1}^{\infty}(2t)^j\right\} \\
&= (1-2t)^{-r/2}\exp\left\{\frac{\vec{\mu}'\mathbf{A}\vec{\mu}}{2}\frac{2t}{1-2t}\right\} \\
&= (1-2t)^{-r/2}\exp\left\{\lambda\frac{2t}{1-2t}\right\}.
\end{aligned}$$

Let $\eta \sim \chi^2(r,\lambda)$. By definition,

$$\eta = X_1^2 + \cdots + X_r^2,$$

where $X_i \sim \mathcal{N}(\mu_i, 1)$ and $X_1, \ldots, X_r$ are independent.

$$\begin{aligned}
M_\eta(t) &= \mathbb{E}[e^{t\eta}] \\
&= \mathbb{E}[e^{t(X_1^2+\cdots+X_r^2)}] \\
&= \prod_{i=1}^{r}\mathbb{E}[e^{tX_i^2}].
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{E}[e^{tX_i^2}] &= \int_{-\infty}^{\infty}(2\pi)^{-1/2}\exp\{tx^2\}\exp\left\{-\frac{(x-\mu_i)^2}{2}\right\}\mathrm{d}x \\
&= \int_{-\infty}^{\infty}(2\pi)^{-1/2}\exp\left\{-\frac{1}{2}(x^2-2tx^2-2\mu_i x+\mu_i^2)\right\}\mathrm{d}x \\
&= (1-2t)^{-1/2}\exp\left\{-\frac{1-2t}{2}\left(\frac{\mu_i^2}{1-2t}-\left(\frac{\mu_i}{1-2t}\right)^2\right)\right\} \\
&\qquad \int_{-\infty}^{\infty}\underbrace{\frac{1}{\sqrt{2\pi}(1-2t)^{-1/2}}\exp\left\{-\frac{(x-\frac{\mu_i^2}{1-2t})^2}{2(1-2t)^{-1}}\right\}}_{\mathcal{N}\left(\frac{\mu_i^2}{1-2t},(1-2t)^{-1}\right)}\mathrm{d}x \\
&= (1-2t)^{-1/2}\exp\left\{\mu_i^2\frac{t}{1-2t}\right\} \\
&= (1-2t)^{-1/2}\exp\left\{\frac{\mu_i^2}{2}\frac{2t}{1-2t}\right\}.
\end{aligned}$$

Hence,

$$
\begin{aligned}
M_\eta(t) &= \prod_{i=1}^{r}(1-2t)^{-1/2}\exp\left\{\frac{\mu_i^2}{2}\frac{2t}{1-2t}\right\} \\
&= (1-2t)^{-r/2}\exp\left\{\underbrace{\frac{1}{2}\sum_{i=1}^{n}\mu_i^2}_{\lambda}\frac{2t}{1-2t}\right\} \\
&= (1-2t)^{-r/2}\exp\left\{\frac{\lambda 2t}{1-2t}\right\},
\end{aligned}
$$

which is the mgf of $\chi^2(r,\lambda)$. By uniqueness of moment generating functions,

$$\vec{X}'\mathbf{A}\vec{X} \sim \chi^2(r,\lambda).$$

($\Longrightarrow$) Assume $\vec{X}'\mathbf{A}\vec{X} \sim \chi^2(r,\lambda)$ for all $\vec{\mu}$. Choose $\vec{\mu}=\vec{0}$, then $\lambda = 0$.

$$
\begin{aligned}
M_{\vec{X}'\mathbf{A}\vec{X}}(t) &= \prod_{i=1}^{n}(1-2t\lambda_i)^{-1/2} \\
&= (1-2t)^{-r/2}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\Longrightarrow \prod_{i=1}^{n}(1-2t\lambda_i) = (1-2t)^r && \text{cancel exponents} \\
&\Longrightarrow \sum_{i=1}^{n}\log(1-2t\lambda_i) = r\log(1-2t) && \text{take logarithm} \\
&\Longrightarrow \sum_{i=1}^{n}\left[\sum_{\ell=1}^{\infty}\frac{(2t\lambda_i)^\ell}{\ell}\right] = r\sum_{\ell=1}^{\infty}\frac{(2t)^\ell}{\ell} && \text{Taylor expansion} \\
&\Longrightarrow \sum_{\ell=1}^{\infty}\frac{(\sum_{i=1}^{n}\lambda_i^\ell - r)(2t)^\ell}{\ell} = 0 && \text{re-order summation}
\end{aligned}
$$

Therefore,

$$\sum_{i=1}^{n}\lambda_i^\ell = r,\ \forall \ell \geq 1.$$

<u>Case 1</u>: If $|\lambda_i| \geq 1$ for some i, then choose $\ell = 2k$ and let $2k \to \infty$, then

$$\lambda_1^{2k} + \cdots + \lambda_i^{2k} + \cdots + \lambda_n^{2k} = r,$$

but the left-hand side is $\infty \neq r$, contradiction. Thus, $|\lambda_i| \leq 1$ for all i.
<u>Case 2</u>: If $|\lambda_i| < 1$ for some i, then choose $\ell = 2k$ and let $k \to \infty$, $|\lambda_i|^{2k} \to 0$. Hence, the total terms with $|\lambda_i| < 1$ will be $n - r$. The equality

$$\lambda_1 + \cdots + \lambda_n = r$$

implies that all the terms with $|\lambda_i| = 1$ are actually λ_i. Why? Let d be the number of i such that $|\lambda_i| < 1$, there will be $n - d$ of $|\lambda_i| = 1$ (fill in details).
Hence, the eigenvalues of $\mathbf{A\Sigma}$ are 1 or 0. Since $\mathbf{\Sigma}$ is positive definite, it follows from Lemma 4.1 that $\mathbf{A\Sigma}$ is idempotent.

EXAMPLE 4.1

Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$. Then,
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

Solution: Let
$$\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}.$$
Then,
$$\vec{X} \sim \text{MN}(\mu\vec{j}, \sigma^2\mathbf{I}).$$
$$\begin{aligned} \frac{(n-1)S^2}{\sigma^2} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \frac{1}{\sigma^2}\vec{X}'(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\vec{X} \\ &= \vec{Y}'\mathbf{A}\vec{Y}, \end{aligned}$$
where $\vec{Y} = \frac{1}{\sigma}\vec{X}$ and $\mathbf{A} = \mathbf{I} - \frac{1}{n}\mathbf{J}$, so
$$\vec{Y} \sim \text{MN}\Big(\frac{\mu}{\sigma}\vec{j}, \mathbf{I}\Big).$$
$$\mathbf{A\Sigma} = (\mathbf{I} - \tfrac{1}{n}\mathbf{J})\mathbf{I} = \mathbf{I} - \tfrac{1}{n}\mathbf{J} = \mathbf{A}.$$
Also,
$$\mathbf{A}^2 = \mathbf{I} - \tfrac{2}{n}\mathbf{J} + \tfrac{1}{n^2}\mathbf{JJ} = \mathbf{I} - \tfrac{1}{n}\mathbf{J} = \mathbf{A}.$$
Therefore, $\mathbf{A\Sigma}$ is idempotent. By Theorem 5.1,
$$\vec{Y}'\mathbf{A}\vec{Y} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(r, \lambda),$$
with
$$r = \text{rank}(A) = \text{rank}(\mathbf{I} - \tfrac{1}{n}\mathbf{J}) = n - 1$$
and
$$\begin{aligned} \lambda &= \frac{1}{2}\vec{\mu}'\mathbf{A}\vec{\mu} \\ &= \frac{1}{2}\frac{\mu}{\sigma}\vec{j}'(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\frac{\mu}{\sigma}\vec{j} \\ &= \frac{\mu^2}{2\sigma^2}(\vec{j}'\vec{j} - \tfrac{1}{n}\vec{j}'\mathbf{J}\vec{j}) \\ &= \frac{\mu^2}{2\sigma^2}(n - \tfrac{1}{n}n^2) \\ &= 0. \end{aligned}$$
Therefore, $\vec{Y}'\mathbf{A}\vec{Y} \sim \chi^2(r, 0) = \chi^2(r)$.

LECTURE 5
23rd January

5 Lecture 5: Criteria for Independence

LEMMA 5.1

Let $\mathbf{A}$ be a symmetric positive semidefinite $n \times n$ matrix with rank r. Then, there exists an $n \times r$ matrix $\mathbf{D}$ with rank r such that

$$\mathbf{A} = \mathbf{D}\mathbf{D}'$$

Proof: If $\mathbf{A}$ is symmetric, then by the spectral theorem

$$\mathbf{A} = \mathbf{Q}' \operatorname{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{Q},$$

where $\mathbf{Q}$ is orthogonal where r of λ_i are non-zero since $\operatorname{rank}(\mathbf{A}) = r$. Without loss of generality, we can assume that $\lambda_1, \ldots, \lambda_r > 0$, $\lambda_j = 0$ for $j > r$. Define

$$\mathbf{D} = \mathbf{Q}' \begin{bmatrix} \mathbf{\Lambda}^{1/2} \\ \mathbf{O} \end{bmatrix}_{n \times r},$$

where

$$\mathbf{\Lambda}^{1/2} = \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r}).$$

Hence,

$$\mathbf{D}\mathbf{D}' = \mathbf{Q}' \begin{bmatrix} \mathbf{\Lambda}^{1/2} \\ \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}^{1/2} & \mathbf{O} \end{bmatrix} \mathbf{Q} = \mathbf{Q}' \begin{bmatrix} \mathbf{\Lambda} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \mathbf{Q} = \mathbf{A}.$$

THEOREM 5.1

Let $\vec{X} \sim MN(\vec{\mu}, \mathbf{\Sigma})$, $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric and $\mathbf{B} \in \mathbb{R}^{k \times n}$. $\vec{X}'\mathbf{A}\vec{X}$ and $\mathbf{B}\vec{X}$ are independent if and only if $\mathbf{B\Sigma A} = \mathbf{O}$.

Proof: We assume that $\mathbf{A}$ is positive semidefinite.
Step 1: Let $r = \operatorname{rank}(\mathbf{A})$, $\mathbf{A} = \mathbf{D}\mathbf{D}'$ from Lemma 5.1. We know that $\mathbf{D}$ is $n \times r$ with $\operatorname{rank}(\mathbf{D}) = r$, and since $\operatorname{rank}(\mathbf{D}\mathbf{D}') = \operatorname{rank}(\mathbf{D}) = r$, then $\mathbf{D}'\mathbf{D}$ is invertible. We will show that

$$\mathbf{B\Sigma A} = \mathbf{O} \iff \mathbf{B\Sigma D} = \mathbf{O}.$$

($\implies$) Note that

$$\begin{aligned} \mathbf{B\Sigma A} = \mathbf{B\Sigma D D}' &= \mathbf{O} \\ \implies \mathbf{B\Sigma D D}'\mathbf{D} &= \mathbf{O} \\ \implies \mathbf{B\Sigma D D}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1} = \mathbf{B\Sigma D} &= \mathbf{O}. \end{aligned}$$

On the other hand, if $\mathbf{B\Sigma D} = \mathbf{O}$, then

$$\mathbf{B\Sigma D D}' = \mathbf{B\Sigma A} = \mathbf{O}.$$

Step 2 (Sufficiency): Assume that $\mathbf{B\Sigma A} = \mathbf{O}$, then $\mathbf{B\Sigma D} = \mathbf{O}$. By direct calculation,

$$\operatorname{Cov}(\mathbf{B}\vec{X}, \vec{X}'\mathbf{D}) = \mathbf{B} \operatorname{Cov}(\vec{X}, \vec{X})\mathbf{D} = \mathbf{B\Sigma D} = \mathbf{O}.$$

Since $\mathbf{B}\vec{X}$ and $\vec{X}'\mathbf{D}$ are multivariate normal, it follows that $\mathbf{B}\vec{X}$ and $\vec{X}'\mathbf{D}$ are independent. Noting that

$$\begin{aligned} \vec{X}'\mathbf{A}\vec{X} &= \vec{X}'\mathbf{D}\mathbf{D}'\vec{X} && \text{Lemma 5.1} \\ &= (\vec{X}'\mathbf{D})(\vec{X}'\mathbf{D})', \end{aligned}$$

which is a function of $\vec{X}'\mathbf{D}$. We know that if X and Y are independent, then for any measurable function $f(X)$ and $g(Y)$ are independent. Hence, $\mathbf{B}\vec{X}$ and $\vec{X}'\mathbf{A}\vec{X}$ are independent.
Step 3 (Necessity): Assume that $\mathbf{B}\vec{X}$ and $\vec{X}'\mathbf{A}\vec{X}$ are independent. By direct calculation,

$$\begin{aligned}
\operatorname{Cov}(\mathbf{B}\vec{X}, \vec{X}'\mathbf{A}\vec{X}) &= \mathbf{B}\operatorname{Cov}(\vec{X}, \vec{X}'\mathbf{A}\vec{X}) \\
&= \mathbf{B}\,\mathbb{E}\big[(\vec{X}-\vec{\mu})(\vec{X}'\mathbf{A}\vec{X} - \mathbb{E}[\vec{X}'\mathbf{A}\vec{X}])\big] \\
&= \mathbf{B}\,\mathbb{E}\big[(\vec{X}-\vec{\mu})(\vec{X}'\mathbf{A}\vec{X} - \vec{\mu}'\mathbf{A}\vec{\mu} - \operatorname{tr}(\mathbf{A}\mathbf{\Sigma}))\big] \\
&= \mathbf{B}\,\mathbb{E}\big[(\vec{X}-\vec{\mu})(\vec{X}'\mathbf{A}\vec{X} - \vec{\mu}'\mathbf{A}\vec{\mu})\big] + \mathbf{B}\underbrace{\mathbb{E}\big[(\vec{X}-\vec{\mu})\big]}_{0}\operatorname{tr}(\mathbf{A}\mathbf{\Sigma}) \\
&= \mathbf{B}\,\mathbb{E}\big[(\vec{X}-\vec{\mu})(\vec{X}'\mathbf{A}\vec{X} - \vec{\mu}\mathbf{A}\vec{\mu})\big] \\
&= \mathbf{B}\,\mathbb{E}\Big[(\vec{X}-\vec{\mu})\big[(\vec{X}-\vec{\mu})'\mathbf{A}(\vec{X}-\vec{\mu}) + 2(\vec{X}-\vec{\mu})'\mathbf{A}\vec{\mu}\big]\Big] \\
&= \mathbf{B}\,\mathbb{E}\Big[(\vec{X}-\vec{\mu})(\vec{X}-\vec{\mu})'\mathbf{A}(\vec{X}-\vec{\mu})\Big] + 2\mathbf{B}\,\mathbb{E}\Big[(\vec{X}-\vec{\mu})'\mathbf{A}\vec{\mu}\Big] \\
&= \mathbf{B}\,\mathbb{E}\Big[(\vec{X}-\vec{\mu})(\vec{X}-\vec{\mu})'\mathbf{A}(\vec{X}-\vec{\mu})\Big] + 2\mathbf{B}\mathbf{\Sigma}\mathbf{A}\vec{\mu}.
\end{aligned}$$

To show that the first term is zero, using the spectral theorem re-write $\mathbf{A}$, define $\vec{Y} = \vec{X} - \vec{\mu}$, and use the fact that the third moments of multivariate normal are 0 (exercise). Hence,

$$\mathbf{B}\mathbf{\Sigma}\mathbf{A}\vec{\mu} = \mathbf{O}.$$

Since $\vec{\mu}$ is arbitrary, it follows that

$$\mathbf{B}\mathbf{\Sigma}\mathbf{A} = \mathbf{O}.$$

THEOREM 5.2

Let $\vec{X} \sim MN(\vec{\mu}, \mathbf{\Sigma})$, $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ be symmetric matrices. $\vec{X}'\mathbf{A}\vec{X}$ and $\vec{X}'\mathbf{B}\vec{X}$ are independent if and only if $\mathbf{A}\mathbf{\Sigma}\mathbf{B} = \mathbf{O}$.

