Cor 6.3.2 Under the conditions of Thm 6.3.2, there exists a nonsingular matrix Q such that $QH = P$, where $PP' = I$ and hence the GLR test of $H_0: H\underline{\beta} = \underline{h}$ vs $H_1: H\underline{\beta} \neq \underline{h}$ is equivalent to the GLR test of $H_0: P\underline{\beta} = \underline{k}$ vs $H_a: P\underline{\beta} \neq \underline{k}$, where $Q\underline{h} = \underline{k}$ and $PP' = I$.

no proof provided actually in class — matrix theory backgr

Proof: For any $q \times p$ matrix H of rank q, $\exists$ nonsingular matrix Q such that $P = QH$ and $PP' = I$, that is, P is the first q rows of an orthogonal matrix.

Then, $H\underline{\beta} = \underline{h} \implies QH\underline{\beta} = Q\underline{h} = k$

$\implies P\underline{\beta} = k$

Then, the parameter space W under $H_0: H\underline{\beta} = \underline{h}$ will be the same as that of $H_0: P\beta = \underline{k}$

$\implies$ LRT test will be equivalent.

Sec 6.4 Special cases for Hypothesis testing of $H_0: H\beta = h$.

(testing a contrast, etc) 1. Testing a linear combination of β:

⟹ Testing $H_0: \ell'\beta = l_0$ vs $H_1: \ell'\beta \neq l_0$ where ℓ is a given $p\times 1$ vector and l_0 is a constant.

In this case, Thm 6.3.1 can be easily applied, noting that $H = \ell'$, $h = l_0$, $q = 1$. In this case, the LRT W (given in (6.3.9)) is:

$$\hat{\sigma}^2_\Omega = \frac{Y'(I-XX^-)Y}{n-p}$$

$$W = \frac{(\ell'\hat\beta - l_0)'[\ell'(X'X)^{-1}\ell]^{-1}(\ell'\hat\beta - l_0)}{Y'(I-XX^-)Y}(n-p)$$

scalar

$$C = (X'X)^{-1}$$

$$= \frac{(\ell'\hat\beta - l_0)'[\ell' C \ell]^{-1}(\ell'\hat\beta - l_0)}{\hat\sigma^2_\Omega}$$

$$= \frac{(\ell'\hat\beta - l_0)^2}{\hat\sigma^2_\Omega(\ell' C \ell)} = \frac{(\ell'\hat\beta - l_0)^2}{\widehat{Var}(\ell'\hat\beta)}$$

$$\lambda_1 = \frac{(\ell'\beta - l_0)^2}{\sigma^2(\ell' C \ell)}$$

$\sim F_{1,n-p,\lambda_1}$ and $W \overset{H_0}{\sim} F_{1,n-p}$

H_0 is rejected if $W \geq F_{\alpha,1,n-p} (= t^2_{\alpha/2,n-p})$

⟹ H_0 is rejected if $\dfrac{(\ell'\hat\beta - l_0)^2}{\hat\sigma^2(\ell' C \ell)} \geq t^2_{\alpha/2,n-p}$

⟹ or Don't reject H_0 if and only if l_0 is in the confidence interval ($1-\alpha$ level):

$$\left(\ell'\hat\beta - t_{\alpha/2,n-p}\sqrt{\hat\sigma^2 \ell' C \ell},\ \ell'\hat\beta + t_{\alpha/2,n-p}\sqrt{\hat\sigma^2(\ell' C \ell)}\right) \quad (6.4.2)$$

Remark It's been shown that the test above is an UMPU test of size α for $H_0: \ell'\beta = l_0$ vs $H_1: \ell'\beta \neq l_0$

uniformly most accurate unbiased

and the above C.I is a UMAU ~~1-α~~ confidence interval on $\ell'\beta$.

2. Testing a subset of the parameter vector β.

specifically.

$\Rightarrow$ We want to test $H_0: \beta_2 = b_2$ vs $H_1: \beta_2 \neq b_2$,

where
$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \begin{array}{l} \leftarrow (p-q)\times 1 \\ \leftarrow q \times 1 \end{array} \qquad q < p$$

Let the matrix X be partitioned accordingly as.
$$X = [\underset{n\times(p-q)}{X_1} \quad \underset{n\times q}{X_2}]$$

Then, the GLM
$$\begin{aligned} Y &= X\beta + \varepsilon \\ &= [X_1 \quad X_2]\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \varepsilon \\ &= X_1\beta_1 + X_2\beta_2 + \varepsilon \end{aligned}$$

Go to next page two

Testing $\beta_2 = b_2$ can be thought as testing $H\beta = h$

with
$$H = [\underset{q\times(p-q)}{0} \quad I_q] \quad \text{and} \quad h = b_2$$

Let $(X'X)^{-1} = C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$ be partitioned accordingly. $C = (X'X)^{-1}$

$$\Rightarrow H(X'X)^{-1}H' = HCH' = [0 \quad I_q]\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}\begin{bmatrix} 0 \\ I_q \end{bmatrix} = C_{22}$$

$\Rightarrow$ The test statistic W ((6.3.9)) is now.

