# Ch 6: THE GENERAL LINEAR MODEL

## 6.2 Mean vectors and covariance matrices

1. (Random vector)

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \quad , \quad x_i = \text{ rand var}$$

2. (Mean vector)

$$E\,x = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} \quad , \quad \mu_i = E\,x_i$$

3. (Covariance matrix)

$$\text{cov}(x, y) = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mn} \end{pmatrix} \quad , \quad \sigma_{ij} = \text{cov}(x_i, y_j)$$

4. (Def) $\text{cov}\,x = cov(x, x)$

5. (Th) $cov(Ax, By) = A\text{cov}(x, y)B'$

   Pf: See text. $\quad cov(au, bv) = ab\text{cov}(u, v)$

6. (Corr) $\text{cov}(Ax) = A\text{cov}(x)A'$

7. (Th) $\text{cov}\, x$ is non-negative definite.

   Pf: $\ell'\text{cov}(x)\ell = \text{cov}(\ell' x) = \text{var}(\ell' x) \geq 0$

## 6.3 The multivariate normal distribution

1. (Def) $y \sim N \iff \ell' y \sim N$ all $\ell \in R^n$

2. (Th) $y \sim N \Longrightarrow a + By \sim N$

   Pf: $\ell'(a + By) = \ell' a + \ell' By \sim N$

3. (Th) $\ell' x \sim \ell' y$ all $\ell \Longrightarrow x \sim y$

   Pf: Deep

4. (Corr) The distribution of a normal random vector is determined by its mean vector $\mu$ and covariance matrix $\Sigma$.

   Pf: Given

$$x \sim N \quad , \quad y \sim N$$

$$Ex = Ey = \mu$$

$$\operatorname{cov} x = cov\, y = \Sigma$$

   we must show $x \sim y$.

   To do this note

$$\ell' x \sim N(\ell' \mu, \ell' \Sigma \ell) \sim \ell' y \qquad \text{all } \ell$$

   Using Th 3, $x \sim y$.

5. (Th) For any $n$-vector $\mu$ and any $n$ by $n$ non-negative definite matrix $\Sigma$, there is a random vector $y$ such that

$$y \sim N$$
$$Ey = \mu$$
$$\operatorname{cov} y = \Sigma$$

Pf: Problem 1. It uses $y = \mu + \Sigma^{1/2} z$.

6. (Note) The notation $y \sim N(\mu, \Sigma)$ makes sense.

7. (Corr) Jointly normal random vectors $y_1, \cdots, y_r$ are independent iff

$$\operatorname{cov}(y_i, y_j) = 0 \quad \text{all } i \neq j$$

Pf: Text

## 6.4 Projections

1. (Def)

$$\mathcal{V} = \text{inner product space of dimension } n$$
$$(u, v) = \text{inner product (e.g. } u'v)$$
$$\|u\| = (u, u)^{1/2} = \text{norm} = \text{length}$$

2. (Def) Let $\mathcal{M}$ be a subspace of $\mathcal{V}$ of dimension $p$.

   $\hat{y}$ = **projection** of $y$ on $\mathcal{M} \stackrel{df}{=}$ closest point

3. (Th) $\hat{y}$ exists and is unique.

   Pf: Text

4. (Th) $\hat{y} \in \mathcal{M}$ is the projection of $y$ on $\mathcal{M}$ iff

$$y - \hat{y} \perp \mathcal{M}$$

Pf: Text

5. (Ex) If

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

then

$$\bar{y} = \begin{pmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix}$$

is the projection of $y$ on the space $\mathcal{C}$ of constant vectors.

Pf: Let $\mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$. Then

$$(\mathbf{1}, y - \bar{y}) = \sum (y_i - \bar{y}) = 0$$

Thus $y - \bar{y} \perp \mathcal{C}$ and hence $\bar{y}$ is the proj of $y$ on $\mathcal{C}$.

6. (Def) If

$$Py = \hat{y}$$

$P$ is called the operator that projects onto $\mathcal{M}$.

