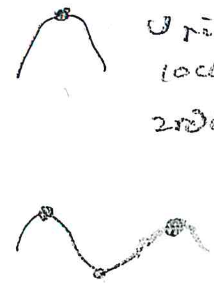
HW 5

$$\log L_p(\sigma^2) = c - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2} Q$$

let $\tau = \sigma^2$

$$= c - \frac{n}{2}\log(\tau) - \frac{1}{2\tau} Q$$

$$\frac{d \log L_p(\tau)}{d\tau} = \frac{-n}{2\tau} + \frac{Q}{2\tau^2} \overset{set}{=} 0$$



or $\quad -n\tau + Q = 0 \quad$ or $n\tau = Q$ unique

$$\hat{\tau} = \frac{Q}{n} = \hat{\sigma}^2 = \frac{\|Y - X\hat{\beta}\|^2}{n} = \frac{\sum_{i=1}^{n} e_i^2}{n}$$

$$= \frac{n-p}{n} MSE.$$

$$\frac{d^2 \log L_p(\tau)}{d\tau^2} = \frac{n}{2\tau^2} - \frac{2Q}{2\tau^3}\Big|_{\hat{\tau}} =$$

$$\frac{n}{2\hat{\tau}^2} - \frac{2n\hat{\tau}}{2\hat{\tau}^3} = \frac{n}{2\hat{\tau}^2} - \frac{2n\hat{\tau}}{2\hat{\tau}^{2}} = \frac{-n}{2\hat{\tau}^2} < 0$$

so $\hat{\tau} = \hat{\sigma}^2_{MLE}$ is the MLE

(unique 1st derivative and 2nd derivative $\big|_{\hat{\tau}} < 0$
$\Rightarrow$ global max)

So the least squares estimator of (β, σ^2) is
$(\hat{\beta}, MSE)$ while the MLE is $(\hat{\beta}, \frac{n-p}{n} MSE)$.

D3.6 25) $\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e}$ X full rank $=$ [$v_0 \cdots v_{p-1}$]. Suppose the columns are orthogonal so $\underline{v}_i^T \underline{v}_j = 0$ $i \neq j$.

Then $X'X = \operatorname{diag}(\underline{v}_i^T \underline{v}_i)$ so

$$(X'X)^{-1} = \operatorname{diag}\left(\frac{1}{\underline{v}_i^T \underline{v}_i}\right) \text{ and}$$

$$\underline{\hat{\beta}} = (X'X)^{-1} X' \underset{\sim}{Y} = \begin{pmatrix} (\underline{v}_0'\underline{v}_0)^{-1} \underline{v}_0' \underline{Y} \\ \vdots \\ (\underline{v}_{p-1}'\underline{v}_{p-1})^{-1} \underline{v}_{p-1}' \underline{Y} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_{p-1} \end{pmatrix}$$

$$\text{So } \hat{\beta}_j = \frac{\underline{v}_j' \underset{\sim}{Y}}{\underline{v}_j' \underline{v}_j}, \quad j = 0, \ldots, p-1$$

26) Under the orthogonal column model of 25), the least squares estimate of $\hat{\beta}_j$ remains unchanged if columns of X other than $\underline{v}_j$ are deleted.

LM

27] Orthogonal designs are often used in experimental design.

Skip §3.7, 3.8 Nonfull rank model

§3.9 and earlier results in ch 3

28] p. 63 Now suppose $\underset{\sim}{Y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon}$

but $\underset{n\times p}{X}$ has rank $r < p$.

Then $P_X = X(X'X)^- X'$ is the unique projection matrix on $C(X)$ and does not depend on $(X'X)^-$.

iii) $\hat{\underline{Y}} = X\hat{\underline{\beta}} = P_X \underset{\sim}{Y}$, $\underline{e} = \underline{Y} - X\hat{\underline{\beta}} = (I - P_X)\underset{\sim}{Y}$ and

RSS $= \underline{e}'\underline{e}$ are unique and so do not depend on $(X'X)^-$.

swap

ii) $\hat{\underline{\beta}} = (X'X)^- X'\underline{Y}$ does depend on $(X'X)^-$ and is not unique.

iv) $\hat{\underline{\beta}}$ is a solution to the normal equations: $X'X\hat{\underline{\beta}} = X'\underbrace{X(X'X)^- X'}_{P_X}\underline{Y} = X'\underline{Y}$

Since $P_X \mathbf{X} = \mathbf{X} \Rightarrow \mathbf{X}' P_X = \mathbf{X}'$

v) p41 rank $(I - P_X) = n - r$

vi) p42 Let $\hat{\underline{\theta}} = \mathbf{X}\hat{\underline{\beta}}$ and $\underline{\theta} = \mathbf{X}\underline{\beta}$.