$$W = \frac{(H\hat{\beta} - h)'[H(X'X)^{-1}H']^{-1}(H\hat{\beta} - h)}{Y'(I - XX^{-})Y}\;\frac{n-p}{q}$$
$$= \frac{(\hat{\beta}_2 - b_2)'C_{22}^{-1}(\hat{\beta}_2 - b_2)}{q\,\hat{\sigma}^2} \sim F'_{q,\,n-p;\,\lambda}$$

with $\lambda = \dfrac{(\underline{\beta}_2 - \underline{b}_2)' C_{22}^{-1} (\underline{\beta}_2 - \underline{b}_2)}{2\sigma^2}$

and $W \overset{H_0}{\sim} F_{q, n-p}$, where

$\hat{\underline{\beta}}_2$ is the last q elements of the MLE: $\hat{\underline{\beta}} = X'\underline{Y}$

Further note that from (Thm 1.3.1, inverse of partition matrix), we can write:

$$C_{22}^{-1} = S_{22} - S_{21} S_{11}^{-1} S_{12} = X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2$$

If we use (6.3.8) to express the test statistic
In this case, there will be computational advantages.
For $\underline{b}_2 = \underline{0}$ case, see (6.4.6) in the book

In chapter 7, we will study some computational details.

Sec 6.5 Confidence Interval Associated with the test $H_0: H\beta = \underline{h}$

We have seen earlier how acceptance of H_0 is equivalent to a confidence interval estimation. If we let $\underline{\theta} = H\beta - \underline{h}$, then $H_0: H\beta = \underline{h}$ vs H_1 can be written as : $H_0: \underline{\theta} = \underline{0}$ vs $H_1: \underline{\theta} \neq \underline{0}$

The test statistic

$$W = \frac{(H\hat{\beta} - \underline{h})'[H(X'X)^{-1}H']^{-1}(H\hat{\beta} - \underline{h})}{q\hat{\sigma}^2}$$ re-written (6.3.9)

$$= \frac{\frac{1}{q}(\hat{\underline{\theta}}' V^{-1} \hat{\underline{\theta}})}{\hat{\sigma}^2}$$

where $\hat{\underline{\theta}} = H\hat{\beta} - \underline{h} = HX^{-}\underline{Y} - \underline{h}$, $V = H(X'X)^{-1}H'$

$\hat{\sigma}^2 = \underline{Y}'(I - XX^{-})\underline{Y}/(n-p)$

H_0 is rejected if $W \geq F_{\alpha, q, n-p}$

or H_0 is accepted if only if $W < F_{\alpha, q, n-p}$.

the later leads to our desire to derive confidence intervals for $\underline{\delta} = H\beta$ or $\delta_1, \dots, \delta_q$, or a linear combination of the δ_i's, $i = 1, \dots, q$.

In general, we have two types of C.I.'s:

1. Individual (one-at-a-time) C.I.s on each δ_i:

In this case, $\delta_i = \underline{k}_i'\beta$, $\underline{k}_i'$ is the ith row of H, $i = 1, \dots, q$

Similar to (6.4.2), we can easily obtain the $1-\alpha$ C.I for θ_i as:

where $\widehat{Var}(\underline{l}_i'\hat{\beta}) = \hat{\sigma}^2 \underline{l}_i' C \underline{l}_i$

$$\underline{k}_i'\hat{\beta} \mp t_{\alpha/2, n-p} \sqrt{\widehat{Var}[\underline{k}_i'\hat{\beta}]}, \quad i = 1, \dots, q \quad (6.5.2)$$

s.t. $P(\delta_i \in C.I.) = 1 - \alpha$, $i = 1, \dots, q$

These intervals (6.5.2) are individual $(1-\alpha)$ level C.I.s

<u>Remark</u>. C.I.s given in (6.4.2) are also one-at-a-time C.I for $\underline{l}'\beta$ for any coefficient $\underline{l}$.

2. Simultaneous C.I.s

The confidence intervals, I_i's, are called $1-\alpha$ level SCIs for all δ_i if $P[\delta_i \in I_i,\ i=1,\cdots,q] = 1-\alpha$

We will derive SCIs for δ_i, $i=1,\cdots,q$ here

We first state the following Thm ~~without proof~~.

Thm 6.5.1 The test statistic W in equation (6.3.9) is equal to W^*, where W^* is given by

$$W^* = \frac{1}{q\hat{\sigma}^2} \max_{\underset{\sim}{\ell} \text{ in } E_q^*} \left[\frac{[\underset{\sim}{\ell}'(H\hat{\beta} - \underset{\sim}{h})]^2}{\underset{\sim}{\ell}'[H(X'X)^{-1}H']\underset{\sim}{\ell}} \right] \quad (6.5.3)$$

where E_q^* is the q-dimentional vector space E_q with $\underset{\sim}{0}$ removed.