7. (Th)

   (i) $P$ is linear

   (ii) $\mathcal{R}(P) = \mathcal{M}$ (range)

   (iii) $\mathcal{N}(P) = \mathcal{M}^{\perp}$ (null)

   (iv) $PP = P$ (idempotent)

   Pf: Text

8. (Th) $y - \hat{y}$ is the projection of $y$ on $\mathcal{M}^{\perp}$.

   Pf: Text

9. (Picture)

10. (Def) $A$ is self-adjoint iff

$$(Au, v) = (u, Av) \quad \text{all } u, v \in \mathcal{V}$$

11. (Th) $A$ is self-adjoint iff its matrix w.r.t. any orthogonal basis of $\mathcal{V}$ is symmetric.

    Pf: Problem 4

12. (Th) $P$ is a projection operator iff it is idempotent and self-adjoint. $(PP = P,\ P = P')$

    Pf: Text

13. (Th) If $P$ is a projection operator

$$\operatorname{rank} P = \operatorname{tr} P$$

    Pf: Problem 5

14. (Corr) If $P$ and $P_1, \cdots, P_k$ are projection operators and if

$$P = \alpha_1 P_1 + \cdots + \alpha_k P_k$$

    Then

$$\operatorname{rank} P = \alpha_1 \operatorname{rank} P_1 + \cdots + \alpha_k \operatorname{rank} P_k$$

15. (Ex) $\text{rank}(I - P) = n - p$

16. (Note) $\|Py\|^2$ is a quadratic form and

$$\text{rank}\,\|Py\|^2 = \text{rank}\,P$$

Pf:

$$\|Py\|^2 = (Py, Py) = (y, PPy) = (y, Py) = y'Py$$

Thus

$$\text{rank}\|Py\|^2 = \text{rank}\,P$$

17. (Th) If $x_1, \cdots, x_n$ is an orthonormal basis of $\mathcal{V}$

$$\|y\|^2 = \sum (x_i, y)^2$$

Pf: Let

$$y = \alpha_1 x_1 + \cdots + \alpha_n x_n$$

Then

$$(x_i, y) = \alpha_i$$

and

$$\begin{aligned} \|y\|^2 &= (y, y) = (\sum \alpha_i x_i, y) \\ &= \sum \alpha_i (x_i, y) = \sum (x_i, y)^2 \end{aligned}$$

## 6.5 The $\chi^2$-theorem

1. (Def) In this section all vectors are column vectors and

$$(u, v) = u'v$$

2. (Recall) If $z_1, \cdots, z_k$ are independent $N(0,1)$

$$z_1^2 + \cdots + z_k^2 \sim \chi^2(k)$$

3. ($\chi^2$-Th) If

$$\begin{aligned} y &\sim N(0, \sigma^2 I) \\ P &= \text{proj operator} \end{aligned}$$

then

$$\|Py\|^2 \sim \sigma^2 \chi^2(\operatorname{rank} P)$$

Pf: Let $\ell_1, \cdots, \ell_p$ be an orthonormal basis of $\mathcal{R}(P)$. Then

$$\operatorname{cov}((\ell_i, y), (\ell_i, y)) = \operatorname{cov}(\ell_i' y, \ell_j' y) = \ell_i'(\sigma^2 I)\ell_j = \sigma^2 \delta_{ij}$$

Thus

$$(\ell_1, y)/\sigma, \cdots, (\ell_p, y)/\sigma$$

are independent $N(0,1)$ and

$$\begin{aligned}\|Py\|^2 &= \sum_{i=1}^{p}(\ell_i, Py)^2 = \sum(\ell_i, y)^2 \\ &= \sigma^2 \sum((\ell_i, y)/\sigma)^2 \sim \sigma^2\chi^2(p) \\ &= \sigma^2\chi^2(\operatorname{rank} P)\end{aligned}$$

## 6.6 The general linear model

### The model

1. (Def) The general linear model is

$$y = \mu + e \quad , \quad \mu \in \mathcal{M}$$

where

$$\begin{aligned} \mathcal{M} &= \text{a } p \text{ dim subspace of } \mathcal{V} \\ \mu & \quad \text{unknown} \\ e & \quad \text{random, unknown} \end{aligned}$$