Proof: Let $\operatorname{rank}(\mathbf{A}) = r$, $\operatorname{rank}(\mathbf{B}) = s$. By the spectral theorem, there are orthogonal matrices $\mathbf{Q}_1$ and $\mathbf{Q}_2$ such that

$$\begin{aligned}
\mathbf{A} &= \mathbf{Q}_1' \operatorname{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{Q}_1, \\
\mathbf{B} &= \mathbf{Q}_2' \operatorname{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n)\mathbf{Q}_2.
\end{aligned}$$

Without loss of generality, we assume that

- $\lambda_1, \ldots, \lambda_r \neq 0$, $\lambda_j = 0$ for $j > r$,
- $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_s \neq 0$, $\tilde{\lambda}_i = 0$ for $i > s$.

Set

$$\mathbf{D}_r = \operatorname{diag}(\lambda_1, \ldots, \lambda_r), \qquad \tilde{\mathbf{D}}_s = \operatorname{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_s).$$

Hence, $\mathbf{Q}_1' = \begin{pmatrix}\mathbf{Q}_{11}' & \mathbf{Q}_{12}'\end{pmatrix}$ with $\mathbf{Q}_{11}'$ being $n \times r$ and $\operatorname{rank}(\mathbf{Q}_{11}') = r$. Then,

$$\begin{aligned}
\mathbf{A} &= \begin{pmatrix}\mathbf{Q}_{11}' & \mathbf{Q}_{12}'\end{pmatrix}\begin{pmatrix}\mathbf{D}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O}\end{pmatrix}\begin{pmatrix}\mathbf{Q}_{11} \\ \mathbf{Q}_{12}\end{pmatrix} \\
&= \mathbf{Q}_{11}'\mathbf{D}_r\mathbf{Q}_{12}.
\end{aligned}$$

Define $\mathbf{Q}_2' = \begin{pmatrix}\tilde{\mathbf{Q}}_{11}' & \tilde{\mathbf{Q}}_{12}'\end{pmatrix}$ to similarly get

$$\mathbf{B} = \tilde{\mathbf{Q}}_{11}'\tilde{\mathbf{D}}_s\tilde{\mathbf{Q}}_{12}.$$

($\Longleftarrow$) "Sufficiency:" Assume that $\mathbf{A\Sigma B} = \mathbf{O}$, so

$$\begin{aligned}
\mathbf{A\Sigma B} = \mathbf{Q}_{11}'\mathbf{D}_r\mathbf{Q}_{11}\mathbf{\Sigma}\tilde{\mathbf{Q}}_{11}'\tilde{\mathbf{D}}_s\tilde{\mathbf{Q}}_{11} &= \mathbf{O} \\
\implies \mathbf{Q}_{11}\mathbf{Q}_{11}'\mathbf{Q}_{11}'\mathbf{D}_r\mathbf{Q}_{11}\mathbf{\Sigma}\tilde{\mathbf{Q}}_{11}'\tilde{\mathbf{D}}_s\tilde{\mathbf{Q}}_{11} &= \mathbf{O} \\
\implies \mathbf{D}_r\mathbf{Q}_{11}\mathbf{\Sigma}\tilde{\mathbf{Q}}_{11}'\tilde{\mathbf{D}}_s &= \mathbf{O} \\
\implies \mathbf{Q}_{11}\mathbf{\Sigma}\tilde{\mathbf{Q}}_{11}' &= \mathbf{O}
\end{aligned}$$

Noting that

$$\operatorname{Cov}(\mathbf{Q}_{11}\vec{X}, \vec{X}'\tilde{\mathbf{Q}}_{11}') = \mathbf{Q}_{11}\mathbf{\Sigma}\tilde{\mathbf{Q}}_{11}' = \mathbf{O}.$$

Therefore, $\mathbf{Q}_{11}\vec{X}$ and $\vec{X}'\tilde{\mathbf{Q}}_{11}'$ are independent. Hence,

$$\begin{aligned}
\vec{X}'\mathbf{A}\vec{X} &= \vec{X}'\mathbf{Q}_{11}\mathbf{D}_r\mathbf{Q}_{11}'\vec{X} \\
&= (\mathbf{Q}_{11}'\vec{X})'\mathbf{D}_r\mathbf{Q}_{11}'\vec{X}
\end{aligned}$$

is a function of $\mathbf{Q}_{11}'\vec{X}$, and similarly $\vec{X}'\mathbf{B}\vec{X}$ is a function of $\vec{X}'\mathbf{Q}_{11}'$. Therefore, $\vec{X}'\mathbf{A}\vec{X}$ is independent of $\vec{X}'\mathbf{B}\vec{X}$.

($\Longrightarrow$) "Necessity:" Assume that $\vec{X}'\mathbf{A}\vec{X}$ and $\vec{X}'\mathbf{B}\vec{X}$ are independent. By Theorem 3.1, we have

$$\begin{aligned}
\operatorname{Var}(\vec{X}'\mathbf{A}\vec{X}) &= 2\operatorname{tr}((\mathbf{A\Sigma})^2) + 4\vec{\mu}'\mathbf{A\Sigma A}\vec{\mu}. \\
\operatorname{Var}(\vec{X}'\mathbf{B}\vec{X}) &= 2\operatorname{tr}((\mathbf{B\Sigma})^2) + 4\vec{\mu}'\mathbf{B\Sigma B}\vec{\mu}.
\end{aligned}$$

Since $(\mathbf{A}+\mathbf{B})$ is symmetric,

$$\begin{aligned}
\operatorname{Var}(\vec{X}'(\mathbf{A}+\mathbf{B})\vec{X}) &= \operatorname{Var}(\vec{X}'\mathbf{A}\vec{X} + \vec{X}'\mathbf{B}\vec{X}) \\
&= \operatorname{Var}(\vec{X}'\mathbf{A}\vec{X}) + \operatorname{Var}(\vec{X}'\mathbf{B}\vec{X}) && \text{by assumption.}
\end{aligned}$$

Hence,

$$2\operatorname{tr}(((\mathbf{A}+\mathbf{B})\mathbf{\Sigma})^2) + 4\vec{\mu}'(\mathbf{A}+\mathbf{B})\mathbf{\Sigma}(\mathbf{A}+\mathbf{B})\vec{\mu} = 2\operatorname{tr}((\mathbf{A\Sigma})^2 + (\mathbf{B\Sigma})^2) + 4\vec{\mu}'(\mathbf{A\Sigma A} + \mathbf{B\Sigma B})\vec{\mu}.$$

Therefore,

$$2\operatorname{tr}((\mathbf{A\Sigma B\Sigma}) + \operatorname{tr}(\mathbf{B\Sigma A\Sigma})) + 4\vec{\mu}'(\mathbf{A\Sigma B} + \mathbf{B\Sigma A})\vec{\mu} = 0.$$

By cyclic property of trace, we obtain

$$\operatorname{tr}(\mathbf{A\Sigma B\Sigma}) = \operatorname{tr}(\mathbf{\Sigma B\Sigma A}) = \operatorname{tr}(\mathbf{B\Sigma A\Sigma}).$$

On the other hand,

$$\vec{\mu}'\mathbf{A\Sigma B}\vec{\mu} + 4\vec{\mu}'\mathbf{A\Sigma B}\vec{\mu} = 0.$$

Choose $\vec{\mu} = \vec{0}$, we get

$$\operatorname{tr}(\mathbf{A\Sigma B\Sigma}) = 0.$$

Thus, $\vec{\mu}'\mathbf{A\Sigma B}\vec{\mu} = 0$ for all $\vec{\mu}$, which implies that

$$\mathbf{A\Sigma B} = \mathbf{O}.$$

EXAMPLE 5.1

Let $\vec{X}' = (X_1, X_2)' \sim \text{MN}(\vec{\mu}, \mathbf{I}_2)$. Show

$$(X_1 - X_2)^2 \text{ is independent of } (X_1 + X_2)^2.$$

Solution:

$$(X_1 - X_2)^2 = X_1^2 - 2X_1X_2 + X_2^2 = X_1^2 - X_1X - 2 - X_2X_1 + X_2^2 = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

$$(X_1 + X_2)^2 = X_1^2 + X_1X_2 + X_2X_1 + X_2^2 = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

Now,

$$\mathbf{A\Sigma B} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \mathbf{O},$$

as required. Therefore, $(X_1 - X_2)^2$ is independent of $(X_1 + X_2)^2$ by Theorem 5.2.

EXAMPLE 5.2

Let $\vec{X}' = (X_1, X_2)' \sim \text{MN}(\vec{\mu}, \mathbf{\Sigma})$. Define

$$\mathbf{B} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Find

$$\mathbf{A} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

such that $\vec{X}'\mathbf{A}\vec{X}$ is independent of $\vec{X}'\mathbf{B}\vec{X}$, where

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}.$$

Solution:

$$\vec{X}'\mathbf{B}\vec{X} = X_1^2 + 4X_1X_2 + X_2^2 = (X_1 + X_2)^2 + 2X_1X_2.$$

$$\begin{aligned} \mathbf{A\Sigma B} &= \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \\ &= \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} 0 & 1.5 \\ 1.5 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1.5b & 1.5a \\ 1.5c & 1.5b \end{pmatrix} \\ &= 1.5 \begin{pmatrix} b & a \\ c & b \end{pmatrix} \\ &= \mathbf{O} \end{aligned}$$

implies that $a = b = c = 0$, so $\mathbf{A} = \mathbf{O}$. Therefore, there is no quadratic form.

6 Lecture 6: Cochran's Theorem

LECTURE 6

26th January

LEMMA 6.1

Let $\mathbf{C} = \mathbf{A} + \mathbf{B}$. Assume that $\mathbf{A}$, $\mathbf{B}$ are both $n \times n$ symmetric. If $\mathbf{C}^2 = \mathbf{C}$, $\mathbf{A}^2 = \mathbf{A}$, $\mathbf{B}$ positive semidefinite,

$$\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) = \text{rank}(\mathbf{C}),$$

then
$$\mathbf{AB} = \mathbf{O}.$$

Proof: Let $\operatorname{rank}(\mathbf{A}) = r$, $\operatorname{rank}(\mathbf{B}) = s$, $\operatorname{rank}(\mathbf{C}) = t = r + s$. If $\mathbf{C}^2 = \mathbf{C}$, then there exists an orthogonal matrix $\mathbf{\Gamma}$ such that
$$\mathbf{\Gamma}'\mathbf{C}\mathbf{\Gamma} = \begin{pmatrix} \mathbf{I}_t & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}.$$

Since $\mathbf{C} = \mathbf{A} + \mathbf{B}$,
$$\mathbf{\Gamma}'\mathbf{A}\mathbf{\Gamma} + \mathbf{\Gamma}'\mathbf{B}\mathbf{\Gamma} = \begin{pmatrix} \mathbf{I}_t & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}.$$

$\mathbf{A}$ and $\mathbf{B}$ are positive semidefinite implies that $\mathbf{\Gamma}'\mathbf{A}\mathbf{\Gamma}$ and $\mathbf{\Gamma}'\mathbf{B}\mathbf{\Gamma}$ are positive semidefinite. If the element on the diagonal is zero, then the corresponding row and columns are zeros. Hence,

$$\mathbf{\Gamma}'\mathbf{A}\mathbf{\Gamma} = \begin{pmatrix} \mathbf{G}_t & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}, \qquad \mathbf{\Gamma}'\mathbf{B}\mathbf{\Gamma} = \begin{pmatrix} \mathbf{H}_t & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}.$$

Since $\mathbf{A}^2 = \mathbf{A}$, we have

$$\begin{aligned}
\mathbf{\Gamma}'\mathbf{A}\mathbf{\Gamma}\mathbf{\Gamma}'\mathbf{A}\mathbf{\Gamma} = \mathbf{\Gamma}'\mathbf{A}\mathbf{\Gamma} &= \begin{pmatrix} \mathbf{G}_t & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \\
\implies \mathbf{\Gamma}'\mathbf{A}\mathbf{\Gamma}\begin{pmatrix} \mathbf{I}_t & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} &= \mathbf{\Gamma}'\mathbf{A}\mathbf{\Gamma} + \mathbf{\Gamma}'\mathbf{A}\mathbf{\Gamma}\mathbf{\Gamma}'\mathbf{B}\mathbf{\Gamma} \\
\implies \begin{pmatrix} \mathbf{G}_t & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}\begin{pmatrix} \mathbf{I}_t & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} &= \begin{pmatrix} \mathbf{G}_t & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} + \begin{pmatrix} \mathbf{G}_t & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}\begin{pmatrix} \mathbf{H}_t & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{G}_t & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} + \begin{pmatrix} \mathbf{G}_r\mathbf{H}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \\
\implies \begin{pmatrix} \mathbf{G}_r\mathbf{H}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} &= \mathbf{\Gamma}'\mathbf{A}\mathbf{\Gamma}\mathbf{\Gamma}'\mathbf{B}\mathbf{\Gamma} \\
&= \mathbf{\Gamma}'\mathbf{A}\mathbf{B}\mathbf{\Gamma} = \mathbf{O}.
\end{aligned}$$

Therefore, $\mathbf{AB} = \mathbf{O}$ since $\mathbf{\Gamma}$ is orthogonal and invertible.

THEOREM 6.1: Cochran

Let $\vec{X} \sim MN(\vec{0}, \mathbf{I}_n)$, $\mathbf{A}_1, \ldots, \mathbf{A}_m$ be symmetric $n \times n$ matrices with $\operatorname{rank}(\mathbf{A}_i) = r_i$, and $\sum_{i=1}^m \mathbf{A}_i = \mathbf{I}_n$. $\vec{X}'\mathbf{A}_i\vec{X} \sim \chi^2(r_i)$ are independent if and only if $\sum_{i=1}^m r_i = n$.