< see 6-32 ① + ② for proof >

Using the above Thm, we can establish the following SCIs on $\ell'(H\beta)$ or $\ell'\delta$

Thm 6.5.2 For the GLM $\underset{\sim}{Y} = X\beta + \varepsilon$, $\varepsilon \sim N_n(\underset{\sim}{0}, \sigma^2 I)$

The set of confidence intervals on $\ell'(H\beta) = \ell'\delta$ ~~given by~~:

$$SCI(\underset{\sim}{\ell}) = \underset{\sim}{\ell}'(H\hat{\beta}) \mp \underbrace{\sqrt{qF_{\alpha,q,n-p}}}_{S} \underbrace{\sqrt{\hat{\sigma}^2 \underset{\sim}{\ell}'[H(X'X)^{-1}H']\underset{\sim}{\ell}}}_{\hat{Var}[\ell'(H\hat{\beta})]} \quad (6.5.6)$$

are SCIs for $\ell'(H\beta)$ with level $1-\alpha$, where $\underset{\sim}{\ell} \in E_q^*$, H is $q\times p$ of rank q

Proof: We know that when $H_0: H\beta = \underset{\sim}{h}$ is true

$$P[W \le F_{\alpha, q, n-p}] = 1-\alpha$$

From Thm 6.5.1, $W = W^*$

$$\Rightarrow 1-\alpha = P[W^* \le F_{\alpha, q, n-p}]$$

$$= P\left[\max_{\underline{\ell} \in E_q^*} \left\{ \frac{[\underline{\ell}' H(\hat{\underline{\beta}} - \underline{\beta})]^2}{\underline{\ell}'[H(X'X)^{-1}H']\underline{\ell}} \frac{1}{q\hat{\sigma}^2} \right\} \le F_{\alpha, q, n-p}\right]$$

$$= P\left[\frac{[\underline{\ell}' H(\hat{\underline{\beta}} - \underline{\beta})]^2}{\underline{\ell}'[H(X'X)^{-1}H']\underline{\ell}} \frac{1}{q\hat{\sigma}^2} \le F_{\alpha, q, n-p}, \text{ for all } \underline{\ell} \in E_q^* \right]$$

$$= P\left[\underline{\ell}'H\hat{\underline{\beta}} - \sqrt{qF}\sqrt{\underline{\ell}'[H(X'X)^{-1}H']\underline{\ell}} \le \underline{\ell}'(H\underline{\beta}) \le \underline{\ell}'H\hat{\underline{\beta}} + \underbrace{\sqrt{qF_{\alpha,q,n-p}}}_{S}\underbrace{\sqrt{\hat{\sigma}^2\underline{\ell}'[H(X'X)^{-1}H']\underline{\ell}}}_{\widehat{Var}[\underline{\ell}'(H\hat{\underline{\beta}})]}\right]$$

$$= P[SCI(\underline{\ell}) \ni \underline{\ell}'(H\underline{\beta}), \ \forall \underline{\ell} \in E_q^*].$$

Note. ① Thm 6.5.2 gives infinite many SCIs.

② The SCIs in Thm 6.5.2 are named as Scheffé SCIs.

③ The SCIs in (6.5.6) can be written as

$$\underline{\ell}'(H\hat{\underline{\beta}}) \mp S\sqrt{\widehat{Var}[\underline{\ell}'(H\hat{\underline{\beta}})]}. \qquad (6.5.10)$$

④. You don't have to use all infinite many SCIs as in (6.5.2). ~~For instance, If~~ In that case, the SCI level is at least $1-\alpha$.

Sec 6.8 The GLM under Normality when $\Sigma \neq \sigma^2 I$.

Now, we consider GLM: $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$
with the assumption that: $\underline{\varepsilon} \sim N_n(\underline{0}, \Sigma)$, $\Sigma \neq \sigma^2 I$

1. When $\Sigma = \sigma^2 V$, V is known and p.d.

As V is p.d., there exists an $n \times n$ nonsingular matrix G s.t. $V = G'G$ (Thm 1.4.2 (2a)). $V^{-1} = G^{-1}G'^{-1}$

Let $\underline{z} = G'^{-1}\underline{y}$, then:

$$\underline{z} \sim N_n(G'^{-1}X\underline{\beta}, \sigma^2 G'^{-1}VG^{-1}) = N_n(G'^{-1}X\underline{\beta}, \sigma^2 G'^{-1}G'GG^{-1})$$
$$= N_n(G'^{-1}X\underline{\beta}, \sigma^2 I).$$

Let $A = G'^{-1}X$, and $\underline{\eta} = G'^{-1}\underline{\varepsilon}$, then the GLM, after multiplying G'^{-1} to it, becomes:

$$G'^{-1}\underline{y} = G'^{-1}X\underline{\beta} + G'^{-1}\underline{\varepsilon}$$

or $\underline{z} = A\underline{\beta} + \underline{\eta}$ with $\underline{\eta} \sim N_n(\underline{0}, \sigma^2 I)$

Note: $A'A = X'G^{-1}G'^{-1}X = X'V^{-1}X$

Then testing of $H_0: H\underline{\beta} = \underline{h}$ vs $H_1: H\underline{\beta} \neq \underline{h}$ can be done by using GLM $\underline{z} = A\underline{\beta} + \underline{\eta}$ with $\underline{\eta} \sim N_n(0, \sigma^2 I)$

The test statistic W given by (6.3.9) is then:

Problem 6.25 #

$$W = \frac{n-p}{q} \frac{(H\hat{\underline{\beta}} - \underline{h})'[H(A'A)^{-1}H']^{-1}(H\hat{\underline{\beta}} - \underline{h})}{\underline{z}'(I - AA^{-})\underline{z}}$$

$$\underset{\hat{\underline{\beta}} = (A'A)^{-1}A'\underline{z}}{\underline{\underline{\text{using}}}} \frac{n-p}{q} \frac{[H(A'A)^{-1}A'\underline{z} - \underline{h}]'[H(A'A)^{-1}H']^{-1}[H(A'A)^{-1}A'\underline{z} - \underline{h}]}{\underline{z}'(I - AA^{-})\underline{z}}$$

$$\underset{A = G'^{-1}X}{\overset{\text{using } \underline{z} = G'^{-1}\underline{y}}{=\!=\!=}} \frac{n-p}{q} \frac{[HX'V^{-1}XX'G^{-1}G'^{-1}\underline{y} - \underline{h}]'[H(X'V^{-1}X)^{-1}H']^{-1} \cdot [H(X'V^{-1}X)^{-1}X'V^{-1}\underline{y} - \underline{h}]}{\underline{y}'G^{-1}(I - G'^{-1}X(X'V^{-1}X)^{-1}X'G^{-1})G'^{-1}\underline{y}}$$

$$= \frac{n-p}{q} \frac{[H(X'V^{-1}X)^{-1}X'V^{-1}\underline{y} - \underline{h}]'[H(X'V^{-1}X)^{-1}H']^{-1}[H(X'V^{-1}X)^{-1}X'V^{-1}\underline{y} - \underline{h}]}{\underline{y}'(V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1})\underline{y}}$$

$W \sim F_{q, n-p, \lambda}$ with $\lambda = \frac{1}{2\sigma^2}(H\underline{\beta} - \underline{h})'[H(A'A)^{-1}H']^{-1}(H\underline{\beta} - \underline{h})$

$= \frac{1}{2\sigma^2}(H\underline{\beta} - \underline{h})'[H(X'V^{-1}X)^{-1}H']^{-1}(H\underline{\beta} - \underline{h})$

From above, we can also write:

$$\hat{\underline{\beta}} = (A'A)^{-1}A'\underline{z} = (X'V^{-1}X)^{-1}X'V^{-1}\underline{y}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}\underline{y}'[V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}]\underline{y}$$

$$\begin{aligned} Cov(\hat{\underline{\beta}}) &= (X'V^{-1}X)^{-1}X'V^{-1}\,Cov(\underline{y})\,[(X'V^{-1}X)^{-1}X'V^{-1}]' \\ &= (X'V^{-1}X)^{-1}X'V^{-1}\sigma^2 V\, V'^{-1}X(X'V^{-1}X)^{-1} \qquad (V' = V) \\ &= \sigma^2(X'V^{-1}X)^{-1}X'V^{-1}X(X'V^{-1}X)^{-1} \\ &= \sigma^2(X'V^{-1}X)^{-1} \end{aligned}$$

Here $\hat{\underline{\beta}}$ and $\hat{\sigma}^2$ are UMVUE's for $\underline{\beta}$ and σ^2 according to Thm 6.2.2.

2. The GLM: Point estimation when Σ is unknown (but p.d.)
This case is more complicated. Let's have the following definition of OLS estimators first.

Definition 6.8.1 Ordinary Least Squares (OLS) Estimators

Consider the model $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, with $E(\underline{\varepsilon}) = \underline{0}$, $Cov(\underline{\varepsilon}) = \Sigma$

For this model, $\tilde{\underline{\beta}} = (X'X)^{-1}X'\underline{y}$ is defined to be the OLS estimator of $\underline{\beta}$, and for any constant vector $\underline{\ell}$ of $p\times 1$, the OLS estimator of $\underline{\ell}'\underline{\beta}$ is defined to be $\underline{\ell}'(X'X)^{-1}X'\underline{y}$.

Note: $\tilde{\underline{\beta}}$ is obtained through the ordinary "least-square" principle, that is $\tilde{\underline{\beta}}$ is such that $(\underline{y} - X\tilde{\underline{\beta}})'(\underline{y} - X\tilde{\underline{\beta}})$ is minimized.

The following Theorem discusses when an UMVUE of $\underset{\sim}{\beta}$ exists for GLM: $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ with $\underset{\sim}{\varepsilon} \sim N_n(0, \Sigma)$.

Thm 6.8.1 For GLM: $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ with $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \Sigma)$. The UMVUE of $\underset{\sim}{\beta}$ is the OLS estimator iff there exists a $p \times p$ nonsingular matrix F s.t.

$$\Sigma X = XF \qquad (6.8.5)$$

Proof: For every known Σ, the UMVUE estimator of $\underset{\sim}{\beta}$ is given by:

$$(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underset{\sim}{y}.$$

The OLS estimator of $\underset{\sim}{\beta}$ is given by $(X'X)^{-1}X'\underset{\sim}{y}$

So, we eventually want to show:

$$(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underset{\sim}{y} = (X'X)^{-1}X'\underset{\sim}{y} \text{ for any } \underset{\sim}{y}.$$

That is, to show:

$$(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} = (X'X)^{-1}X' \qquad (*)$$

for any $\underset{\sim}{y}$.

Assume the above equation $(*)$ is true.

multiply $X'X$ to the left of both sides $\longrightarrow$ $X'X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} = X'$

$\Longrightarrow X = \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'X \qquad (\Sigma' = \Sigma)$

$\Longrightarrow \Sigma X = X\underbrace{(X'\Sigma^{-1}X)^{-1}X'X}_{F}$

$= XF$

F is $p\times p$ and nonsingular.