2. (Possible assumptions)

$$\begin{array}{ll} E\,e = 0 & \text{unbiased} \\ \operatorname{cov} e = \sigma^2 I & \text{uncorrelated, const var} \\ e \sim N & \text{normality} \end{array}$$

3. (Ex) Given $x = (x_i)$, let

$$\begin{aligned} \mathcal{M} &= \{(\mu_i) : \mu_i = \alpha + \beta x_i \text{ , some } \alpha, \beta\} \\ &= \text{the simple linear regression model} \end{aligned}$$

4. (Ex)

$$\begin{aligned} \mathcal{M} &= \{(\mu_{ij}) : \mu_{ij} = \alpha_i + \beta_j \text{ some } \alpha_i, \beta_j\} \\ &= \text{the additive two-way ANOVA model} \end{aligned}$$

**Least squares**

1. (Def) $\hat{y} \in \mathcal{M}$ is the **least squares estimate** of $\mu$ if $\mu = \hat{y}$ minimizes
$$Q(\mu) = \|y - \mu\|^2 \quad , \quad \mu \in \mathcal{M}$$
2. (Note) $\hat{y}$ = proj of $y$ on $\mathcal{M}$
3. (Fitted model) $y = \hat{y} + \hat{e}$

**Statistical properties**

1. (Def) Let $P =$ proj on $\mathcal{M}$ and $Q = I - P$, then

$$\hat{y} = Py \quad , \quad \hat{e} = Qy$$

2. (Th) If $E\,e = 0$, then

$$\begin{aligned} E\hat{y} &= \mu \qquad (\ E\hat{y} = EPy = P\mu = \mu\ ) \\ E\hat{e} &= 0 \qquad (\ E\hat{e} = E(y - \hat{y}) = \mu - \mu = 0\ ) \end{aligned}$$

3. (Th) If $\operatorname{cov} e = \sigma^2 I$, then

$$\begin{aligned} \operatorname{cov}\hat{y} &= \sigma^2 P \\ \operatorname{cov}\hat{e} &= \sigma^2 Q \\ \operatorname{cov}(\hat{y}, \hat{e}) &= 0 \end{aligned}$$

If, moreover, $Ee = 0$, then

$$E\|\hat{e}\|^2 = (n-p)\sigma^2$$

Pf: Text. For example

$$\begin{aligned} \operatorname{cov}(\hat{y}, \hat{e}) &= \operatorname{cov}(Py, Qy) = P\operatorname{cov}(y, y)Q' \\ &= P(\sigma^2 I)Q = \sigma^2 P(I - P) = 0 \end{aligned}$$

4. (Corr) Let RSS $= \|\hat{e}\|^2$, then

$$\hat{\sigma}^2 = \text{RSS}/(n-p)$$

is an unbiased estimate of $\sigma^2$.

5. (Th 6.11) If $e \sim N(0, \sigma^2 I)$, then

(i) $\hat{y}$ and $\hat{e}$ are normal and independent

(ii) $\|\hat{e}\|^2 \sim \sigma^2\chi^2(n-p)$

Pf:

$$\hat{y} = Py \ , \ \hat{e} = Qy \Longrightarrow \hat{y} \ , \ \hat{e} \text{ normal}$$

$$cov(\hat{y}, \hat{e}) = 0 \Longrightarrow \hat{y} \ , \ \hat{e} \text{ indep}$$

Since

$$\hat{e} = Qy = Q(\mu + e) = Qe$$

by the $\chi^2$-th

$$\begin{aligned} \|\hat{e}\|^2 &= \|Qe\|^2 \sim \sigma^2\chi^2(\text{rank}\, Q) \\ &= \sigma^2\chi^2(n-p) \end{aligned}$$

6. (Prob 10) $y_1, \cdots, y_n$ indep $N(\mu, \sigma^2)$

$$y_i = \mu + e_i$$

$$y = \mu + e \quad , \quad \mu = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} \quad , \quad \mathcal{M} = [\mathbf{1}]$$

$$\hat{y} = \begin{pmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix} \quad , \qquad \text{Ex 6.1}$$

(a)Th 6.11 $\Longrightarrow \hat{y}, \hat{e}$ are independent. Since

$$\bar{y} = \hat{y}_1 = \text{fnc}(\hat{y})$$
$$s_y^2 = \tfrac{1}{n-1}\Sigma(y_i - \bar{y})^2 = \tfrac{1}{n-1}\|\hat{e}\|^2 = \text{fnc}(\hat{e})$$

it follows that $\bar{y}$ and $s_y^2$ are independent.