Proof: ($\Longleftarrow$) "Sufficiency" Assume $\sum_{i=1}^m r_i = n$. For each $i = 1, \ldots, m$, set

$$\mathbf{B}_i = \mathbf{I} - \mathbf{A}_i,$$

with $\operatorname{rank}(\mathbf{B}_i) = s_i$. We claim that $s_i = n - r_i$.

$$\begin{aligned}
s_i &= \operatorname{rank}(\mathbf{I} - \mathbf{A}_i) \\
&= \operatorname{rank}\left(\sum_{j \neq i} \mathbf{A}_j\right) \\
&\leq \sum_{j \neq i} \operatorname{rank}(\mathbf{A}_j) \\
&= n - r_i.
\end{aligned}$$

By definition, $\mathbf{I} = \mathbf{A}_i + \mathbf{B}_i \implies \operatorname{rank}(\mathbf{I}) = n$. So,

$$\begin{aligned}\operatorname{rank}(\mathbf{I}) &= n \\ &= \operatorname{rank}(\mathbf{A}_i + \mathbf{B}_i) \\ &\leq \operatorname{rank}(\mathbf{A}_i) + \operatorname{rank}(\mathbf{B}_i).\end{aligned}$$

Therefore, $\operatorname{rank}(\mathbf{B}_i) \geq n - r_i \implies s_i = n - r_i$ for all i. Hence,

$$|\lambda \mathbf{I} - \mathbf{B}_i| = 0$$

have r_i roots being 0. Noting that

$$\begin{aligned}|\lambda \mathbf{I} - \mathbf{B}_i| &= |(\lambda - 1)\mathbf{I} - \mathbf{A}_i| \\ &= |\tilde{\lambda}\mathbf{I} - \mathbf{A}_i|\end{aligned}$$

have r_i roots being 1. Since $\operatorname{rank}(\mathbf{A}_i) = r_i$, it follows that all other roots of $\mathbf{A}_i$ are 0. Hence,

$$\mathbf{A}_i = \mathbf{A}_i \mathbf{\Sigma} = \mathbf{A}_i \mathbf{I}$$

is idempotent by Lemma 4.1. Write

$$\mathbf{I} = \mathbf{A}_1 + (\mathbf{A}_2 + \cdots + \mathbf{A}_m).$$

Since $\mathbf{I}^2 = \mathbf{I}$, $\mathbf{A}_1^2 = \mathbf{A}_1$, $\mathbf{A}_2 + \cdots + \mathbf{A}_m$ is positive semidefinite, it follows from Lemma 6.1 that

$$\mathbf{A}_1(\mathbf{A}_2 + \cdots + \mathbf{A}_m) = \mathbf{O}.$$

This implies that

$$\mathbf{I}^2 = \mathbf{A}_1 + (\mathbf{A}_2 + \cdots \mathbf{A}_m),$$

which implies that $\mathbf{A}_2 + \cdots + \mathbf{A}_m$ is idempotent. Applying Lemma 6.1 to

$$\mathbf{A}_2 + \cdots + \mathbf{A}_m = \mathbf{A}_2 + (\mathbf{A}_3 + \cdots + \mathbf{A}_m)$$

it follows that

$$\mathbf{A}_2(\mathbf{A}_3 + \cdots + \mathbf{A}_m) = \mathbf{O}.$$

By induction, we get

$$\mathbf{A}_{m-1}\mathbf{A}_m = \mathbf{O}.$$

By re-labeling, we get

$$\mathbf{A}_i\mathbf{A}_j = 0 \ \forall i \neq j.$$

Since

$$\mathbf{A}_i \mathbf{\Sigma} \mathbf{A}_j = \mathbf{A}_i \mathbf{A}_j = \mathbf{O},$$

it follows from Theorem 5.2 that

$$\vec{X}'\mathbf{A}_i\vec{X} \text{ and } \vec{X}'\mathbf{A}_j\vec{X}$$

are independent. The fact that

$$\vec{X}'\mathbf{A}_i\vec{X} \sim \chi^2(r_i)$$

follows from Theorem 4.1.

LECTURE 7
30th January

7 Lecture 7: Full Rank Regression

Model:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon,$$

where x_i are **predictors**, Y is the **response**, and ε is noise. If we have $i = 1, \ldots, n$ observations, then the model becomes:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i = \mathbf{X}\vec{\beta} + \vec{\varepsilon},$$

where

$$\vec{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}, \quad \vec{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Assumptions:

(1) $\mathbb{E}[\varepsilon_i] = 0$.

(2) $\mathrm{Var}(\varepsilon_i) = \sigma^2$.

(3) $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$.

Full rank assumption:

(1) $k < n$;

(2) $\mathrm{rank}(\mathbf{X}) = k + 1$.

Method 1: Least Squares Method

$$L = \sum_{i=1}^{n} \varepsilon_i^2 = \vec{\varepsilon}'\vec{\varepsilon} = (\vec{Y} - \mathbf{X}\vec{\beta})'(\vec{Y} - \mathbf{X}\vec{\beta}).$$

If we minimize with respect to $\vec{\beta}$, we get

$$\frac{\partial L}{\partial \vec{\beta}} = \begin{pmatrix} \frac{\partial L}{\partial \beta_0} \\ \vdots \\ \frac{\partial L}{\partial \beta_k} \end{pmatrix} = -2\mathbf{X}\vec{Y} + 2\mathbf{X}'\mathbf{X}\vec{\beta} = 0 \implies \hat{\vec{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\vec{Y}.$$

THEOREM 7.1

$\hat{\vec{\beta}}$ is an unbiased estimator of $\vec{\beta}$.

Proof:

$$\begin{aligned} \mathbb{E}[\hat{\vec{\beta}}] &= \mathbb{E}\Big[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\vec{Y}\Big] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathbb{E}[\vec{Y}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathbb{E}\Big[\mathbf{X}\vec{\beta} + \vec{\varepsilon}\Big] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\vec{\beta} + \vec{0} \\ &= \vec{\beta}. \end{aligned}$$

THEOREM 7.2

If $\mathrm{Cov}(\vec{Y}, \vec{Y}) = \sigma^2 \mathbf{I}$*, then* $\mathrm{Cov}(\hat{\vec{\beta}}, \hat{\vec{\beta}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$.

Proof:

$$
\begin{aligned}
\mathrm{Cov}(\hat{\vec{\beta}}, \hat{\vec{\beta}}) &= \mathbb{E}\Big[\big(\hat{\vec{\beta}} - \mathbb{E}[\hat{\vec{\beta}}]\big)\big(\hat{\vec{\beta}} - \mathbb{E}[\hat{\vec{\beta}}]\big)'\Big] \\
&= \mathbb{E}\Big[\big((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\vec{Y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathbb{E}[\vec{Y}]\big)\big((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\vec{Y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathbb{E}[\vec{Y}]\big)'\Big] \\
&= \mathbb{E}\Big[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\big(\vec{Y} - \mathbb{E}[\vec{Y}]\big)(\vec{Y} - \mathbb{E}[\vec{Y}])'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\Big] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathbb{E}\Big[(\vec{Y} - \mathbb{E}[\vec{Y}])(\vec{Y} - \mathbb{E}[\vec{Y}])'\Big]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.
\end{aligned}
$$

Estimation of σ^2

- Residual:

$$
\begin{aligned}
(\vec{Y} - \hat{\vec{Y}}) &= \hat{\vec{\varepsilon}} \\
&= (\vec{Y} - \mathbf{X}\hat{\vec{\beta}}) \\
&= (\vec{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\vec{Y}) \\
&= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\vec{Y}.
\end{aligned}
$$

- Let $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ be the **hat matrix**.
- Since $\mathbf{H}$ is idempotent, we may write

$$
\begin{aligned}
\mathrm{SSE} &= \hat{\vec{\varepsilon}}'\hat{\vec{\varepsilon}} \\
&= \|\vec{Y} - \hat{\vec{Y}}\|^2 \\
&= (\vec{Y} - \hat{\vec{Y}})'(\vec{Y} - \hat{\vec{Y}}) \\
&= \vec{Y}'(\mathbf{I} - \mathbf{H})\vec{Y}.
\end{aligned}
$$

THEOREM 7.3

$$
S^2 = \frac{SSE}{n - (k+1)}
$$

is an unbiased estimator of σ^2.

Proof:

$$\begin{aligned}
\mathbb{E}[S^2] &= \frac{1}{n-(k+1)}\,\mathbb{E}[\vec{Y}'(\mathbf{I}-\mathbf{H})\vec{Y}] \\
&= \frac{1}{n-(k+1)}\left[\operatorname{tr}((\mathbf{I}-\mathbf{H})\mathbf{\Sigma}) + \vec{\mu}'(\mathbf{I}-\mathbf{H})\vec{\mu}\right] \\
&= \frac{\sigma^2}{n-(k+1)}\operatorname{tr}(\mathbf{I}-\mathbf{H}) \\
&= \frac{\sigma^2}{n-(k+1)}\big(\operatorname{tr}(\mathbf{I}) - \operatorname{tr}(\mathbf{H})\big) \\
&= \frac{\sigma^2}{n-(k+1)}\big(\operatorname{tr}(\mathbf{I}) - \operatorname{rank}(\mathbf{H})\big) \\
&= \frac{\sigma^2}{n-(k+1)}\big(n-(k+1)\big) \\
&= \sigma^2.
\end{aligned}$$

Note that $\vec{\mu} = \mathbf{X}\vec{\beta} \implies \vec{\mu}' = \vec{\beta}'\mathbf{X}'$, so

$$\begin{aligned}
\mathbf{X}'(\mathbf{I}-\mathbf{H})\mathbf{X} &= \mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{H}\mathbf{X} \\
&= \mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \\
&= \mathbf{O}.
\end{aligned}$$

Maximum Likelihood Estimators for $\vec{\beta}$ and σ^2

THEOREM 7.4

If $\vec{Y} \sim MN(\mathbf{X}\vec{\beta}, \sigma^2\mathbf{I}_n)$, where $\mathbf{X}$ is $n \times (k+1)$ of rank $k+1 < n$, then the maximum likelihood estimators of $\vec{\beta}$ and σ^2 are

$$\hat{\vec{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\vec{Y}, \qquad \hat{\sigma}^2 = \frac{SSE}{n} = \frac{(\vec{Y}-\mathbf{X}\hat{\vec{\beta}})'(\vec{Y}-\mathbf{X}\hat{\vec{\beta}})}{n}.$$

Proof: The likelihood function is given by the multivariate normal density

$$\begin{aligned}
L(\vec{\beta}, \sigma^2) &= f(\vec{Y}; \vec{\beta}, \sigma^2) \\
&= \frac{1}{(2\pi)^{n/2}|\sigma^2\mathbf{I}|^{1/2}}\exp\left\{-\frac{(\vec{Y}-\mathbf{X}\vec{\beta})'(\sigma^2\mathbf{I})^{-1}(\vec{Y}-\mathbf{X}\vec{\beta})}{2}\right\} \\
&= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{(\vec{Y}-\mathbf{X}\vec{\beta})'(\vec{Y}-\mathbf{X}\vec{\beta})}{2\sigma^2}\right\}.
\end{aligned}$$

The log-likelihood function is

$$\begin{aligned}
\ell(\vec{\beta}, \sigma^2) &= \ln\Big(L(\vec{\beta}, \sigma^2)\Big) \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(\vec{Y}-\mathbf{X}\vec{\beta})'(\vec{Y}-\mathbf{X}\vec{\beta}) \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(\vec{Y}'\vec{Y} - 2\vec{Y}'\mathbf{X}\vec{\beta} + \vec{\beta}'\mathbf{X}'\mathbf{X}\vec{\beta}).
\end{aligned}$$

Taking the derivative with respect to $\vec{\beta}$ yields

$$\begin{aligned}\vec{0} &= \frac{\partial \ell(\vec{\beta}, \sigma^2)}{\partial \vec{\beta}} = -\frac{1}{2\sigma^2}(-2\mathbf{X}'\vec{Y} + 2\mathbf{X}'\mathbf{X}\vec{\beta}) \\ \vec{0} &= 2\mathbf{X}'\vec{Y} - 2\mathbf{X}'\mathbf{X}\vec{\beta} \\ \mathbf{X}'\mathbf{X}\vec{\beta} &= \mathbf{X}'\vec{Y} \\ \hat{\vec{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\vec{Y}.\end{aligned}$$

Taking the derivative with respect to σ^2 yields

$$\begin{aligned}0 &= \frac{\partial \ell(\vec{\beta}, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\vec{Y} - \mathbf{X}\vec{\beta})'(\vec{Y} - \mathbf{X}\vec{\beta}) \\ \frac{n(2\sigma^4)}{2\sigma^2} &= (\vec{Y} - \mathbf{X}\vec{\beta})'(\vec{Y} - \mathbf{X}\vec{\beta}) \\ \hat{\sigma}^2 &= \frac{(\vec{Y} - \mathbf{X}\hat{\vec{\beta}})'(\vec{Y} - \mathbf{X}\hat{\vec{\beta}})}{n}.\end{aligned}$$

Properties of $\hat{\vec{\beta}}$ and $\hat{\sigma}^2$

THEOREM 7.5

If $\vec{Y} \sim MN(\mathbf{X}\vec{\beta}, \sigma^2\mathbf{I}_n)$, where $\mathbf{X}$ is $n \times (k+1)$ of rank $k+1 < n$, and $\vec{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$, then the maximum likelihood estimators of $\hat{\vec{\beta}}$ and $\hat{\sigma}^2$ given in Theorem 7.1 have the following distributional properties:

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

(2) $\dfrac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2\big(n - (k+1)\big)$.