Now, assume $\Sigma X = XF$, reverse about to get:

$X' = X'X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$

with $F = (X'\Sigma^{-1}X)^{-1}X'X$

$\Longrightarrow F' = X'X(X'\Sigma^{-1}X)^{-1}$

$F'^{-1} = X'\Sigma^{-1}X(X'X)^{-1}$

$\Longrightarrow F'^{-1}X'X = X'\Sigma^{-1}X \qquad ①$

and $F'^{-1}X' = X'\Sigma^{-1} \qquad ②$

Then, $(X'X)^{-1}X' = (X'X)^{-1}\underbrace{F'F'^{-1}}_{I}X' = (F'^{-1}X'X)^{-1}F'^{-1}X'$

$\overset{①②}{=} (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} \Longrightarrow (*)$ holds. //

<u>Cor 6.8.1.1</u> Consider the GLM model in Thm 6.8.1 and let Σ be defined by $\Sigma = X\Delta_1X' + (I-XX^-)\Delta_2(I-XX^-) + \theta I$ where Δ_1, Δ_2 and θ are restricted only so that Σ is p.d. and $\Delta_1XX' + \theta I$ is nonsingular. Then the UMVU estimator of $\underline{\ell}'\beta$ is equal to the OLS estimator of $\underline{\ell}'\beta$ for any $p\times 1$ constant vector $\underline{\ell}$.

Proof: $\Sigma X = X\Delta_1X'X + (I-XX^-)\Delta_2(I-XX^-)X + \theta IX$

$= X\Delta_1X'X + (I-XX^-)\Delta_2(X - \underbrace{XX^-X}_{X})+\theta X$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad \underbrace{\qquad\qquad}_{0}$

$= X\Delta_1X'X + \theta X$

$= X(\Delta_1X'X + \theta I)$

As $\Delta_1 X'X + \theta I$ is nonsingular. Let $F = \Delta_1 X'X + \theta I$

Then: $\Sigma X = XF$

$\overset{\text{Thm 6.8.1}}{\Longrightarrow}$ The UMVU estimator of β (in $\underline{Y} = X\beta + \underline{\varepsilon}$) is given by $(X'X)^{-1}X'Y$ (the OLS estimator).

$\Longrightarrow$ For a constant vector $\underline{\ell}$, $\underline{\ell}'(X'X)^{-1}X'\underline{Y}$ is the OLS estimator and it is the UMVU estimator of $\underline{\ell}'\beta$. //.

<u>Cor 6.5.1.2</u> For the GLM: $\underline{Y} = X\beta + \underline{\varepsilon}$, $\underline{\varepsilon} \sim N_n(\underline{0}, \Sigma)$, $X = [\underline{1}, x_2, \cdots, x_p]$, that is, the model has an intercept term, say β_0. If Σ is defined by $\Sigma = \alpha^2(1-\rho)I + \alpha^2\rho J$ for $-1/(n-1) < \rho < 1$, then $\underline{\ell}'\hat{\beta} = \underline{\ell}'(X'X)^{-1}X'\underline{Y}$ is the UMVU estimator of $\underline{\ell}'\beta$ for any $p \times 1$ constant vector $\underline{\ell}$

Proof: Let (In Cor 6.5.1.1) $\Delta_2 = 0$, $\Delta_1 = \begin{pmatrix} \alpha^2\rho & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{p\times p}$ and $\theta = \alpha^2(1-\rho)$

and then

$$X\Delta_1 X' + \theta I$$

$$= \begin{bmatrix} 1 & x_{12} & \cdots & x_{1p} \\ 1 & x_{22} & & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n2} & & x_{np} \end{bmatrix} \Delta_1 X' + \theta I$$

$$= \begin{bmatrix} \alpha^2\rho & 0 & \cdots & 0 \\ \alpha^2\rho & 0 & \cdots & 0 \\ \vdots & & & \\ \alpha^2\rho & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & & & \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix} + \alpha^2(1-\rho)I$$

$$= \begin{bmatrix} \alpha^2\rho & \alpha^2\rho & \cdots & \alpha^2\rho \\ \alpha^2\rho & \alpha^2\rho & & \alpha^2\rho \\ \alpha^2\rho & \alpha^2\rho & \cdots & \alpha^2\rho \end{bmatrix} + \alpha^2(1-\rho)I$$

$$= \alpha^2\rho J + \alpha^2(1-\rho)I = \alpha^2 \begin{bmatrix} 1 & \rho & \rho \\ \rho & \ddots & \\ & & 1 \end{bmatrix}$$

$$= \Sigma$$

To ensure Σ be p.d., we must have $-\frac{1}{n-1} < \rho < 1$ ... ①

see proof for on 6-38

Let $F = \Delta_1 X'X + \alpha^2(1-\rho)I$, F is $p \times p$

then $$XF = \underline{X\Delta_1 X'}X + \alpha^2(1-\rho)X$$
$$= \alpha^2\rho J X + \alpha^2(1-\rho)IX$$
$$= [\alpha^2\rho J + \alpha^2(1-\rho)I]\, X$$
$$= \Sigma X$$