(b) Since $p = 1$, by Th 6.11

$$(n-1)s_y^2 = \|\hat{e}\|^2 \sim \sigma^2\chi^2(n-1)$$

7. (Note) The two-way tables $\mu_{ij}$ of the form

$$\mu_{ij} = \alpha_i \quad , \quad i = 1, \cdots, r \ ; \ j = 1, \cdots, c$$

for some $\alpha_i$ are a linear space of dimension $r$.

Pf:

$$\begin{pmatrix} \alpha_1 & \cdots & \alpha_1 \\ \vdots & & \vdots \\ \alpha_r & \cdots & \alpha_r \end{pmatrix} \longleftrightarrow \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_r \end{pmatrix}$$

Thus

$$\dim \mathcal{M} = \dim R^r = r$$

8. (Prob 14) $y_{ij} = \alpha_i + e_{ij}$ , $i = 1, \cdots, r$ ; $j = 1, \cdots, c$. Then

$$\mathcal{M} = \{(\mu_{ij}) : \mu_{ij} = \alpha_i \text{ for some } \alpha_i\}$$

$$\dim \mathcal{M} = r$$

(a)

$$\|y - \mu\|^2 = \Sigma_i(\Sigma_j(y_{ij} - \alpha_i)^2)$$

$$\hat{\alpha}_i = \bar{y}_{i.}$$

(b)

$$\bar{y}_{i.} = \hat{\alpha}_i = \hat{y}_{i1} = \text{fnc}(\hat{y})$$

$$\text{SSW} = \Sigma\Sigma(y_{ij} - \bar{y}_{i.})^2 = \|y - \hat{y}\|^2 = \text{fnc}(y - \hat{y})$$

Now

$$\begin{aligned} \text{Th 6.11} &\implies \hat{y}, y - \hat{y} \ \text{ indep} \\ &\implies (\bar{y}_{1.}, \cdots, \bar{y}_{r.}), \text{SSW indep} \\ &\implies \bar{y}_{1.}, \cdots, \bar{y}_{r.}, \text{SSW indep} \end{aligned}$$

(c) Th 6.11 $\Longrightarrow \text{SSW} = \|y - \hat{y}\|^2 \sim \sigma^2 \chi^2(rc - r)$

## 6.7 The linear regression model

1. (Model) $y = X\beta + e$ , $X =$ a given $n \times p$ matrix

2. (Note) This is the GLM with

$$\mathcal{M} = \{\mu : \mu = X\beta \text{ some } \beta\} = \mathcal{C}(X) = \text{ col sp of } X$$

3. (Def) $\hat{\beta}$ is the **least squares estimate** of $\beta$ if $\beta = \hat{\beta}$ minimizes

$$Q(\beta) = \|y - X\beta\|^2$$

4. (Note)

$$\hat{\beta} \text{ is the LS est of } \beta \iff X\hat{\beta} = \hat{y} = Py$$

5. (Normal equations) $\hat{\beta}$ is a LS est of $\beta$ iff

$$X'X\hat{\beta} = X'y$$

   Pf: $\hat{\beta}$ is a LS est $\iff y - X\hat{\beta} \perp \mathcal{M} \iff y - X\hat{\beta} \perp X \iff X'(y - X\hat{\beta}) = 0 \iff X'X\hat{\beta} = X'y$

6. (Th) If $X'X$ is nonsingular,

   (i) $\hat{\beta} = (X'X)^{-1}X'y$

and if $Ee = 0$

(ii) $E\hat{\beta} = \beta$

and if $\operatorname{cov} e = \sigma^2 I$

(iii) $\operatorname{cov}\hat{\beta} = \sigma^2(X'X)^{-1}$

and if $e \sim N$

(iv) $\hat{\beta}$ and $\hat{\sigma}^2$ are independent

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

Pf: (i), (ii), (v) are easy.