Among the class of linear unbiased estimators of $\underline{c}'\underline{\theta}$, $\underline{c}'\hat{\underline{\theta}}$ is BLUE

(but need $E[\underline{c}'\hat{\underline{\theta}}] = \underline{c}'\underline{\theta}$),

vii) p44 If $Cov(\underline{Y}) = \sigma^2 I_n$, then

$$MSE = \frac{RSS}{n-r} = \frac{e'e}{n-r}$$ is an unbiased

estimator of σ^2.

viii) p63 Let $\underset{n \times r}{\mathbf{X}_1}$ be the matrix formed by r linearly independent columns of $\mathbf{X}$ or by a basis for $C(\mathbf{X})$. Then

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

39) * p64 In the non full rank $\mathbf{X}$ model rank $(\mathbf{X}) = r < p$, $\hat{\underline{\beta}}$ is not unique so $\underline{\beta}$ is not estimable.

30) * p64 $\underline{a}'\underline{\beta}$ is <u>estimable</u> LM 41

if it has a linear unbiased estimator $\underline{b}'\underline{Y}$ so $E\, \underline{b}'\underline{Y} = \underline{a}'\underline{\beta}$.

31) The term "estimable" is misleading since there are nonestimable quantities $\underline{a}'\underline{\beta}$ that can be estimated with biased estimators.

32) * p64 i) $\underline{a}'\underline{\beta}$ is estimable iff $\underline{a}' = \underline{b}'X$ iff $\underline{a} = X'\underline{b}$ iff $\underline{a} \in C(X')$.

ii) If $\underline{a}'\underline{\beta}$ is estimable and a least squares estimator $\hat{\underline{\beta}}$ is any solution to the normal equations $X'X\hat{\underline{\beta}} = X'\underline{Y}$ then $\underline{a}'\hat{\underline{\beta}}$ is unique and $\underline{a}'\hat{\underline{\beta}}$ is the BLUE of $\underline{a}'\underline{\beta}$.

33) p65 Can avoid nonestimable functions by using a full rank model instead of a nonfull rank model. Delete columns of $\tilde{X}$, say, resulting in a full rank matrix X.

Skip § 3.9.3, 3.9.4

34) If $\hat{\theta}$ is an unbiased estimator of θ then $E\hat{\theta} = \theta \ \forall \theta$ in the parameter space.

35) Gauss Markov Theorem also see 11)

Let $Y = X\beta + \epsilon$, X full rank, $E\epsilon = 0$ and $\mathrm{cov}(\epsilon) = \sigma^2 I$. Then $a'\hat{\beta}$ is the BLUE of $a'\beta$ for every constant $p \times 1$ vector a.

(Some say $\hat{\beta}$ is the BLUE of β.)

Proof) $a'\hat{\beta} = \underbrace{a'(X'X)^{-1}X'}_{d'} Y = d'Y$

is linear and $V(a'\hat{\beta}) = V(d'Y) = \sigma^2 d'd$ $= \mathrm{cov}(a'\hat{\beta}) = a'\mathrm{cov}(\hat{\beta})a$

$= a'(X'X)^{-1}X'\sigma^2 I X(X'X)^{-1}a = \sigma^2 a'(X'X)^{-1}a$

Let $\underset{\sim}{c}'\underset{\sim}{Y}$ be any other linear unbiased estimator of $\underset{\sim}{a}'\underset{\sim}{\beta}$. Then

$$E(\underset{\sim}{c}'\underset{\sim}{Y}) = \underset{\sim}{a}'\underset{\sim}{\beta} = \underset{\sim}{c}' E\underset{\sim}{Y} = \underset{\sim}{c}' X \underset{\sim}{\beta},$$

for any $\underset{\sim}{\beta} \in \mathbb{R}^p$ the parameter space of β. Hence $\underline{a}' = \underline{c}' X$.

Now $cov(\underline{c}'\underline{Y}, \underline{a}'\hat{\underline{\beta}}) = cov(\underline{c}'\underline{Y}, \underline{d}'\underline{Y})$

$$= \underset{\sim}{c}' \, cov(\underline{Y}) \underset{\sim}{d} = \sigma^2 \underline{c}'\underline{d}$$

$$= \sigma^2 \underbrace{\underset{\sim}{c}' X}_{\underline{a}'} \overbrace{(X'X)^{-1} \underset{\sim}{a}}^{\underline{d}} = \sigma^2 \underline{a}'(X'X)^{-1}\underline{a}$$

$$= V(\underline{a}'\hat{\underline{\beta}}) = V(\underline{d}'\underline{Y}) \text{ by } (*). \quad (**)$$