To ensure F be nonsingular, let's see:

$$F = \Delta_1 X'X + \alpha^2(1-\rho)I = \begin{pmatrix} \alpha^2\rho & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{12} & x_{22} & \cdots & x_{n2} \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{pmatrix}\begin{pmatrix} 1 & x_{12} & \cdots \\ 1 & x_{22} & \cdots \\ & x_{n} & \cdots \end{pmatrix}$$
$$+ \alpha^2(1-\rho)I$$

$$= \begin{pmatrix} \alpha^2\rho & \alpha^2\rho & \cdots & \alpha^2\rho \\ 0 & 0 & & 0 \\ 0 & 0 & & 0 \end{pmatrix}\begin{pmatrix} 1 & x_{12} & \cdots & x_{1p} \\ 1 & x_{22} & \cdots & x_{2p} \\ 1 & x_{n2} & \cdots & x_{np} \end{pmatrix} + \alpha^2(1-\rho)I$$

$$= \begin{pmatrix} n\alpha^2\rho & \alpha^2\rho\sum_{j=1}^{n} x_{j2} & \cdots & \alpha^2\rho\sum_{j=1}^{n} x_{jp} \\ 0 & 0 & & 0 \\ 0 & & & 0 \end{pmatrix} + \alpha^2(1-\rho)I$$

$$= \begin{pmatrix} n\alpha^2\rho + \alpha^2(1-\rho) & \alpha^2\rho\Sigma x_{j2} & \cdots & \alpha^2\rho\Sigma x_{jp} \\ 0 & \alpha^2(1-\rho) & & \\ 0 & 0 & \cdots & \alpha^2(1-\rho) \end{pmatrix}$$

$\Longrightarrow \mathrm{Det}(F) = [n\alpha^2\rho + \alpha^2(1-\rho)]\alpha^{2(p-1)}(1-\rho)^{p-1} \overset{\text{want}}{\neq} 0$

$\Longrightarrow \rho \neq 1$ and $n\alpha^2\rho + \alpha^2(1-\rho) = (n-1)\alpha^2\rho + \alpha^2 \overset{\text{want}}{\neq} 0$

$\Longrightarrow$ $\rho \neq -\frac{1}{n-1}$, and $\rho \neq 1$ $\Longrightarrow$ with ①, we are guranteed that $\rho \neq -\frac{1}{n-1}$.

$\Longrightarrow$ When $-\frac{1}{n-1} < \rho < 1$, $\Sigma X = XF$ holds, and Σ p.d.

$\xrightarrow{\text{Cor 6.8.1.1}}$ $\ell'\hat{\beta}$ is the UMVUE. //

$$\Sigma = \sigma^2 \underbrace{\begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & & & \vdots \\ \rho & \cdots & \cdots & 1 \end{bmatrix}}_{A} = \sigma^2 A$$

as $\sigma^2 > 0$, Σ p.d. iff A is p.d.

We need the necessary and sufficient condition for A to be p.d. From Thm 1.4.2 (3b), we just need to find the condition that each principal minor of A has positive determinant.

Now: $a_{11} = 1 > 0$

$$\begin{vmatrix} 1 & \rho \\ \rho & 1 \end{vmatrix} = 1-\rho^2 > 0 \text{ iff } \rho^2 < 1 \text{ or } -1<\rho<1 \quad \cdots (1)$$

$$\begin{vmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{vmatrix} = (1-\rho)^2(1+2\rho) > 0 \text{ iff } 1+2\rho>0 \text{ or } \rho > -\tfrac{1}{2} \quad (\text{as } 1-\rho>0 \text{ by } (1)) \quad \cdots (2)$$

$\vdots$

$$|A| = (1-\rho)^{n-1}(1+(n-1)\rho) > 0 \text{ iff } 1+(n-1)\rho > 0 \ (\text{as } 1-\rho>0 \text{ by } (1)) \quad \cdots (n)$$

Combining (1), (2), $\cdots$, (n), we get:

$-1<\rho<1$ and $\rho > -\frac{1}{n-1}$ and hence $-\frac{1}{n-1} < \rho < 1$

That is, ~~all~~ the determinants of all minor pricipals of A are positive iff $-\frac{1}{n-1} < \rho < 1$.

Subsection 6.8.3 Student read

Sec 6.9 Examination of Assumptions

Sec 6.9.1 Residual analysis

Defn 6.9.1 For the GLM $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$. The vector $\underline{r}$ defined by $\underline{r} = \underline{Y} - \hat{\underline{Y}} = \underline{Y} - X\hat{\underline{\beta}}$, where $\hat{\underline{\beta}} = X^-\underline{Y}$, is called the vector of residuals.

Thm 6.9.1 $\underline{r} = (I - XX^-)\underline{Y} = (I - XX^-)\underline{\varepsilon}$

Proof:
$$
\begin{aligned}
\underline{r} &= \cancel{(I - XX^-)} \; \underline{Y} - X\hat{\underline{\beta}} \\
&= \underline{Y} - X(X^-\underline{Y}) \\
&= \underline{Y} - XX^-\underline{Y} \\
&= (I - XX^-)\underline{Y} \quad \checkmark \\
&= (I - XX^-)(X\underline{\beta} + \underline{\varepsilon}) \\
&= X\underline{\beta} - \underbrace{XX^-X}_{X}\underline{\beta} - XX^-\underline{\varepsilon} + \underline{\varepsilon} \\
&= X\underline{\beta} - X\underline{\beta} + \underline{\varepsilon} - XX^-\underline{\varepsilon} \\
&= (I - XX^-)\underline{\varepsilon} \quad \checkmark
\end{aligned}
$$

use Props of COMP 2020 open book

Thm 6.9.2 Under the assumption of $\underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I)$