(iii)

$$\begin{aligned}
\operatorname{cov}\hat{\beta} &= \operatorname{cov}((X'X)^{-1}X'y) \\
&= (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}$$

(iv)

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y = (X'X)^{-1}X'(\hat{y} + \hat{e}) \\
&= (X'X)^{-1}X'\hat{y} = \mathrm{fnc}(\hat{y}) \\
\hat{\sigma}^2 &= \|\hat{e}\|^2/(n-p) = \mathrm{fnc}(\hat{e})
\end{aligned}$$

Th 6.10 $\Longrightarrow$ $\hat{y}$ and $\hat{e}$ indep $\Longrightarrow$ $\hat{\beta}$ and $\hat{\sigma}^2$ indep

**Regression model $t$-statistic**

1. (Note) If $\operatorname{cov} e = \sigma^2 I$, then

$$\operatorname{var}(c'\hat\beta) = \sigma^2 c'(X'X)^{-1}c.$$

Pf:

$$\begin{aligned}\operatorname{var} c'\hat\beta &= \operatorname{cov} c'\hat\beta = c'(\operatorname{cov}\hat\beta)c \\ &= c'(\sigma^2(X'X)^{-1})c = \sigma^2 c'(X'X)^{-1}c\end{aligned}$$

2. (Def)

$$\begin{aligned}\widehat{\operatorname{var}}\, c'\hat\beta &= \hat\sigma^2\, c'(X'X)^{-1}c \\ \widehat{\operatorname{std}}\, c'\hat\beta &= (\widehat{\operatorname{var}}\, c'\hat\beta)^{1/2}\end{aligned}$$

3. (Th) $e \sim N(0, \sigma^2 I)$ and $X'X$ invertable imply

$$\frac{c'\hat\beta - c'\beta}{\widehat{\operatorname{std}}\, c'\hat\beta} \sim \mathcal{T}(n-p)$$

Pf:

$$\begin{aligned}\text{num} &\sim N(0, \sigma^2 c'(X'X)^{-1}c) \\ (\text{den})^2 &\sim \text{const}\ \chi^2(n-p) \text{ indep of num} \\ E(\text{den})^2 &= \sigma^2 c'(X'X)^{-1}c = \text{var num}\end{aligned}$$

Use the characterization on page 30.

4. (Note) All linear regression $t$-inferences follow from Th 3.

## 6.8 The fundamental $F$-test

1. (Restricted model) Let $\mathcal{H}$ be a $q$ dim subspace of $\mathcal{M}$

$$\begin{aligned} y &= \mu + e \ , \quad \mu \in \mathcal{H} \\ \hat{\hat{y}} &= \text{ LS est of } \mu \text{ under this model} \end{aligned}$$

2. (Scheffe's famous picture)

3. (Def) Let $H$ be the operator that projects onto $\mathcal{H}$.

4. (Lemma) $HP = H$

Pf:

$$\begin{aligned}\mathcal{M}^\perp &\subset \mathcal{H}^\perp \\ (I-H)(I-P) &= I-P \\ HP &= H\end{aligned}$$

5. (Th 6.15) If $e \sim N(0, \sigma^2 I)$

   (i) $\hat{y} - \hat{\hat{y}}$ and $\hat{e}$ are normal and independent

   (ii) $\|\hat{y} - \hat{\hat{y}}\|^2 \sim \sigma^2 \chi^2(p-q) \quad \text{if } \mu \in \mathcal{H}$

Pf: Since

$$\begin{aligned}\hat{\hat{y}} &= Hy = HPy = H\hat{y} \\ \hat{y} - \hat{\hat{y}} &= \text{fnc}\, \hat{y} \\ \text{Th 6.11} &\Longrightarrow \text{(i)}\end{aligned}$$