(3) $\hat{\vec{\beta}}$ *and* $\hat{\sigma}^2$ *are independent.*

Proof:

(1) Note that $\vec{Y} \sim \text{MN}(\vec{\mu}, \mathbf{\Sigma}) \implies \mathbf{A}\vec{Y} \sim \text{MN}(\mathbf{A}\vec{\mu}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}')$. Let $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\vec{\mu} = \mathbf{X}\vec{\beta}$, and $\mathbf{\Sigma} = \sigma^2\mathbf{I}_n$. Now,

$$\begin{aligned}\mathbf{A}\vec{Y} = \hat{\vec{\beta}} &\sim \text{MN}\Big((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\vec{\beta}, (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}_n\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\Big) \\ &\sim \text{MN}(\vec{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}).\end{aligned}$$

(2) Note that

$$\begin{aligned}\frac{n\hat{\sigma}^2}{\sigma^2} &= \frac{\text{SSE}}{\sigma^2} \\ &= \frac{\vec{Y}'(\mathbf{I} - \mathbf{H})\vec{Y}}{\sigma^2} \\ &= \vec{W}'(\mathbf{I} - \mathbf{H})\vec{W},\end{aligned}$$

where $\vec{W} = \frac{\vec{Y}}{\sigma} \sim \text{MN}\Big(\frac{\mathbf{X}\vec{\beta}}{\sigma}, \mathbf{I}\Big)$. It follows from Theorem 4.1 that

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(r, \lambda),$$

with $r = \operatorname{rank}(\mathbf{I} - \mathbf{H}) = \operatorname{tr}(\mathbf{I} - \mathbf{H}) = n - (k+1)$ and

$$\begin{aligned}
\lambda &= \frac{1}{2}\vec{\mu}'\mathbf{A}\vec{\mu} \\
&= \frac{1}{2}\left(\frac{\mathbf{X}\vec{\beta}}{\sigma}\right)'(\mathbf{I} - \mathbf{H})\frac{\mathbf{X}\vec{\beta}}{\sigma} \\
&= \frac{1}{2\sigma^2}\vec{\beta}'\left[\mathbf{X}'(\mathbf{I} - \mathbf{H})\mathbf{X}\right]\vec{\beta} \\
&= 0,
\end{aligned}$$

where $\vec{\mu} = \mathbb{E}[\vec{W}]$ and $\mathbf{A} = \mathbf{I} - \mathbf{H}$.

(3) Note that $\hat{\vec{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\vec{Y}$ and

$$\hat{\sigma}^2 = \frac{\text{SSE}}{n} = \frac{1}{n}\vec{Y}'(\mathbf{I} - \mathbf{H})\vec{Y} = \vec{Y}'\left(\frac{\mathbf{I} - \mathbf{H}}{n}\right)\vec{Y}.$$

Let $\vec{Y} = \vec{X}$, $\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}$, and $\mathbf{A} = \frac{\mathbf{I}-\mathbf{H}}{n}$. Relabelling,

$$\hat{\vec{\beta}} = \mathbf{B}\vec{X},$$

$$\hat{\sigma}^2 = \vec{X}'\mathbf{A}\vec{X}.$$

Now,

$$\begin{aligned}
(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\left(\frac{\mathbf{I} - \mathbf{H}}{n}\right) &= \frac{1}{n}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] \\
&= 0.
\end{aligned}$$

The result follows from Theorem 5.1.

LECTURE 8
31st January

8 Lecture 8: Test of Overall Regression

DEFINITION 8.1: Sum of Squares Total, Residual, Error

$$\begin{aligned}
\bar{Y} &= \frac{1}{n}\sum_{i=1}^{n} Y_i, \\
\text{SST} &= \sum_{i=1}^{n}(Y_i - \bar{Y})^2, \\
\text{SSR} &= \sum_{i=1}^{n}(\hat{Y} - \bar{Y})^2, \\
\text{SSE} &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.
\end{aligned}$$

SST is the **sum of squares total**, SSR is the **sum of squares residual**, and SSE is the **sum of squares error**.

THEOREM 8.1

(i) $SST = \vec{Y}'(\mathbf{I} - \frac{1}{n}\mathbf{J})\vec{Y}$.

(ii) $SSR = \vec{Y}'(\mathbf{H} - \frac{1}{n}\mathbf{J})\vec{Y}$.

(iii) $SSE = \vec{Y}'(\mathbf{I} - \mathbf{H})\vec{Y}$.

Hence, $SST = SSR + SSE$.

Proof:

(i) Sum of Squares Total:

$$\begin{aligned}
\text{SST} &= \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \\
&= (\vec{Y} - \bar{Y}\vec{j})'(\vec{Y} - \bar{Y}\vec{j}) \\
&= \vec{Y}'\vec{Y} - 2\bar{Y}\vec{Y}'\vec{j} + \bar{Y}^2\vec{j}'\vec{j} \\
&= \vec{Y}'\vec{Y} - 2n\bar{Y}^2 + n\bar{Y}^2 \\
&= \vec{Y}'\vec{Y} - n\bar{Y}^2 \\
&= \vec{Y}'\vec{Y} - \tfrac{1}{n}\vec{Y}'\vec{j}\vec{j}'\vec{Y} \\
&= \vec{Y}'(\mathbf{I} - \tfrac{1}{n}\mathbf{J})\vec{Y}.
\end{aligned}$$

(ii) Sum of Squares Regression:

$$\begin{aligned}
\text{SSR} &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \\
&= (\hat{\vec{Y}} - \bar{Y}\vec{j})'(\hat{\vec{Y}} - \bar{Y}\vec{j}) \\
&= (\mathbf{H}\vec{Y} - \bar{Y}\vec{j})'(\mathbf{H}\vec{Y} - \bar{Y}\vec{j}) \\
&= \vec{Y}'\mathbf{H}\mathbf{H}\vec{Y} - 2\bar{Y}\vec{Y}'\mathbf{H}\vec{j} + \bar{Y}^2\vec{j}'\vec{j} \\
&= \vec{Y}'\mathbf{H}^2\vec{Y} - \tfrac{2}{n}\vec{Y}'\mathbf{H}\vec{j}\vec{j}'\vec{Y} + \tfrac{1}{n}\vec{Y}'\mathbf{J}\vec{Y} \\
&= \vec{Y}'\mathbf{H}\vec{Y} - \tfrac{2}{n}\vec{Y}'\mathbf{H}\mathbf{J}\vec{Y} + \tfrac{1}{n}\vec{Y}'\mathbf{J}\vec{Y} \\
&= \vec{Y}'\mathbf{H}\vec{Y} - \tfrac{2}{n}\vec{Y}'\mathbf{J}\vec{Y} + \tfrac{1}{n}\vec{Y}'\mathbf{J}\vec{Y} \\
&= \vec{Y}'\mathbf{H}\vec{Y} - \tfrac{1}{n}\vec{Y}'\mathbf{J}\vec{Y} \\
&= \vec{Y}'(\mathbf{H} - \tfrac{1}{n}\mathbf{J})\vec{Y}
\end{aligned}$$

since $\mathbf{H}^2 = \mathbf{H}$, $\mathbf{HX} = \mathbf{X}$, and $\mathbf{H}\vec{j} = \vec{j} \implies \mathbf{HJ} = \mathbf{J}$.

(iii) Sum of Squares Error:

$$\begin{aligned}
\text{SSE} &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i) \\
&= (\vec{Y} - \hat{\vec{Y}})'(\vec{Y} - \hat{\vec{Y}}) \\
&= (\vec{Y} - \mathbf{H}\vec{Y})'(\vec{Y} - \mathbf{H}\vec{Y}) \\
&= \big((\mathbf{I} - \mathbf{H})\vec{Y}\big)'(\mathbf{I} - \mathbf{H})\vec{Y} \\
&= \vec{Y}'(\mathbf{I} - \mathbf{H})^2\vec{Y} \\
&= \vec{Y}'(\mathbf{I} - \mathbf{H})\vec{Y}
\end{aligned}$$

since $(\mathbf{I}-\mathbf{H})^2 = \mathbf{I}-\mathbf{H}$.

Therefore,

$$\mathbf{I}-\mathbf{H}+\mathbf{H}-\tfrac{1}{n}\mathbf{J} = \mathbf{I}-\tfrac{1}{n}\mathbf{J} \implies \text{SST} = \text{SSR} + \text{SSE}.$$

THEOREM 8.2

(1) $SSR/\sigma^2 \sim \chi^2(k,\lambda)$ with

$$\begin{aligned}\lambda &= \frac{1}{2\sigma^2}(\mathbf{X}\vec{\beta})'(\mathbf{H}-\tfrac{1}{n}\mathbf{J})\mathbf{X}\vec{\beta}\\ &= \frac{1}{2\sigma^2}(\mathbf{X}_1\vec{\beta}_1)'(\mathbf{H}-\tfrac{1}{n}\mathbf{J})\mathbf{X}_1\vec{\beta}_1,\end{aligned}$$

where

$$\vec{\beta}_1 = \begin{pmatrix}\beta_1\\ \vdots \\ \beta_k\end{pmatrix}, \qquad \mathbf{X}_1 = \begin{pmatrix}X_{11} & \cdots & X_{1k}\\ \vdots & \ddots & \vdots\\ X_{n1} & \cdots & X_{nk}\end{pmatrix}.$$

(2) SSR and SSE are independent.

Proof:

(1) First result:

$$\frac{\text{SSR}}{\sigma^2} = \frac{\vec{Y}'}{\sigma}(\mathbf{H}-\tfrac{1}{n}\mathbf{J})\frac{\vec{Y}}{\sigma}$$

with

$$\frac{\vec{Y}}{\sigma} \sim \text{MN}\left(\frac{\mathbf{X}\vec{\beta}}{\sigma}, \mathbf{I}\right).$$

Since

$$\begin{aligned}(\mathbf{H}-\tfrac{1}{n}\mathbf{J})^2 &= \mathbf{H}^2 - 2\tfrac{1}{n}\mathbf{H}\mathbf{J} + \frac{1}{n^2}\mathbf{J}^2\\ &= \mathbf{H} - 2\tfrac{1}{n}\mathbf{J} + \tfrac{1}{n}\mathbf{J}\\ &= \mathbf{H} - \tfrac{1}{n}J\end{aligned}$$

it follows from Theorem 4.1 that $\text{SSR}/\sigma^2 \sim \chi^2(r,\lambda)$ with

$$r = \text{rank}(\mathbf{H}-\tfrac{1}{n}\mathbf{J}) = \text{tr}(\mathbf{H}-\tfrac{1}{n}\mathbf{J}) = k+1-1 = k$$

and

$$\lambda = \frac{1}{n}\left(\frac{\mathbf{X}\vec{\beta}}{\sigma}\right)'(\mathbf{H}-\tfrac{1}{n}\mathbf{J})\frac{\mathbf{X}\vec{\beta}}{\sigma}.$$

Write

$$\vec{\beta} = \begin{pmatrix}\beta_0\\ \vec{\beta}_1\end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix}\vec{j} & \mathbf{X}_1\end{pmatrix}.$$

Then, $\mathbf{X}\vec{\beta} = \beta_0\vec{j} + \mathbf{X}_1\vec{\beta}_1$ implies

$$\begin{aligned}\lambda &= \frac{1}{2}\left(\frac{\beta_0\vec{j}+\mathbf{X}_1\vec{\beta}_1}{\sigma}\right)'(\mathbf{H}-\tfrac{1}{n}\mathbf{J})\frac{\beta_0\vec{j}+\mathbf{X}_1\vec{\beta}_1}{\sigma}\\ &= \frac{1}{2\sigma^2}\left[(\mathbf{X}_1\vec{\beta}_1)(\mathbf{H}-\tfrac{1}{n}\mathbf{J})(\mathbf{X}_1\vec{\beta}_1) + \beta_0^2\vec{j}'(\mathbf{H}-\tfrac{1}{n}\mathbf{J})\vec{j} + 2\beta_0(\mathbf{H}-\tfrac{1}{n}\mathbf{J})\vec{j}\right]\\ &= \frac{1}{2\sigma^2}(\mathbf{X}_1\vec{\beta}_1)'(\mathbf{H}-\tfrac{1}{n}\mathbf{J})\mathbf{X}_1\vec{\beta}_1\end{aligned}$$

since

$$(\mathbf{H} - \tfrac{1}{n}\mathbf{J})\vec{j} = \mathbf{H}\vec{j} - \tfrac{1}{n}\mathbf{J}\vec{j} = \vec{j} - \tfrac{1}{n}n\vec{j} = 0.$$

(2) Note that

$$\begin{aligned} \text{SSR} &= \vec{Y}'(\mathbf{H} - \tfrac{1}{n}\mathbf{J})\vec{Y} = \vec{Y}'\mathbf{A}\vec{Y}. \\ \text{SSE} &= \vec{Y}'(\mathbf{I} - \mathbf{H})\vec{Y} = \vec{Y}'\mathbf{B}\vec{Y}. \end{aligned}$$

Note that $\vec{Y} \sim \text{MN}(\mathbf{X}\vec{\beta}, \sigma^2\mathbf{I})$. Note that $\mathbf{H} - \frac{1}{n}\mathbf{J}$ and $\mathbf{I} - \mathbf{H}$ are symmetric matrices of the same dimension.

$$\begin{aligned} \mathbf{A}\sigma^2\mathbf{I}\mathbf{B} &= \sigma^2\mathbf{A}\mathbf{B} \\ &= \sigma^2(\mathbf{H} - \tfrac{1}{n}\mathbf{J})(\mathbf{I} - \mathbf{H}) \\ &= \sigma^2(\mathbf{H} - \tfrac{1}{n}\mathbf{J} - \mathbf{H}^2 + \tfrac{1}{n}\mathbf{J}\mathbf{H}) \\ &= \sigma^2(\mathbf{H} - \tfrac{1}{n}\mathbf{J} - \mathbf{H} + \tfrac{1}{n}\mathbf{J}) \\ &= \sigma^2\mathbf{O} \\ &= \mathbf{O}. \end{aligned}$$

Therefore, SSR and SSE are independent by Theorem 5.2.

REMARK

By Theorem 7.4, we have $\text{SSE}/\sigma^2 \sim \chi^2(n - (k+1))$, and $\text{SST}/\sigma^2 \sim \chi^2(n-1, \lambda)$.