$\underline{r} \sim N_n(\underline{0}, M\sigma^2)$, $M = I - XX^-$

and $\underline{r}$ and $\hat{\underline{\beta}}$ are independent

Proof: When $\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I)$

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \sim N(X\underline{\beta}, \sigma^2 I)$$

6-40

or $\underset{\sim}{r} = (I - XX^-)\underset{\sim}{\varepsilon}$

$$\sim N(\underset{\sim}{0}, (I - XX^-)\sigma^2 I (I - XX^-)') = N(\underset{\sim}{0}, \sigma^2 M).$$

$$\Longrightarrow \quad \underset{\sim}{r} = (I - XX^-)\underset{\sim}{Y} \sim N(\underset{\sim}{\mu}, \Sigma)$$

with
$$\underset{\sim}{\mu} = (I - XX^-)X\beta = X\beta - \underbrace{XX'X}_{X}\beta = \underset{\sim}{0}$$

$$\begin{aligned}
\Sigma &= (I - XX^-)(\sigma^2 I)(I - XX^-)' \\
&= \sigma^2 (I - XX^-)(I - X^{-\prime}X') \\
&= \sigma^2 [I - XX^- - X^{-\prime}X' + \underbrace{XX^-}_{I}X^{-\prime}X' \\
&= \sigma^2 [I - XX^- \cancel{- X(X'X)^{-1}X'} + \cancel{X^{-\prime}X'}\,] \\
&= \sigma^2 [I - X(X'X)^{-1}X'] = \sigma^2 [I - XX^-] \\
&= \sigma^2 M
\end{aligned}$$

$\underset{\sim}{r} = (I - XX^-)\underset{\sim}{Y} \qquad \hat{\beta} = \underbrace{(X'X)^{-1}X'}_{B}\underset{\sim}{Y}$

$$\underset{\sim}{r}'\underset{\sim}{r} \;\; = \;\; \underset{\sim}{Y}'\underbrace{(I - XX^-)}_{A}\underset{\sim}{Y}$$

use this proof

Consider ~~two~~ a ~~dual~~ quadratic form

$\hat{\beta}$ is linear form of $\underset{\sim}{Y}$.

$$\xrightarrow{\text{Thm 4.5.1}} \quad B\Sigma A = (X'X)^{-1}X'(\sigma^2 I)(I - XX^-)$$
$$= \sigma^2 [(X'X)^{-1}X' - \underbrace{(X'X)^{-1}X'X}_{I}(X'X)^{-1}X')$$
$$= \sigma^2 \, 0 = 0$$

$\hat{\beta}$ and $\underset{\sim}{Y}'(I - XX^-)\underset{\sim}{Y}$ are independent, or $\hat{\beta}$ and $\underset{\sim}{r}'\underset{\sim}{r}$ are independent. As $\underset{\sim}{r}'\underset{\sim}{r}$ is a function of only $\underset{\sim}{r}$ $\Rightarrow$ $\hat{\beta}$ and $\underset{\sim}{r}$ are independent. //

Alternative Proof:

$$\underset{\sim}{r} \sim N(\underset{\sim}{0}, (I - XX^-)(I - XX^-)\sigma^2 I) = N_n(\underset{\sim}{0}, (I - XX^-)\sigma^2)$$

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

~~PTO~~ $\Longrightarrow$ 6-40 #1

Consider:

$$\begin{pmatrix} \underset{\sim}{r} \\ \hat{\beta} \end{pmatrix} = \begin{bmatrix} (I - XX^-)\underset{\sim}{y} \\ X^-\underset{\sim}{y} \end{bmatrix} = \underbrace{\begin{bmatrix} I - XX^- \\ X^- \end{bmatrix}}_{B}\underset{\sim}{y} \sim N_{n+p}\left(\begin{pmatrix} I - XX^- \\ X^- \end{pmatrix}X\beta,\ \begin{bmatrix} I - XX^- \\ X^- \end{bmatrix}\sigma^2 I \begin{bmatrix} I - XX^- \\ X^- \end{bmatrix}'\right)$$

That is, we have $\underset{\sim}{\gamma} = (I - XX^-)\underset{\sim}{y} \sim N(0, \sigma^2 M)$

$= N(0, \sigma^2(I - XX^-))$ of rank $(I - XX^-)$

$= trace(I) - trace(XX^-)$

$= n - p$

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

$$\Rightarrow \begin{pmatrix} \gamma \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} (I - XX^-)\underset{\sim}{y} \\ X^- \underset{\sim}{y} \end{pmatrix} = \begin{pmatrix} I - XX^- \\ X^- \end{pmatrix} \underset{\sim}{y}$$

$$\sim N\left(\begin{bmatrix} I - XX^- \\ X^- \end{bmatrix} X\underset{\sim}{\beta}, \begin{bmatrix} I - XX^- \\ X^- \end{bmatrix} \sigma^2 \begin{bmatrix} I - XX^- \\ X^- \end{bmatrix}' \right)$$

$$= N\left(\begin{bmatrix} I - XX^- \\ X^- \end{bmatrix} X\underset{\sim}{\beta}, \sigma^2 B \right) \text{ of rank } \begin{bmatrix} I - XX^- \\ X^- \end{bmatrix} \underset{\sim}{X}$$

where $B = \begin{bmatrix} I - XX^- \\ X^- \end{bmatrix} \begin{bmatrix} I - XX^- \\ X^- \end{bmatrix}'$