Now

$$(P-H)(P-H) = P - PH - HP + H = P - H$$

$\Longrightarrow P - H$ is idempotent

$\Longrightarrow P - H$ is a projection of rank $p - q$

If $\mu \in \mathcal{H}$

$$\hat{y} - \hat{\hat{y}} = (P - H)y = (P - H)(\mu + e) = (P - H)e$$

By the $\chi^2$-Th,

$$\|\hat{y} - \hat{\hat{y}}\|^2 = \|(P - H)e\|^2 \sim \sigma^2 \chi^2(p - q)$$

6. (Fundamential $F$-test) If $y \sim N(\mu, \sigma^2 I)$ and $\mu \in \mathcal{H}$

$$F = \frac{n - p}{p - q} \frac{\|\hat{y} - \hat{\hat{y}}\|^2}{\|y - \hat{y}\|^2} \sim \mathcal{F}(p - q, n - p)$$

Pf:

$$\begin{aligned} \|\hat{y} - \hat{\hat{y}}\|^2 &\sim \sigma^2 \chi^2(p - q) \\ \|y - \hat{y}\|^2 &\sim \sigma^2 \chi^2(n - p) \quad , \quad \text{(Th 6.11)} \\ \|\hat{y} - \hat{\hat{y}}\|^2 \quad &, \quad \|y - \hat{y}\|^2 \quad \text{indep} \end{aligned}$$

Use the standard characterization of the $F$-distribution.

7. (Alternate form) Let

$$\text{RSS} = \|y - \hat{y}\|^2 \quad , \quad \text{RSS}_o = \|y - \hat{\hat{y}}\|^2$$

Then

$$F = \frac{n - p}{p - q} \frac{\text{RSS}_o - \text{RSS}}{\text{RSS}}$$

Pf: From Scheffe's picture

$$\|\hat{y}-\hat{\hat{y}}\|^2 = \|\hat{\hat{e}}\|^2 - \|\hat{e}\|^2 = \text{RSS}_o - \text{RSS}$$

8. (Prob 15) From Prob 14

$$y_{ij} = \alpha_i + e_{ij}$$

$$\hat{y}_{ij} = \bar{y}_{i.}$$

$$\text{SSW} = \sum\sum(y_{ij} - \bar{y}_{i.})^2 = \|y - \hat{y}\|^2$$

$$\dim \mathcal{M} = r$$

(a) Under the hypothesis $\alpha_1 = \cdots = \alpha_r$

$$\begin{aligned} y_{ij} &= \nu + e_{ij} \quad , \quad \hat{\hat{y}}_{ij} = \bar{y}_{..} \\ \text{SSB} &= \sum\sum(\bar{y}_{i.} - \bar{y}_{..})^2 = \|\hat{y} - \hat{\hat{y}}\|^2 \\ \dim \mathcal{H} &= 1 \end{aligned}$$

Thus

$$F = \frac{n-p}{p-q}\,\frac{\|\hat{y}-\hat{\hat{y}}\|^2}{\|y-\hat{y}\|^2} = \frac{rc-r}{r-1}\,\frac{\text{SSB}}{\text{SSW}}$$

(b) By Th 6.16

$$F \sim \mathcal{F}(p-q, n-p) = \mathcal{F}(r-1, rc-r)$$

9. (Prob 16) Show

$$\begin{aligned} \hat{y}_{ij} &= (y_{ij} + y_{ij})/2 \\ y_{ij} - \hat{y}_{ij} &= \quad ? \\ \hat{\hat{y}}_{ij} &= \delta_{ij} y_{ij} \\ \hat{y}_{ij} - \hat{\hat{y}}_{ij} &= \quad ? \end{aligned}$$

## 6.9 The Gauss Markov Theorem

1. (Def)

$$\begin{aligned} \gamma &= \ell'\mu = \text{ linear parameter} \\ \hat{\gamma} &= \ell'\hat{y} = \text{ least squares est of } \gamma \end{aligned}$$

2. (Gauss-Markov Th) If $Ee = 0$ and $\operatorname{cov} e = \sigma^2 I$, then the LS estimate of any linear parameter $\gamma = \ell'\mu$ has minimum variance amoung all linear unbiased estimates of $\gamma$. (BLUE)

   Pf: Let $\tilde{\gamma} = m'y =$ any linear unbiased estimator.