REMARK

$$\begin{aligned} \mathbb{E}[\text{SSR}] &= \sigma^2\,\mathbb{E}\left[\frac{\text{SSR}}{\sigma^2}\right] \\ &= \sigma^2 \sum_{i=1}^{k} \mathbb{E}[X_i^2] && \text{where } X_i \sim \mathcal{N}(\mu_i, 1) \\ &= \sigma^2\left(\sum_{i=1}^{n} \big(\text{Var}(X_i) + \mu_i^2\big)\right) \\ &= \sigma^2(k + 2\lambda) \\ &= k\sigma^2 + 2\sigma^2\lambda \\ &= k\sigma^2 + (\mathbf{X}_1\vec{\beta}_1)'(\mathbf{H} - \tfrac{1}{n}\mathbf{J})\mathbf{X}_1\vec{\beta}_1. \end{aligned}$$

ANOVA Table for Hypothesis Test of H_0: $\vec{\beta}_1 = \vec{0}$ versus H_A: $\vec{\beta}_1 \neq \vec{0}$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Expected Mean Square
Due to $\vec{\beta}_1$	k	SSR	MSR	$\sigma^2 + \frac{1}{k}(\mathbf{X}_1\vec{\beta}_1)'(\mathbf{H} - \frac{1}{n}\mathbf{J})\vec{X}_1\vec{\beta}_1$
Error	$n - (k+1)$	SSE	MSE	σ^2
Total	$n - 1$	SST		

Note that under H_0, $\frac{1}{k}(\mathbf{X}_1\vec{\beta}_1)'(\mathbf{H}-\frac{1}{n}\mathbf{J})\vec{X}_1\vec{\beta}_1 = 0$.

$$\begin{aligned}
\mathbb{E}[\text{MSR}] &= \mathbb{E}\left[\frac{\text{SSR}}{k}\right] \\
&= \frac{1}{k}\left[\sigma^2 k + (\mathbf{X}_1\vec{\beta}_1)(\mathbf{H}-\tfrac{1}{n}\mathbf{J})\vec{X}_1\vec{\beta}_1\right] \\
&= \sigma^2 + \frac{1}{k}(\mathbf{X}_1\vec{\beta}_1)'(\mathbf{H}-\tfrac{1}{n}\mathbf{J})\vec{X}_1\vec{\beta}_1. \\
\mathbb{E}[\text{MSE}] &= \mathbb{E}\left[\frac{\text{SSE}}{n-(k+1)}\right] \\
&= \frac{(n-(k+1))\sigma^2}{n-(k+1)} \\
&= \sigma^2.
\end{aligned}$$

Test Statistic:

$$F = \frac{\text{SSR}/k}{\text{SSE}/(n-(k+1))} \sim F(k, n-(k+1)).$$

Reject H_0 if $F > F_\alpha(k, n-(k+1))$. If H_0 holds, then $\lambda = 0$ implies SSR $\sim \chi^2(k)$ and SSE $\sim \chi^2(n-(k+1))$. Furthermore, note that if X and Y are independent, then $f(X)$ and $g(Y)$ are independent, so

$$\begin{aligned}
\mathbb{E}[F] &= \mathbb{E}\left[\frac{\text{SSR}/k}{\text{SSE}/(n-(k+1))}\right] \\
&= \frac{n-(k+1)}{k}\,\mathbb{E}[\text{SSR}]\,\mathbb{E}\left[\frac{1}{\text{SSE}}\right].
\end{aligned}$$

$$\mathbb{E}\left[\frac{\text{SSR}}{k}\right] = \sigma^2 \iff \mathbb{E}[\text{SSR}] = k\sigma^2.$$

Hold tight for the hard part (you can skip this part if you know the mean of the *inverse-chi-squared distribution*),

$$\begin{aligned}
\mathbb{E}\left[\frac{\sigma^2}{\text{SSE}}\right] &= \int_0^\infty \frac{1}{y}\frac{1}{2^{(n-k-1)/2}\Gamma\left(\frac{n-k-1}{2}\right)} y^{(n-k-1)/2-1} e^{-y/2}\,\mathrm{d}y \\
&= 2^{-(n-k-1)/2}\frac{\Gamma\left(\frac{n-k-1}{2}-1\right)}{\Gamma\left(\frac{n-k-1}{2}\right)} \times \\
&\qquad 2^{(n-k-1)/2-1}\underbrace{\frac{1}{\Gamma\left(\frac{n-k-1}{2}-1\right)2^{(n-k-1)/2-1}}\int_0^\infty y^{((n-k-1)/2-1)-1}e^{-y/2}\,\mathrm{d}y}_{=1 \text{ by Gamma distribution}} \\
&= \frac{1}{2}\frac{1}{(n-k-1)/2-1} \\
&= \frac{1}{n-k-3}.
\end{aligned}$$

Hence,

$$\mathbb{E}\left[\frac{1}{\text{SSE}}\right] = \frac{1}{\sigma^2(n-k-3)}.$$

Therefore,

$$\mathbb{E}[F] = \frac{n-k-1}{k}\,\mathbb{E}[\text{SSR}]\,\mathbb{E}\left[\frac{1}{\text{SSE}}\right] = \frac{n-k-1}{k}k\sigma^2\frac{1}{\sigma^2(n-k-3)} = \frac{n-k-1}{n-k-3}.$$

LECTURE 9
6th February

9 Lecture 9: Lack of Fit

Consider the case of studying blood pressure and its relationship to height and weight. Clearly, people of the same height and weight can have different blood pressures. In other words, the same predictor values may correspond to different response values. This type of variation is called pure error. To detect poor model fit, we would need to distinguish between variation caused by the model and pure error.

General Framework

Let $m \geq 1$ and $n_1, \ldots, n_m \geq 1$ such that $\sum_{i=1}^m n_i = n$. For $i = 1, \ldots, n$, we have

$$Y_{ir} = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ik} + \varepsilon_{ir},\ r = 1, \ldots, n_i.$$

In matrix notation, we write $\mathbf{Y} = \mathbf{X}\vec{\beta}$, where $\vec{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}$ in the usual way, and

$$\vec{Y}' = \begin{pmatrix} Y_{11} & \cdots & Y_{1n_1} \cdots & Y_{m1} & \cdots & Y_{mn_m} \end{pmatrix},$$

$$\mathbf{X} = \begin{pmatrix} n_1 \left\{ \begin{matrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \ddots & \vdots \\ 1 & x_{11} & \cdots & x_{1k} \end{matrix} \right\} \\ \vdots \\ n_m \left\{ \begin{matrix} 1 & x_{m1} & \cdots & x_{mk} \\ \vdots & \ddots & \ddots & \vdots \\ 1 & x_{m1} & \cdots & x_{mk} \end{matrix} \right\} \end{pmatrix},$$

$$\vec{\varepsilon}' = \begin{pmatrix} \varepsilon_{11} & \cdots & \varepsilon_{1n_1} & \cdots & \varepsilon_{m1} & \cdots & \varepsilon_{mn_m} \end{pmatrix}.$$

We write Y_{ij} for $i = 1, \ldots, m$ (m groups) and $j = 1, \ldots, n_i$ (number of observations in group i). The sample average of group i is defined by

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij},\ i = 1, \ldots m.$$

The fitted values are $\hat{\vec{Y}} = \mathbf{X}\hat{\vec{\beta}}$, so $\hat{Y}_{ij}$ is the same for all $j = 1, \ldots, n_i$, hence we may write $\hat{Y}_{ij}$ as $\hat{Y}_i$.

$$\begin{aligned} \text{SSE} &= \sum_{i=1}^m \sum_{j=1}^{n_i} (Y_{ij} - \hat{Y}_{ij})^2 \\ &= \sum_{i=1}^m \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i + \bar{Y}_i - \hat{Y}_{ij})^2 \\ &= \sum_{i=1}^m \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^m \sum_{j=1}^{n_i} (\hat{Y}_{ij} - \bar{Y}_i)^2 - 2 \sum_{i=1}^m \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)(\bar{Y}_i - \hat{Y}_{ij}) \\ &= \sum_{i=1}^m \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^m \sum_{j=1}^{n_i} (\hat{Y}_{ij} - \bar{Y}_i)^2 \\ &= \sum_{i=1}^m \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^m n_i (\hat{Y}_i - \bar{Y}_i)^2 \\ &= \text{SSPE} + \text{SSLF} \end{aligned}$$

since $\hat{Y}_{ij}$ is independent of j. Therefore,

$$\text{SST} = \text{SSR} + \text{SSE} = \text{SSR} + \text{SSPE} + \text{SSLF}.$$

- Degrees of freedom of SSLF: $m-(k+1)$.
- Degrees of freedom of SSPE: $(n_1-1)+\cdots+(n_m-1)=n-m$.

The first test is a test of linear relationship, but if we wanted to determine how good that relationship is, we will need the following hypothesis test. If the linear model fits well, then SSLF should be small.

- $\mathrm{H_0}$: The model is adequate.
- $\mathrm{H_A}$: The model is not adequate.

Test statistic:

$$F=\frac{\text{SSLF}/(m-k-1)}{\text{SSPE}/(n-m)}\sim F(m-k-1,n-m).$$

If we reject $\mathrm{H_0}$, that means there's too much variation within the group. Reject $\mathrm{H_0}$ when $F>F_\alpha(m-k-1,n-m)$.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Due to $\vec{\beta}_1$	k	SSR	MSR	MSR/MSE
Error	$n-(k+1)$	SSE	MSE	
Lack of Fit	$m-k-1$	SSLF	MSLF	MSLF/MSPE
Pure Error	$n-m$	SSPE	MSPE	
Total	$n-1$	SST		

Selection of Predictors

We observe that the number of predictors always improves the estimates, but becomes less efficient. To find a reasonable number of predictors, one needs to compare models by adding or dropping predictors.

- Partition $\vec{\beta}$ as $\vec{\beta}=\begin{pmatrix}\beta_0\\ \vdots\\ \beta_k\end{pmatrix}=\begin{pmatrix}\vec{\beta}_I\\ \vec{\beta}_{II}\end{pmatrix}$, where $\vec{\beta}_I=\begin{pmatrix}\beta_0\\ \vdots\\ \beta_\ell\end{pmatrix}$ and $\vec{\beta}_{II}=\begin{pmatrix}\beta_{\ell+1}\\ \vdots\\ \beta_k\end{pmatrix}$ for $1\le\ell<k$.
- Partition $\mathbf{X}$ as $\mathbf{X}=\begin{pmatrix}\mathbf{X}_I & \mathbf{X}_{II}\end{pmatrix}$, where $\mathbf{X}_I\in\mathbb{R}^{n\times(\ell+1)}$ and $\mathbf{X}_{II}\in\mathbb{R}^{n\times(k-\ell)}$ for $1\le\ell<k$.

The **full model** is

$$\vec{Y}=\mathbf{X}\vec{\beta}+\vec{\varepsilon}=\begin{pmatrix}\mathbf{X}_I & \mathbf{X}_{II}\end{pmatrix}\begin{pmatrix}\vec{\beta}_I\\ \vec{\beta}_{II}\end{pmatrix}=\mathbf{X}_I\vec{\beta}_I+\mathbf{X}_{II}\vec{\beta}_{II}.$$

The **reduced model** is

$$\vec{Y}=\mathbf{X}_I\vec{\beta}_I+\vec{\varepsilon}^*.$$

Let

- $\mathbf{H}=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.
- $\mathbf{H}_1=\mathbf{X}_I(\mathbf{X}_I'\mathbf{X}_I)^{-1}\mathbf{X}_I'$.

Define

- $\text{SS}(\vec{\beta})=\text{SSR(full)}=\vec{Y}'(\mathbf{H}-\frac{1}{n}\mathbf{J})\vec{Y}$.
- $\text{SS}(\vec{\beta}_I)=\text{SSR(reduced)}=\vec{Y}'(\mathbf{H}_I-\frac{1}{n}\mathbf{J})\vec{Y}$.
- $\text{SS}(\vec{\beta}_{II}\mid\vec{\beta}_I)=\text{SS}(\vec{\beta})-\text{SS}(\vec{\beta}_I)=\vec{Y}'(\mathbf{H}-\mathbf{H}_I)\vec{Y}$.

Comparing the full model and the reduced model, we test $\mathrm{H_0}$: $\vec{\beta}_{II}=\vec{0}$ versus $\mathrm{H_A}$: $\vec{\beta}_{II}\neq\vec{0}$. Under $\mathrm{H_0}$, $x_{\ell+1},\ldots,x_k$ do not add predictive value to the model that includes $x_1,\ldots,x_\ell$ already.

THEOREM 9.1

$\mathbf{H} - \mathbf{H}_1$ *is idempotent.*

Proof: Assignment 2.

Model misspecification:

- Leaving out $\vec{\beta}_{II}$ when it should be included, results in underfitting.
- Including $\vec{\beta}_{II}$ when it should be dropped, results in overfitting.

LECTURE 10
9th February

10 Lecture 10: Determination of Predictors and Generalized Inverse

THEOREM 10.1

Let $\vec{Y} \sim MN(\mathbf{X}\vec{\beta}, \sigma^2\mathbf{I})$, $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{H}_1 = \mathbf{X}_I(\mathbf{X}_I'\mathbf{X}_I)^{-1}\mathbf{X}_I'$. *Then,*

(1) $\vec{Y}'(\mathbf{I} - \mathbf{H})\vec{Y}/\sigma^2 = SSE/\sigma^2 \sim \chi^2(n-k-1)$.

(2) $\vec{Y}'(\mathbf{H} - \mathbf{H}_1)\vec{Y}/\sigma^2 = \frac{SS(\vec{\beta}_{II}|\vec{\beta}_I)}{\sigma^2} \sim \chi^2(k-\ell, \tilde{\lambda})$, *where*

$$\tilde{\lambda} = \vec{\beta}_{II}'\mathbf{X}_{II}'(\mathbf{I} - \mathbf{H}_1)\mathbf{X}_{II}\vec{\beta}_{II}/2\sigma^2.$$

(3) $\vec{Y}'(\mathbf{I} - \mathbf{H})\vec{Y}$ *and* $\vec{Y}'(\mathbf{H} - \mathbf{H}_1)\vec{Y}$ *are independent.*

Proof:

(1) Earlier proof.