Now $\underline{0 \leq} V(\underline{c}'\underline{Y} - \underset{\sim}{a}'\hat{\underline{\beta}}) = V(\underline{c}'\underline{Y} - \underline{d}'\underline{Y})$

$$= V(\underline{c}'\underline{Y}) + V(\underline{d}'\underline{Y}) - 2\, cov(\underline{c}'\underline{Y}, \underline{d}'\underline{Y})$$

$$V(Z-X) = V(Z) + V(X) - 2\, cov(Z,X) \quad (= cov(Z-X, Z-X)),$$

by (**) $= V(\underline{c}'\underline{Y}) + V(\underline{d}'\underline{Y}) - 2V(\underline{d}'\underline{Y}) =$

$\underline{V(\underline{c}'\underline{Y}) - V(\underline{d}'\underline{Y})}$. So $0 \leq V(\underline{c}'\underline{Y}) - V(\underline{d}'\underline{Y})$

or $V(\underset{\sim}{c}'\underset{\sim}{Y}) \ge V(\underset{\sim}{d}'\underset{\sim}{Y}) = V(\underset{\sim}{a}'\hat{\underset{\sim}{\beta}})$.

$\boxed{//}$

§3.10 38) The <u>least squares</u> (LS) or <u>ordinary least squares</u> (OLS) model is $\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e}$, $\underset{\sim}{X}$ full rank

$E(\underset{\sim}{e}) = \underset{\sim}{0}$, $\quad cov(\underset{\sim}{e}) = \sigma^2 \underset{\sim}{I}$.

38) The <u>generalized least squares</u> (GLS) model is $\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e}$, $\underset{\sim}{X}$ full rank,

$E(\underset{\sim}{e}) = \underset{\sim}{0}$, $\quad cov(\underset{\sim}{e}) = \sigma^2 \underset{\sim}{V} > 0$,

so $\underset{\sim}{V} = \underset{\sim}{V}^T$ is positive definite is <u>known</u>.

38) The <u>weighted least squares</u> (WLS) model is the special case of the GLS model where $\underset{\sim}{V} = diag(v_1, \ldots, v_n)$ with weights $w_i = 1/v_i$ for $i = 1, \ldots, n$.

39) The feasible generalized least squares (FGLS) model is the same as the GLS model except $\underset{\sim}{V} = V(\underset{\sim}{\theta})$ is a function of an unknown $q \times 1$ vector of parameters. An estimator of V is $\hat{V} = V(\hat{\underset{\sim}{\theta}})$.

matrix not variance

The feasible weighted least squares (FWLS) model is a special case where $V = V(\underset{\sim}{\theta})$ is diagonal. Then the estimated weights

$$\hat{w}_i = \frac{1}{\hat{v}_i} = \frac{1}{v_i(\hat{\underset{\sim}{\theta}})}.$$

40) The GLS or WLS estimator is

$$\hat{\underset{\sim}{\beta}}_{WLS} \text{ or } \hat{\underset{\sim}{\beta}}_{GLS} = (X^T V^{-1} X)^{-1} X^T V^{-1} \underset{\sim}{Y}$$

The FGLS or FWLS estimator is

$$\hat{\underset{\sim}{\beta}}_{FWLS} \text{ or } \hat{\underset{\sim}{\beta}}_{FGLS} = (X^T \hat{V}^{-1} X)^{-1} X^T \hat{V}^{-1} \underset{\sim}{Y}.$$

The fitted values are $\hat{Y}_{AB} = X \hat{\beta}_{AB}$ where

ABI is GLS, FGLS, WLS or FWLS.

4.1) It can be shown that the GLS estimator minimizes the criterion $Q_{GLS}(\underset{\sim}{\eta}) = (\underset{\sim}{Y} - X\underset{\sim}{\eta})^T V^{-1} (\underset{\sim}{Y} - X\underset{\sim}{\eta})$.

4.2) Seber and Lee P 66-68 transform GLS into OLS using an $n \times n$ nonsingular matrix K such that $V = K K^T$. K is found using the Cholesky decomposition.

4.3) Th) Let $V = RR$ where $R = V^{\frac{1}{2}} = R^T > 0$. Let $\underset{\sim}{Z} = R^{-1}\underset{\sim}{Y}$, $U = R^{-1}X$ and $\underset{\sim}{a} = R^{-1}\underset{\sim}{e}$.

R is found using the spectral theorem (singular value decomposition).

a) $\underset{\sim}{Z} = U\underset{\sim}{\beta} + \underset{\sim}{a}$ follows the OLS model since $E\underset{\sim}{a} = \underset{\sim}{0}$ and $Cov(\underset{\sim}{a}) = \sigma^2 I_n$

b) The GLS estimator $\hat{\beta}_{GLS}$ can be obtained from the OLS regression