$$\begin{aligned} \ell'\mu &= E\hat{\gamma} = E\tilde{\gamma} = m'\mu \quad , \quad \text{all } \mu \in \mathcal{M} \\ \ell - m &\perp \mathcal{M} \Longrightarrow P(\ell - m) = 0 \Longrightarrow P\ell = Pm \\ \operatorname{var}\hat{\gamma} &= \operatorname{var}\ell'\hat{y} = \sigma^2\ell' P\ell \\ &= \sigma^2\|P\ell\|^2 = \sigma^2\|Pm\|^2 \\ &\leq \sigma^2\|m\|^2 = \operatorname{var}(m'y) = \operatorname{var}\tilde{\gamma} \end{aligned}$$

3. (Note) If in the linear regression model $X'X$ is non-singular,

   $c'\beta$ is a linear parameter

$c'\hat{\beta}$ is its LS estimate

Pf: Let $\ell = X(X'X)^{-1}c$

$$\begin{aligned} \ell'\mu &= \ell'X\beta = c'(X'X)^{-1}X'X\beta = c'\beta \\ \ell'\hat{y} &= \ell'X\hat{\beta} = c'(X'X)^{-1}X'X\hat{\beta} = c'\hat{\beta} \end{aligned}$$

4. (Corr) If $X'X$ is non-singular, $Ee = 0$, and $\operatorname{cov} e = \sigma^2 I$, then

   $c'\hat{\beta}$ is the BLUE of $c'\beta$

5. (Note) Least squares is the best way to estimate when $\operatorname{cov} e = \sigma^2 I$.

## 6.10 Generalized least squares and weighting

1. (Generalized inner-product)

$$(u, v)_W = u'Wv \quad , \quad W \text{ pos def}$$

2. (Generalized LS) Minimize w.r.t. $\mu \in \mathcal{M}$

$$\|y - \mu\|_W^2 = (y - \mu)'W(y - \mu)$$

The value of $\mu$ that minimizes this is the GLS estimate of $\mu$. It is denoted by $\hat{y}_W$.

3. (Note) When $W$ is diagonal GLS is weighted LS

$$\|y - \mu\|_W^2 = \sum w_i(y_i - \mu_i)^2$$

4. (Comment) In this section all omitted proofs are in the text.

5. (Th) For the linear regression model

$$y = X\beta + e$$

the GLS est of $\beta$ is

$$\hat{\beta}_W = (X'WX)^{-1}X'Wy$$

6. (Gauss-Markov Th) Assume $Ee = 0$, $\operatorname{cov} e = \sigma^2 A$, and $A$ is postive definite. If $W = A^{-1}$, then $c'\hat{\beta}_W$ is the minimum variance linear unbiased estimate (BLUE) of $c'\beta$.

7. (Normal theory) If

$$y = X\beta + e \quad , \quad e \sim N(0, \sigma^2 A) \quad , \quad W = A^{-1}$$

then

$$\hat{\beta}_W \sim N(\beta, \sigma^2 (X'WX)^{-1})$$

Pf:Recall

$$\hat{\beta} = (X'WX)^{-1} X'Wy$$

Thus

$$\begin{aligned} \hat{\beta}_W & \quad \text{is normal} \\ E\hat{\beta}_W &= \beta \\ \operatorname{cov} \hat{\beta}_W &= (X'WX)^{-1} X'W(\sigma^2 A)WX(X'WX)^{-1} \\ &= \sigma^2 (X'WX)^{-1} \end{aligned}$$

8. (F-test) If $\mu \in \mathcal{H}$, $e \sim N(0, \sigma^2 A)$, and $W = A^{-1}$

$$\frac{n-p}{p-q} \frac{\|\hat{y}_W - \hat{\hat{y}}_W\|_W^2}{\|y - \hat{y}_W\|_W^2} \sim \mathcal{F}(p-q, n-p)$$

9. (Corr)If $\mu \in \mathcal{H}$ and

$$\begin{aligned}
\text{RSS} &= \|y - \hat{y}_W\|_W^2 \\
\text{RSS}_o &= \|y - \hat{\hat{y}}_W\|_W^2
\end{aligned}$$

then

$$\frac{n-p}{p-q} \frac{\text{RSS}_o - \text{RSS}}{RSS} \sim \mathcal{F}(p-q, n-p)$$

10. (Note) Computer software uses the results of this section for weighted and generalized least squares.