(2)

$$\begin{aligned}\vec{Y}'(\mathbf{H} - \mathbf{H}_1)\vec{Y}/\sigma^2 &= \frac{\text{SS(full} \mid \text{reduced)}}{\sigma^2} \\ &= \frac{\text{SS}(\vec{\beta}_{II} \mid \vec{\beta}_I)}{\sigma^2}.\end{aligned}$$

By Theorem 9.1, $\mathbf{H} - \mathbf{H}_1$ is idempotent, so

$$\vec{Y}'(\mathbf{H} - \mathbf{H}_1)\vec{Y} \sim \chi^2(r, \tilde{\lambda})$$

$r = \text{rank}(\mathbf{H} - \mathbf{H}_1)$, and

$$\tilde{\lambda} = \frac{1}{2\sigma^2}(\mathbf{X}\vec{\beta})'(\mathbf{H} - \mathbf{H}_1)\mathbf{X}\vec{\beta}.$$

By direct calculation, we have

$$\begin{aligned} r &= \text{rank}(\mathbf{H} - \mathbf{H}_1) \\ &= \text{tr}(\mathbf{H} - \mathbf{H}_1) && \text{since } \mathbf{H} - \mathbf{H}_1 \text{ is idempotent} \\ &= \text{tr}(\mathbf{H}) - \text{tr}(\mathbf{H}_1) \\ &= \text{tr}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') - \text{tr}(\mathbf{X}_I(\mathbf{X}_I'\mathbf{X}_I)^{-1}\mathbf{X}_I') \\ &= \text{tr}(\mathbf{I}_{k+1}) - \text{tr}(\mathbf{I}_{\ell+1}) && \text{cyclic property} \\ &= k - \ell. \end{aligned}$$

Noting that $\mathbf{H}\mathbf{X} = \mathbf{X}$, $\mathbf{H}_1\mathbf{X}_I = \mathbf{X}_I$, and

$$\mathbf{X}\vec{\beta} = \mathbf{X}_I\vec{\beta}_I + \mathbf{X}_{II}\vec{\beta}_{II}$$

it follows that

$$\begin{aligned}
2\sigma^2\tilde{\lambda} &= (\mathbf{X}\vec{\beta})'(\mathbf{H} - \mathbf{H}_1)\mathbf{X}\vec{\beta} \\
&= (\mathbf{X}\vec{\beta})'\mathbf{H}\mathbf{X}\vec{\beta} - (\mathbf{X}\vec{\beta})'\mathbf{H}_1\mathbf{X}\vec{\beta} \\
&= (\mathbf{X}\vec{\beta})'(\mathbf{X}\vec{\beta}) - (\mathbf{X}\vec{\beta})'(\mathbf{H}_1\mathbf{X}_I\vec{\beta}_I + \mathbf{H}_1\mathbf{X}_{II}\vec{\beta}_{II}) \\
&= (\mathbf{X}\vec{\beta})'\left[\mathbf{X}\vec{\beta} - \mathbf{H}_1\mathbf{X}_I\vec{\beta}_I - \mathbf{H}_1\mathbf{X}_{II}\vec{\beta}_{II}\right] \\
&= (\mathbf{X}\vec{\beta})'\left[\mathbf{X}_I\vec{\beta}_I + \mathbf{X}_{II}\vec{\beta}_{II} - \mathbf{X}_I\vec{\beta}_I - \mathbf{H}_1\mathbf{X}_{II}\vec{\beta}_{II}\right] \\
&= (\mathbf{X}\vec{\beta})'(\mathbf{X}_{II}\vec{\beta}_{II} - \mathbf{H}_1\mathbf{X}_{II}\vec{\beta}_{II}) \\
&= (\mathbf{X}\vec{\beta})'(\mathbf{I} - \mathbf{H}_1)\mathbf{X}_{II}\vec{\beta}_{II} \\
&= (\mathbf{X}_I\vec{\beta}_I + \mathbf{X}_{II}\vec{\beta}_{II})'(\mathbf{I} - \mathbf{H}_1)\mathbf{X}_{II}\vec{\beta}_{II} \\
&= (\mathbf{X}_{II}\vec{\beta}_{II})'(\mathbf{I} - \mathbf{H}_1)\vec{X}_{II}\vec{\beta}_{II} + (\mathbf{X}_I\vec{\beta}_I)(\mathbf{I} - \mathbf{H}_1)\mathbf{X}_{II}\vec{\beta}_{II}.
\end{aligned}$$

It remains to show that $(\mathbf{X}_I\vec{\beta}_I)(\mathbf{I} - \mathbf{H}_1)\mathbf{X}_{II}\vec{\beta}_{II} = \mathbf{O}$.

$$\begin{aligned}
(\mathbf{X}_I\vec{\beta}_I)'(\mathbf{I} - \mathbf{H}_1)\mathbf{X}_{II}\vec{\beta}_{II} &= (\mathbf{X}_I\vec{\beta}_I)'\mathbf{X}_{II}\vec{\beta}_{II} - \underbrace{(\mathbf{X}_I\vec{\beta}_I)'\mathbf{H}_1}_{\text{see below}}\mathbf{X}_{II}\vec{\beta}_{II} \\
&= (\mathbf{X}_I\vec{\beta}_I)'\mathbf{X}_{II}\vec{\beta}_{II} - (\mathbf{X}_I\vec{\beta}_I)'\mathbf{X}_{II}\vec{\beta}_{II} \\
&= \mathbf{O}
\end{aligned}$$

since $\left\{\left[(\mathbf{X}_I\vec{\beta}_I)'\mathbf{H}_1\right]'\right\}' = \{\mathbf{H}_1\mathbf{X}_I\vec{\beta}_I\}' = (\mathbf{X}_I\vec{\beta}_I)'$.

(3) A2Q5.

ANOVA for Model Selection

H_0: $\vec{\beta}_{II} = \vec{0}$ versus H_A: $\vec{\beta}_{II} \neq \vec{0}$.

Source	df	SS	MS	Statistics
Due to $\vec{\beta}$	k	SSR(full)	SSR(full)$/k$	
Due to $\vec{\beta}_I$	ℓ	SSR(reduced)	SSR(reduced)$/\ell$	F
Due to $\vec{\beta}_{II} \mid \vec{\beta}_I$	$k - \ell$	SSR(full | reduced)	SSR(full | reduced)$/(k - \ell)$	
Error	$n - (k + 1)$	SSE	MSE	
Total	$n - 1$	SST		

where

$$F = \frac{\text{SSR(full} \mid \text{reduced)}/(k - \ell)}{\text{SSE}/(n - k - 1)}$$

If we reject H_0, then the full model is better than the reduced model. Reject H_0 when $F > F_\alpha(k - \ell, n - k - 1)$.

Regression for Models without Full Rank

DEFINITION 10.1: Generalized Inverse

Let $\mathbf{A} \in \mathbb{R}^{n\times k}$. The **generalized inverse** (g-inverse) of $\mathbf{A}$ is *any* $\mathbf{G} \in \mathbb{R}^{k\times n}$ satisfying

$$\mathbf{AGA} = \mathbf{A}.$$

We say $\mathbf{G}$ is a g-inverse of $\mathbf{A}$.

EXAMPLE 10.1

Let $\mathbf{A} = \begin{pmatrix}1\\2\\3\\4\end{pmatrix}$. Find a g-inverse of $\mathbf{A}$.

Solution:

$$\begin{pmatrix}1\\2\\3\\4\end{pmatrix}\begin{pmatrix}a & b & c & d\end{pmatrix}\begin{pmatrix}1\\2\\3\\4\end{pmatrix} = \begin{pmatrix}1\\2\\3\\4\end{pmatrix}.$$

If we pick $a = 1, b = c = d = 0$, then $\mathbf{G} = \begin{pmatrix}1 & 0 & 0 & 0\end{pmatrix}$ is a g-inverse of $\mathbf{A}$. Also, $\tilde{\mathbf{G}} = \begin{pmatrix}1/2 & 1/4 & 0 & 0\end{pmatrix}$ is another g-inverse of $\mathbf{A}$. Hence, we can see that g-inverses are not unique.

REMARK — Basic Facts

(1) If $\mathbf{A}$ is invertible, then the g-inverse of $\mathbf{A}$ is unique and given by $\mathbf{A}^{-1}$.

- **Proof**: Let $\mathbf{G}$ be any g-inverse of $\mathbf{A}$, then $\mathbf{AGA} = \mathbf{A}$.

$$\mathbf{A}^{-1}\mathbf{AGAA}^{-1} = \mathbf{A}^{-1}\mathbf{AA}^{-1} \implies \mathbf{G} = \mathbf{A}^{-1}.$$

  Clearly, $\mathbf{AA}^{-1}\mathbf{A} = \mathbf{A}$.

(2) If $\mathbf{G}$ is a g-inverse of $\mathbf{A}$, then for any $\mathbf{C} \in \mathbb{R}^{k\times n}$, $\mathbf{G}_1 = \mathbf{G} + \mathbf{C} - \mathbf{GACAG}$ is also a g-inverse of $\mathbf{A}$.

- **Proof**: Note that

$$\begin{aligned}\mathbf{AG}_1\mathbf{A} &= \mathbf{AGA} + \mathbf{ACA} - \mathbf{AGACAGA}\\ &= \mathbf{AGA} + \mathbf{ACA} - \mathbf{ACA}\\ &= \mathbf{AGA}\\ &= \mathbf{A}.\end{aligned}$$

LEMMA 10.1

Every matrix $\mathbf{A}$ has at least one g-inverse.

Proof: Let $\mathbf{A} \in \mathbb{R}^{n\times k}$ with $\operatorname{rank}(\mathbf{A}) = r < \min\{n, k\}$. Then,

$$\mathbf{A} = \begin{pmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{pmatrix},$$

where $\mathbf{A}_{11} \in \mathbb{R}^{r\times r}$ with $\operatorname{rank}(\mathbf{A}_{11}) = r$.

Claim:

$$\mathbf{G} = \begin{pmatrix} \mathbf{A}_{11}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}_{k\times n}$$

is a g-inverse of $\mathbf{A}$.

$$\begin{aligned} \mathbf{AGA} &= \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{I}_r & \mathbf{O} \\ \mathbf{A}_{21}\mathbf{A}_{11} & \mathbf{O} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{pmatrix}. \end{aligned}$$

Since $\operatorname{rank}(\mathbf{A}) = r = \operatorname{rank}(\mathbf{A}_{11})$, it follows that $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$ are linear combinations of $\mathbf{A}_{11}$ and $\mathbf{A}_{12}$. Thus, one can find a $\mathbf{B} \in \mathbb{R}^{(n-r)\times r}$ such that

$$\begin{pmatrix} \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \mathbf{B} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \end{pmatrix}.$$

Hence,

$$\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = \mathbf{B}\mathbf{A}_{11}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = \mathbf{B}\mathbf{A}_{12} = \mathbf{A}_{22}.$$

Therefore, $\mathbf{AGA} = \mathbf{A}$.

LECTURE 11
13th February

11 Lecture 11: g-inverse

Algorithm for Finding a g-inverse

Let $\mathbf{A} \in \mathbb{R}^{n\times k}$ with $\operatorname{rank}(\mathbf{A}) = r < \min\{n, k\}$.

- Step 1: Find an invertible sub-matrix $\mathbf{M} \in \mathbb{R}^{r\times r}$.
- Step 2: Compute $(\mathbf{M}^{-1})'$.
- Step 3: Replace $\mathbf{M}$ with $(\mathbf{M}^{-1})'$ in $\mathbf{A}$.
- Step 4: Set all other elements in $\mathbf{A}$ to be 0.
- Step 5: Transpose the resulting matrix to $\mathbf{G} \in \mathbb{R}^{k\times n}$.

EXAMPLE 11.1

Compute a g-inverse of $\mathbf{A} = \begin{pmatrix} 4 & 1 & 2 & 0 \\ 1 & 1 & 5 & 15 \\ 3 & 1 & 3 & 5 \end{pmatrix}$.

Solution: Note that $n = 3$ and $k = 4$. Let

$$\begin{aligned} \vec{v}_1 &= \begin{pmatrix} 4 & 1 & 2 & 0 \end{pmatrix}, \\ \vec{v}_2 &= \begin{pmatrix} 1 & 1 & 5 & 15 \end{pmatrix}, \\ \vec{v}_3 &= \begin{pmatrix} 3 & 1 & 3 & 5 \end{pmatrix}. \end{aligned}$$

$\vec{v}_1$ and $\vec{v}_2$ are linearly independent since

$$a\vec{v}_1 + b\vec{v}_2 = \vec{0} \implies 15b = 0 \text{ and } 4a = 0 \implies a = b = 0.$$

Also, $3\vec{v}_3 = 2\vec{v}_1 + \vec{v}_2$. Therefore, $\operatorname{rank}(\mathbf{A}) = 2$. Now,

- Step 1:
$$\mathbf{M} = \begin{pmatrix} 4 & 0 \\ 3 & 5 \end{pmatrix}.$$

- Step 2:
$$(\mathbf{M}^{-1})' = \frac{1}{20}\begin{pmatrix} 5 & -3 \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} 5/20 & -3/20 \\ 0 & 4/20 \end{pmatrix}.$$

- Step 3:
$$\begin{pmatrix} 5/20 & 1 & 2 & -3/20 \\ 1 & 1 & 5 & 15 \\ 0 & 1 & 3 & 4/20 \end{pmatrix}.$$

- Step 4:
$$\begin{pmatrix} 5/20 & 0 & 0 & -3/20 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4/20 \end{pmatrix}.$$

- Step 5:
$$\begin{pmatrix} 5/20 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -3/20 & 0 & 4/20 \end{pmatrix}.$$

Verify that $\mathbf{AGA} = \mathbf{A}$.

THEOREM 11.1

Let $\mathbf{A} \in \mathbb{R}^{n\times k}$ with $\operatorname{rank}(\mathbf{A}) = r < \min\{n, k\}$ and $\mathbf{G}$ be a g-inverse of $\mathbf{A}$. Let $\mathbf{F}$ be a g-inverse of $\mathbf{A}'\mathbf{A}$. Then,

(1) $\mathbf{G}'$ *is a g-inverse of* $\mathbf{A}'$.

(2) $\operatorname{rank}(\mathbf{GA}) = \operatorname{rank}(\mathbf{AG}) = \operatorname{rank}(\mathbf{A}) = r$.

(3) $\mathbf{A} = \mathbf{AFA}'\mathbf{A}$ *and* $\mathbf{A}' = \mathbf{A}'\mathbf{AFA}'$. *This means that* $\mathbf{FA}'$ *is a g-inverse of* $\mathbf{A}$.

Proof:

(1) Since $\mathbf{AGA} = \mathbf{A}$, we have that $(\mathbf{AGA})' = \mathbf{A}'\mathbf{G}'\mathbf{A}' = \mathbf{A}'$. Therefore, $\mathbf{G}'$ is a g-inverse of $\mathbf{A}'$.

(2) Since $\mathbf{AGA} = \mathbf{A}$, $\operatorname{rank}(\mathbf{A}) \leq \operatorname{rank}(\mathbf{GA}) \leq \operatorname{rank}(\mathbf{A})$. Similarly, $\operatorname{rank}(\mathbf{A}) \leq \operatorname{rank}(\mathbf{AG}) \leq \operatorname{rank}(\mathbf{A})$. Therefore,
$$\operatorname{rank}(\mathbf{AG}) = \operatorname{rank}(\mathbf{GA}) = \operatorname{rank}(\mathbf{A}).$$

(3) Since $\mathbf{F}$ is a g-inverse of $\mathbf{A}'\mathbf{A}$, we have $\mathbf{A}'\mathbf{AFA}'\mathbf{A} = \mathbf{A}'\mathbf{A}$. Rearranging,
$$\begin{aligned} \mathbf{A}'\mathbf{AFA}'\mathbf{A} - \mathbf{A}'\mathbf{A} &= \mathbf{O} \\ (\mathbf{A}'\mathbf{AFA}' - \mathbf{A}')\mathbf{A} &= \mathbf{O} \\ \mathbf{A}'(\mathbf{AFA}'\mathbf{A} - \mathbf{A}) &= \mathbf{O}. \end{aligned}$$

Note that $(\mathbf{AFA'A})' = \mathbf{A'AF'A}$ and

$$\begin{aligned}\mathbf{A'AF'A'}(\mathbf{AFA'A} - \mathbf{A}) &= \mathbf{A'AF'}\underbrace{\mathbf{A'AFA'A}}_{\mathbf{A'A}} - \mathbf{A'AF'A'A} \\ &= \mathbf{A'AF'A'A} - \mathbf{A'AF'A'A} \\ &= \mathbf{O}.\end{aligned}$$

Therefore,

$$(\mathbf{AFA'A} - \mathbf{A})'(\mathbf{AFA'A} - \mathbf{A}) = \mathbf{O}.$$

Hence,

$$\mathbf{AFA'A} - \mathbf{A} = \mathbf{O} \implies \mathbf{AFA'A} = \mathbf{A}.$$

Similarly, $\mathbf{A'AFA'A} = \mathbf{A'A}$, which implies

$$(\mathbf{A'AFA'} - \mathbf{A'})\mathbf{A} = \mathbf{O}.$$

By direct calculation,

$$(\mathbf{A'AFA'} - \mathbf{A'})\mathbf{AF'A'A} = \mathbf{A'AF'A'A} - \mathbf{A'AF'A'A} = \mathbf{O}.$$

Therefore,

$$(\mathbf{A'AFA'} - \mathbf{A'})(\mathbf{A'AFA'} - \mathbf{A'})' = \mathbf{O}.$$

Hence,

$$\mathbf{A'AFA'} = \mathbf{A'}.$$

$\mathbf{AF}$ is a g-inverse of $\mathbf{A'}$ and $\mathbf{FA'}$ is a g-inverse of $\mathbf{A}$.

THEOREM 11.2

Let $\mathbf{F}$ be a g-inverse of $\mathbf{A'A}$.

(1) $\mathbf{F'}$ *is a g-inverse of* $\mathbf{A'A}$.

(2) $\operatorname{rank}(\mathbf{AFA'}) = \operatorname{rank}(\mathbf{A})$.

(3) Let $\tilde{\mathbf{F}}$ be any g-inverse of $\mathbf{A'A}$, then $\mathbf{A'FA} = \mathbf{A'}\tilde{\mathbf{F}}\mathbf{A}$.

(4) $\mathbf{AFA'}$ *is symmetric.*

Proof:

(1) Using Theorem 11.1, $\mathbf{A'AFA'A} = \mathbf{A'A}$, so

$$\mathbf{A'A} = (\mathbf{A'A})' = \mathbf{A'AF'A'A}.$$

(2) By Theorem 11.1, we have $\mathbf{A} = \mathbf{AFA'A}$. It follows that

$$\operatorname{rank}(\mathbf{A}) \leq \operatorname{rank}(\mathbf{AFA'}) \leq \operatorname{rank}(\mathbf{A}).$$

Therefore, $\operatorname{rank}(\mathbf{AFA'}) = \operatorname{rank}(\mathbf{A})$.

(3) Let $\tilde{\mathbf{F}}$ be any g-inverse of $\mathbf{A'A}$. Then, $\mathbf{A} = \mathbf{AFA'A} = \mathbf{A}\tilde{\mathbf{F}}\mathbf{A'A}$ by Theorem 11.1, so

$$(\mathbf{AFA'} - \mathbf{A}\tilde{\mathbf{F}}\mathbf{A'})\mathbf{A} = \mathbf{O}.$$

Therefore,

$$\begin{aligned}(\mathbf{AFA}' - \mathbf{A\tilde{F}A}')(\mathbf{AFA}' - \mathbf{A\tilde{F}A}')' &= (\mathbf{AFA}' - \mathbf{A\tilde{F}A}')(\mathbf{AF}'\mathbf{A}' - \mathbf{A\tilde{F}}'\mathbf{A}')' \\ &= (\mathbf{AFA}' - \mathbf{A\tilde{F}A}')\mathbf{A}(\mathbf{F}'\mathbf{A}' - \mathbf{\tilde{F}}'\mathbf{A}') \\ &= (\underbrace{\mathbf{AFA}'\mathbf{A}}_{\mathbf{A}} - \underbrace{\mathbf{A\tilde{F}A}'\mathbf{A}}_{\mathbf{A}})(\mathbf{F}'\mathbf{A}' - \mathbf{\tilde{F}}'\mathbf{A}') \\ &= \mathbf{O}.\end{aligned}$$

Hence, $\mathbf{AFA}' = \mathbf{A\tilde{F}A}'$.

(4) By (1), $\mathbf{F}'$ is a g-inverse of $\mathbf{A}'\mathbf{A}$. Hence, $\mathbf{AFA}' = \mathbf{AF}'\mathbf{A}' = (\mathbf{AFA}')'$. Therefore, $\mathbf{AFA}'$ is symmetric.

THEOREM 11.3

Let $\mathbf{A} \in \mathbb{R}^{n\times k}$. Consider the system of equations

$$\mathbf{A}\vec{x} = \vec{y}.$$

(1) If $\vec{x}_0$ is a solution of the system of equations, then $\mathbf{GA}\vec{x}_0$ is also a solution of the system of equations for any g-inverse $\mathbf{G}$ of $\mathbf{A}$.

(2) Let $\mathbf{G}$ be a g-inverse of $\mathbf{A}$, then for any $\vec{z} \in \mathbb{R}^k$,

$$\mathbf{G}\vec{y} + (\mathbf{GA} - \mathbf{I})\vec{z}$$

is a solution of the system of equations.

(3) Every solution can be written in the form of (2).

Proof:

(1) $\vec{x}_0$ is a solution implies that $\mathbf{A}\vec{x}_0 = \vec{y}$. However,

$$\mathbf{A}(\mathbf{GA}\vec{x}_0) = (\mathbf{AGA})\vec{x}_0 = \mathbf{A}\vec{x}_0 = \vec{y}.$$

(2) Note that

$$\{\mathbf{G}\vec{y} + (\mathbf{GA} - \mathbf{I})\vec{z} \mid \vec{z} \in \mathbb{R}^k\}.$$

So,

$$\begin{aligned}\mathbf{A}(\mathbf{G}\vec{y} + (\mathbf{GA} - \mathbf{I})\vec{z}) &= \mathbf{AG}\vec{y} + (\mathbf{AGA}\vec{z} - \mathbf{A}\vec{z}) \\ &= \mathbf{AG}\vec{y} \\ &= \mathbf{AG}(\mathbf{A}\vec{x}) \\ &= \mathbf{AGA}\vec{x} \\ &= \mathbf{A}\vec{x} \\ &= \vec{y}.\end{aligned}$$

(3) Let $\vec{x}_0$ be any solution, so $\mathbf{A}\vec{x}_0 = \vec{y}$. Choose $\vec{z} = (\mathbf{GA} - \mathbf{I})\vec{x}_0$.

$$\begin{aligned}\mathbf{G}\vec{y} + (\mathbf{GA} - \mathbf{I})\vec{z} &= \mathbf{G}\vec{y} + (\mathbf{GAGA} - 2\mathbf{GA} + \mathbf{I})\vec{x}_0 \\ &= \mathbf{G}\vec{y} + (-\mathbf{GA} + \mathbf{I})\vec{x}_0 \\ &= \mathbf{G}\vec{y} - \mathbf{GA}\vec{x}_0 + \vec{x}_0 \\ &= \mathbf{GA}\vec{x}_0 - \mathbf{GA}\vec{x}_0 + \vec{x}_0 \\ &= \vec{x}_0.\end{aligned}$$

12 Lecture 12: Regression Without Full Rank

DEFINITION 12.1: Estimable

Let $\mathbf{A} \in \mathbb{R}^{n\times k}$ with $\operatorname{rank}(\mathbf{A}) = r < \min\{n, k\}$, $\vec{x} \in \mathbb{R}^k$. Given a $\vec{b} \in \mathbb{R}^{k\times 1}$, the quantity $\vec{b}'\vec{x}$ is called **estimable** if its value is the same for every solution of $\mathbf{A}\vec{x} = \vec{y}$.

THEOREM 12.1

Let $\mathbf{A} \in \mathbb{R}^{n\times k}$, $\vec{b} \in \mathbb{R}^k$, *and* $\mathbf{G}$ *be a g-inverse of* $\mathbf{A}$.

$$\vec{b}'\vec{x} \textit{ is estimable} \iff \vec{b}'\mathbf{G}\mathbf{A} = \vec{b}'.$$

Proof: Let $\vec{x}_1$ and $\vec{x}_2$ be any two solutions of $\mathbf{A}\vec{x} = \vec{y}$. By Theorem 11.3, there exists $\vec{z}_1, \vec{z}_2 \in \mathbb{R}^k$ such that

$$\begin{aligned}\vec{x}_1 &= \mathbf{G}\vec{y} + (\mathbf{G}\mathbf{A} - \mathbf{I})\vec{z}_1\\ \vec{x}_2 &= \mathbf{G}\vec{y} + (\mathbf{G}\mathbf{A} - \mathbf{I})\vec{z}_2.\end{aligned}$$

($\Longleftarrow$) *Sufficiency*: Assume that $\vec{b}'\mathbf{G}\mathbf{A} = \vec{b}'$.

$$\begin{aligned}&\implies \vec{b}'\mathbf{G}\mathbf{A} - \vec{b}' = \vec{0}'\\ &\implies \vec{b}'(\mathbf{G}\mathbf{A} - \mathbf{I}) = \vec{0}'\\ &\implies \vec{b}'(\mathbf{G}\mathbf{A} - \mathbf{I})\vec{z} = 0\end{aligned}$$

Now,

$$\begin{aligned}\vec{b}'\vec{x}_1 - \vec{b}'\vec{x}_2 &= \vec{b}'(\vec{x}_1 - \vec{x}_2)\\ &= \vec{b}'\left[(\mathbf{G}\mathbf{A} - \mathbf{I})(\vec{z}_1 - \vec{z}_2)\right]\\ &= \vec{b}'(\mathbf{G}\mathbf{A} - \mathbf{I})(\vec{z}_1 - \vec{z}_2)\\ &= 0.\end{aligned}$$

Hence, $\vec{b}'\vec{x}_1 = \vec{b}'\vec{x}_2 \implies \vec{b}'\vec{x}$ is estimable.
($\Longrightarrow$) *Necessity*: Assume that $\vec{b}'\vec{x}$ is estimable. Let $\vec{x}_0$ be a solution of $\mathbf{A}\vec{x} = \vec{y}$. Choose $\vec{z}_0$ such that

$$\vec{x}_0 = \mathbf{G}\vec{y} + (\mathbf{G}\mathbf{A} - \mathbf{I})\vec{z}_0.$$

Let $z \in \mathbb{R}^k$. By definition,

$$\vec{x} = \mathbf{G}\vec{y} + (\mathbf{G}\mathbf{A} - \mathbf{I})(\vec{z}_0 - \vec{z}).$$

Hence, $\vec{x}$ is a solution of $\mathbf{A}\vec{x} = \vec{y}$, and so

$$\begin{aligned}\vec{b}'\vec{x} &= \vec{b}'\left[\mathbf{G}\vec{y} + (\mathbf{G}\mathbf{A} - \mathbf{I})\vec{z}_0 + (\mathbf{G}\mathbf{A} - \mathbf{I})\vec{z}\right]\\ &= \vec{b}'\vec{x}_0 + \vec{b}_1(\mathbf{G}\mathbf{A} - \mathbf{I})\vec{z}.\end{aligned}$$

So, $\vec{b}'(\mathbf{G}\mathbf{A} - \mathbf{I})\vec{z} = 0$ for all $\vec{z}$, therefore $\vec{b}'(\mathbf{G}\mathbf{A} - \mathbf{I}) = \vec{0}' \implies \vec{b}'\mathbf{G}\mathbf{A} = \vec{b}'$.

Non-Full Rank Regression

In regression, $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\varepsilon}$ and our normal equation is $\mathbf{X}'\mathbf{X}\vec{\beta} = \mathbf{X}'\vec{Y}$. We consider the case where $\mathbf{X}'\mathbf{X}$ is not invertible. Let $\mathbf{F}$ be a g-inverse of $\mathbf{X}'\mathbf{X}$, so

$$\vec{\beta}_0 = \mathbf{F}\mathbf{X}'\vec{Y}.$$

<u>Claim</u>: $\vec{\beta}_0$ is a solution of

$$\mathbf{X}'\mathbf{X}\vec{\beta} = \mathbf{X}'\vec{Y}.$$

By Theorem 11.1,

$$\mathbf{X}'\mathbf{X}\vec{\beta}_0 = \mathbf{X}'\mathbf{X}\mathbf{F}\mathbf{X}'\vec{Y} = \mathbf{X}'\vec{Y}.$$

However, $\vec{\beta}_0$ may not be a good estimator.

REMARK — Properties of $\vec{\beta}_0$

(1) Expectation: $\mathbb{E}[\vec{\beta}_0] = \mathbb{E}[\mathbf{F}\mathbf{X}'\vec{Y}] = \mathbf{F}\mathbf{X}'\,\mathbb{E}[\vec{Y}] = \mathbf{F}\mathbf{X}'\mathbf{X}\vec{\beta} \neq \vec{\beta}$ in general.

(2) Variance:

$$\begin{aligned}
\operatorname{Var}(\vec{\beta}_0) &= \operatorname{Cov}(\mathbf{F}\mathbf{X}\vec{Y}, \mathbf{F}\mathbf{X}\vec{Y}) \\
&= \mathbf{F}\mathbf{X}'\operatorname{Cov}(\vec{Y}, \vec{Y})\mathbf{X}\mathbf{F}' \\
&= \sigma^2\mathbf{F}\mathbf{X}'\mathbf{X}\mathbf{F}'.
\end{aligned}$$

(3) Fitted values:

$$\begin{aligned}
\hat{\vec{Y}} &= \mathbf{X}\vec{\beta}_0 \\
&= \mathbf{X}\mathbf{F}\mathbf{X}'\vec{Y} \\
&= (\mathbf{X}\mathbf{F}\mathbf{X}')\vec{Y}.
\end{aligned}$$

Note that by Theorem 11.2 (3), $\hat{\vec{Y}}$ does not depend on which g-inverse ($\mathbf{F}$) we use.

(4) SSE:

$$\begin{aligned}
\text{SSE} &= (\vec{Y} - \hat{\vec{Y}})'(\vec{Y} - \hat{\vec{Y}}) \\
&= (\vec{Y} - \mathbf{X}\mathbf{F}\mathbf{X}'\vec{Y})'(\vec{Y} - \mathbf{X}\mathbf{F}\mathbf{X}'\vec{Y}) \\
&= \vec{Y}'(\mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}')'(\mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}')\mathbf{Y} \\
&= \vec{Y}'(\mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}')(\mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}')\mathbf{Y} && \text{Theorem 11.2 (4)} \\
&= \vec{Y}'(\mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}')\vec{Y},
\end{aligned}$$

where

$$\begin{aligned}
(\mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}')(\mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}') &= \mathbf{I} - 2\mathbf{X}\mathbf{F}\mathbf{X}' + \underbrace{\mathbf{X}\mathbf{F}\mathbf{X}'\mathbf{X}}_{\mathbf{X}}\mathbf{F}\mathbf{X} \\
&= \mathbf{I} - 2\mathbf{X}\mathbf{F}\mathbf{X}' + \mathbf{X}\mathbf{F}\mathbf{X}' && \text{Theorem 11.1 (3)} \\
&= \mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}'.
\end{aligned}$$

EXAMPLE 12.1

Weights of Six Plants		
Types of Plants		
Normal	Off-type	Aberrant
101	84	32
105	88	
94		
300	172	32

Relation between weight and types. Let Y = weight, x_1 = normal, x_2 = off-type, x_3 = aberrant, where all the covariates are binary. Y_{ij} = observation of j^{th} plant of type i for $i = 1, 2, 3$; $n_1 = 3$, $n_2 = 2$, $n_3 = 1$, and $n = n_1 + n_2 + n_3 = 6$.

$$\vec{Y}' = \begin{pmatrix} Y_{11} & Y_{12} & Y_{13} & Y_{21} & Y_{22} & Y_{23} \end{pmatrix}.$$

Regression model:

$$Y_{ij} = \beta_0 + \beta_i + \varepsilon_{ij}.$$

- β_0 = population mean;
- β_i = effect of type i on the weight;
- ε_{ij} = random error of observation Y_{ij}.

Explicitly, we have

$$\begin{aligned}
Y_{11} &= \beta_0 + \beta_1 + 0\beta_2 + 0\beta_3 + \varepsilon_{11} \\
Y_{12} &= \beta_0 + \beta_1 + 0\beta_2 + 0\beta_3 + \varepsilon_{12} \\
Y_{13} &= \beta_0 + \beta_1 + 0\beta_2 + 0\beta_3 + \varepsilon_{13} \\
Y_{21} &= \beta_0 + 0\beta_1 + \beta_2 + 0\beta_3 + \varepsilon_{21} \\
Y_{22} &= \beta_0 + 0\beta_1 + \beta_2 + 0\beta_3 + \varepsilon_{22} \\
Y_{23} &= \beta_0 + 0\beta_1 + 0\beta_2 + \beta_3 + \varepsilon_{23}.
\end{aligned}$$

Therefore, $\vec{Y} = \mathbf{X}\vec{\beta}$ with

$$\mathbf{X}\vec{\beta} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Note that $\text{rank}(\mathbf{X}) = 3 < \min\{6, 4\}$ so $\mathbf{X}$ is not full rank.

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 6 & 3 & 2 & 1 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

$$\mathbf{X}'\vec{Y} = \begin{pmatrix} Y_{0.} \\ Y_{1.} \\ Y_{2.} \\ Y_{33} \end{pmatrix} = \begin{pmatrix} 504 \\ 300 \\ 172 \\ 32 \end{pmatrix}.$$

Normal equation $\mathbf{X}'\mathbf{X}\vec{\beta} = \mathbf{X}'\vec{Y}$. g-inverse of $\mathbf{X}'\mathbf{X}$:

$$\mathbf{M} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \implies \mathbf{F} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Note that

$$\vec{\beta}_0 = \mathbf{F}\mathbf{X}'\vec{Y} = \begin{pmatrix} 0 \\ 100 \\ 86 \\ 32 \end{pmatrix}$$

is one solution. However, we cannot claim that $\vec{\beta}_0$ is an estimate of $\vec{\beta}$. By direct calculation,

$$\mathbf{F}\mathbf{X}'\mathbf{X} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 6 & 3 & 2 & 1 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 7 & 3 & 2 & 2 \end{pmatrix}.$$

Choose $\vec{b}' = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix}$ and compute

$$\vec{b}'\mathbf{F}\mathbf{X}'\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix} = \vec{b}',$$

so $\vec{b}'\vec{\beta} = \beta_0 + \beta_1$ is estimable. Therefore, $\hat{\beta}_0 + \hat{\beta}_1$ is an estimator of $\beta_0 + \beta_1$.

LECTURE 13
27th February

13 Lecture 13: Regression Without Full Rank (Continued)

THEOREM 13.1

Assume that $\operatorname{rank}(\mathbf{X}) = r$. *Then,*

$$\hat{\sigma}^2 = s^2 = \frac{SSE}{n-r}$$

is an unbiased estimator of σ^2.

Proof:

THEOREM 13.2

Assume that $\operatorname{rank}(\mathbf{X}) = r$ *and* $\vec{Y} \sim MN(\mathbf{X}\vec{\beta}, \sigma^2\mathbf{I})$.

(1) $\vec{\beta}_0 = \mathbf{F}\mathbf{X}\vec{Y} \sim MN(\mathbf{F}\mathbf{X}'\mathbf{X}\vec{\beta}, \sigma^2\mathbf{F}\mathbf{X}'\mathbf{X}\mathbf{F}')$.

(2) $\vec{\beta}_0$ *and* $\hat{\sigma}^2$ *are independent.*

(3) $SSE/\sigma^2 \sim \chi^2(n-r)$.

(4) $SSR/\sigma^2 \sim \chi^2(r-1, \lambda)$, *where*

$$\lambda = \frac{1}{2\sigma^2}(\mathbf{X}\vec{\beta})'(\mathbf{X}\mathbf{F}\mathbf{X}' - \tfrac{1}{n}\mathbf{J})\mathbf{X}\vec{\beta}.$$

(5) SSE and SSR are independent.

Proof:

(1) Trivial.

(2) $\vec{\beta}_0 = \mathbf{F}\mathbf{X}'\vec{Y}$ and

$$\hat{\sigma}^2 = \frac{\text{SSE}}{n-r} = \frac{1}{n-r}\vec{Y}'(\mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}')\vec{Y}.$$

By direct calculation,

$$\begin{aligned}
&\mathbf{F}\mathbf{X}'\sigma^2\mathbf{I}(\mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}') \\
&\qquad = \frac{\sigma^2}{n-r}\mathbf{F}\mathbf{X}'(\mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}') \\
&\qquad = \frac{\sigma^2}{n-r}[\mathbf{F}\mathbf{X}' - \mathbf{F}\mathbf{X}'\mathbf{X}\mathbf{F}\mathbf{X}'] \\
&\qquad = \frac{\sigma^2}{n-r}[\mathbf{F}\mathbf{X}' - \mathbf{F}\mathbf{X}'] && \text{Theorem 11.1} \\
&\qquad = \mathbf{O}.
\end{aligned}$$

Therefore, $\vec{\beta}_0$ and $\hat{\sigma}^2$ are independent by Theorem 5.1.

(3) Note that

$$\frac{\text{SSE}}{\sigma^2} = \frac{\vec{Y}'}{\sigma}(\mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}')\frac{\vec{Y}}{\sigma}$$

and

$$\frac{\vec{Y}}{\sigma} \sim \text{MN}\left(\frac{\mathbf{X}\vec{\beta}}{\sigma}, \mathbf{I}\right).$$

Note that $\mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}'$ is idempotent (see properties of $\vec{\beta}_0$). Now,

$$\begin{aligned}
\lambda &= \frac{1}{2}\left(\frac{\mathbf{X}\vec{\beta}}{\sigma}\right)'(\mathbf{I} - \mathbf{X}\mathbf{F}\mathbf{X}')\frac{\mathbf{X}\vec{\beta}}{\sigma} \\
&= \frac{1}{2\sigma^2}\vec{\beta}'[\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{X}\mathbf{F}\mathbf{X}'\mathbf{X}]\vec{\beta} \\
&= \frac{1}{2\sigma^2}\vec{\beta}'[\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{X}]\vec{\beta} \\
&= 0.
\end{aligned}$$

The result follows by Theorem 4.2.

(4) Note that

$$\frac{\text{SSR}}{\sigma^2} = \frac{\vec{Y}'}{\sigma}(\mathbf{X}\mathbf{F}\mathbf{X}' - \tfrac{1}{n}\mathbf{J})\frac{\vec{Y}}{\sigma}.$$

We need to show that $\mathbf{X}\mathbf{F}\mathbf{X}' - \frac{1}{n}\mathbf{J}$ is idempotent.

$$\begin{aligned}
(\mathbf{X}\mathbf{F}\mathbf{X}' - \tfrac{1}{n}\mathbf{J})^2 &= (\mathbf{X}\mathbf{F}\mathbf{X}' - \tfrac{1}{n}\mathbf{J})(\mathbf{X}\mathbf{F}\mathbf{X}' - \tfrac{1}{n}\mathbf{J}) \\
&= \mathbf{X}\mathbf{F}\mathbf{X}'\mathbf{X}\mathbf{F}\mathbf{X}' - \tfrac{1}{n}\mathbf{J}\mathbf{X}\mathbf{F}\mathbf{X}' - \tfrac{1}{n}\mathbf{X}\mathbf{F}\mathbf{X}'\mathbf{J} + \tfrac{1}{n^2}\mathbf{J}\mathbf{J} \\
&= \mathbf{X}\mathbf{F}\mathbf{X}' - \tfrac{1}{n}\mathbf{X}\mathbf{F}\mathbf{X}'\mathbf{J} - \tfrac{1}{n}\mathbf{J}\mathbf{X}\mathbf{F}\mathbf{X}' + \tfrac{1}{n}\mathbf{J}.
\end{aligned}$$

We know that $\mathbf{X}\mathbf{F}\mathbf{X}'\mathbf{X} = \mathbf{X}$, so partitioning we see

$$\mathbf{X}\mathbf{F}\mathbf{X}'\begin{pmatrix}\vec{j} & \mathbf{X}_1\end{pmatrix} = \mathbf{X},$$

which implies that $\mathbf{X}\mathbf{F}\mathbf{X}'\vec{j} = \vec{j}$. Therefore, $\mathbf{X}\mathbf{F}\mathbf{X}'\mathbf{J} = \mathbf{J}$.

(5) $\text{SSE} = \vec{Y}'(\mathbf{I} - \mathbf{XFX}')\vec{Y}$ and $\text{SSR} = \vec{Y}'(\mathbf{XFX}' - \frac{1}{n}\mathbf{J})\vec{Y}$.

$$\begin{aligned}(\mathbf{I} - \mathbf{XFX}')(\mathbf{XFX}' - \tfrac{1}{n}\mathbf{J}) &= \mathbf{XFX}' - \tfrac{1}{n}\mathbf{J} - \mathbf{XFX}'\mathbf{XFX} + \tfrac{1}{n}\mathbf{XFX}'\mathbf{J} \\ &= \mathbf{XFX}' - \tfrac{1}{n}\mathbf{J} - \mathbf{XFX}' + \tfrac{1}{n}\mathbf{J} \\ &= \mathbf{O}.\end{aligned}$$

The result follows by Theorem 5.2.

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Due to $\vec{\beta}_1$	r	SSR	MSR	$\frac{\text{SSR}/(r-1)}{\text{SSE}/(n-r)}$
Error	$n - r$	SSE	MSE	
Total	$n - 1$	SST		

THEOREM 13.3

Let $\vec{\beta} = \begin{pmatrix}\beta_0 \\ \vdots \\ \beta_k\end{pmatrix} \in \mathbb{R}^{k+1}$ *and* $\mathbf{F}$ *is a* g*-inverse of* $\mathbf{X}'\mathbf{X}$. *Then,* $\vec{b}'\vec{\beta}$ *is estimable if and only if one of the following hold:*

(1) $\vec{b}'\mathbf{FX}'\mathbf{X} = \vec{b}'$.

(2) There exists $\vec{a} \in \mathbb{R}^n$ *such that* $\vec{b}' = \vec{a}'\mathbf{X}$.

(3) There exists $\vec{c} \in \mathbb{R}^{k+1}$ *such that*
$$\vec{b}' = \vec{c}'\mathbf{X}'\mathbf{X}.$$

Proof:

(1) Earlier proof.

(2) To show (1) $\implies$ (2), choose $\vec{a}' = \vec{b}'\mathbf{FX}'$, which implies that $\vec{a}'\mathbf{X} = \vec{b}'$. To show (2) $\implies$ (1), note that $\vec{b}' = \vec{a}'\mathbf{X}$, and multiply to get $\vec{b}'\mathbf{FX}'\mathbf{X} = \vec{a}'\mathbf{XFX}'\mathbf{X} = \vec{a}'\mathbf{X}$.

(3) To show (1) $\implies$ (3), choose $\vec{c} = \vec{b}'\mathbf{F}$. To show (3) $\implies$ (1),

$$\begin{aligned}\vec{b}' &= \vec{c}'\mathbf{X}'\mathbf{X} \\ \vec{b}'\mathbf{FX}'\mathbf{X} &= \vec{c}'\mathbf{X}'\mathbf{XFX}'\mathbf{X} \\ &= \vec{c}'\mathbf{X}'\mathbf{X} \\ &= \vec{b}'.\end{aligned}$$

REMARK

Assume that $\vec{b}'\vec{\beta}$ is estimable. Let $\vec{\beta}_0 = \mathbf{FX}'\vec{Y}$. $\vec{b}'\vec{\beta}_0$ is a good estimator of $\vec{b}'\vec{\beta}$.

The expectation of $\vec{b}'\vec{\beta}_0$ is

$$\begin{aligned}\mathbb{E}[\vec{b}'\vec{\beta}_0] &= \vec{b}'\mathbf{F}\mathbf{X}'\,\mathbb{E}[\vec{\beta}_0] \\ &= \vec{b}'\mathbf{F}\vec{X}'\mathbf{X}\vec{\beta} \\ &= \vec{\beta}'\vec{\beta}.\end{aligned}$$

Hence, $\vec{b}'\vec{\beta}_0$ is an unbiased estimator of $\vec{b}'\vec{\beta}$. The variance of $\vec{b}'\vec{\beta}_0$ is

$$\begin{aligned}\mathrm{Var}(\vec{b}'\vec{\beta}_0) &= \vec{b}'\,\mathrm{Var}(\vec{\beta}_0)\vec{b} \\ &= \vec{b}'\sigma^2\mathbf{F}\mathbf{X}'\mathbf{X}\mathbf{F}'\vec{b} \\ &= \sigma^2\vec{b}'\mathbf{F}\mathbf{X}'\mathbf{X}'\mathbf{F}'\vec{b} \\ &= \sigma^2\vec{b}'\mathbf{F}'\vec{b} \\ &= \sigma^2\vec{b}'\mathbf{F}\vec{b}\end{aligned}$$

by Theorem 11.2 (3).
Hence, $\vec{b}'\vec{\beta}_0$ is an estimator for $\vec{b}'\vec{\beta}$.

THEOREM 13.4

If $\vec{Y} \sim MN(\mathbf{X}\vec{\beta}, \sigma^2\mathbf{I})$, *then*

$$\vec{b}'\vec{\beta}_0 \sim MN(\vec{b}'\vec{\beta}, \sigma^2\vec{b}'\mathbf{F}\vec{b}),$$

and

$$\frac{\vec{b}'\vec{\beta}_0 - \vec{b}'\vec{\beta}}{\sqrt{\sigma^2\vec{b}'\mathbf{F}\vec{b}}} \sim \mathcal{N}(0,1).$$

However, since σ^2 is unknown, this quantity is a t-distribution.