$$= \begin{bmatrix} I - XX^- \\ X^- \end{bmatrix} [(I - XX^-)' \quad X^{-\prime}]$$

$$= \begin{bmatrix} (I - XX^-)(I - XX^-)' & (I - XX^-)X^{-\prime} \\ X^-(I - XX^-)' & X^- X^{-\prime} \end{bmatrix}$$

with $(I - XX^-)X^{-\prime} = X^{-\prime} - XX^- X^{-\prime}$

$= X^{-\prime} - X(X'X)^{-1}X'X(X'X)^{-1}$

$= X^{-\prime} - X(X'X)^{-1} = X^{-\prime} - X^{-\prime} = 0_{p \times n}$

and $X^-(I - XX^-)' = [(I - XX^-)X^{-\prime}]' = 0_{n \times p}$

$\Rightarrow Cov(\hat{\gamma}, \hat{\beta}) = 0$, since $\begin{pmatrix} \hat{\gamma} \\ \hat{\beta} \end{pmatrix}$ is multivariate normal

$\xrightarrow{\text{Prop(6)a.}}$ $\hat{\gamma} \perp\!\!\!\perp \hat{\beta}$

Sec 6.10 Inference about the GLM: Case 2

In case 2, we don't assume the normality on $\underline{\varepsilon}$, but just assume: $E(\underline{\varepsilon})=\underline{0}$, $Cov(\underline{\varepsilon})=\sigma^2 I$.

For this case, we can not make estimation for β and σ^2 using likelihood estimation (like we did in 6.1 – 6.9) as no distributions on $\underline{\varepsilon}$ are available.

We will, instead, using a very old (but useful) criterion in mathematics, the least squares principle to do the estimation.

Thm 6.10.1 In the GLM $\underline{Y}=X\beta+\underline{\varepsilon}$, where $E(\underline{\varepsilon})=\underline{0}$ and $cov(\underline{\varepsilon})=\sigma^2 I$, the least squares estimators of β and σ^2 are given by:

$$\Longrightarrow \hat{\underline{\beta}} = CX'\underline{Y} = (X'X)^{-1}X'\underline{Y} = X^{-}\underline{Y}, \text{ with } C=(X'X)^{-1},$$

$$\hat{\sigma}^2 = \frac{1}{n-p}(\underline{Y}-X\hat{\underline{\beta}})'(\underline{Y}-X\hat{\underline{\beta}}) = \frac{1}{n-p}\underline{Y}'(I-XX^{-})\underline{Y}.$$

Proof: The least square estimator, $\underline{b}$, is, that

$$S=\underline{r}'\underline{r} = \sum_{i=1}^{n} r_i^2 = (\underline{Y}-X\underline{b})'(\underline{Y}-X\underline{b}) \text{ is minimized}$$

$$\frac{dS}{d\underline{b}} = \cancel{-2X'(\underline{Y}-X\underline{b})} = -2X'\underline{Y} + 2X'X\underline{b} \overset{\text{set}}{=} 0$$

$$\Longrightarrow X'X\underline{b} = X'\underline{Y} \Longrightarrow \underline{b} = (X'X)^{-1}X'\underline{Y} = X^{-}\underline{Y} = \hat{\underline{\beta}}$$

As $E[(\underline{Y}-X\hat{\underline{\beta}})'(\underline{Y}-X\hat{\underline{\beta}})] = (n-p)\sigma^2$

$= E[\underline{y}'(I-XX^{-})\underline{y}] \overset{\text{Thm 4.6.1}}{=} \text{trace}[(I-XX^{-})\sigma^2 I] + (X\beta)'(I-XX^{-})X\beta = \sigma^2(n-p)+0$

$$\Longrightarrow \hat{\sigma}^2 = \frac{1}{n-p}(\underline{Y}-X\hat{\underline{\beta}})'(\underline{Y}-X\hat{\underline{\beta}}) = \frac{1}{n-p}\underline{Y}'(I-XX^{-})\underline{Y}$$

is an unbiased estimator of σ^2 (obtained through LS)

$\Longrightarrow$ The least squares estimator of $\theta_1 = \underline{\ell}'\beta$ is given by $\underline{\ell}'\hat{\beta} = \ell' X^- \underline{y}$

starts new on 11/28

~~linear~~ linear $\qquad T = (\underline{\ell}' X^- + \underline{b}')\underline{y} + a_0$

(2). Consider any estimator of the form $\underline{a}'\underline{Y} + a_0$ where $\underline{b}$ and a_0 are to be determined so that T is an unbiased estimator of $\underline{\ell}'\beta$ and has min. variance among all unbiased estimator of $\underline{\ell}'\beta$

~~Let $\underline{a}' = \underline{\ell}' X^- + \underline{b}'$, then we determin $\underline{b}$ and a_0 instead~~ known

Now. $\cancel{E[\underline{a}'\underline{Y} + a_0]} = \overbrace{E[(\underline{\ell}'X^- + \underline{b}')\underline{Y} + a_0]}^{E[T]}$

$= E[\underline{\ell}'X^-\underline{Y} + \underline{b}'\underline{Y} + a_0]$

$= E[\underline{\ell}'X^-\underline{Y}] + \underline{b}'E(\underline{Y}) + a_0$

$= \underline{\ell}'X^- E(\underline{Y}) + \underline{b}'E(\underline{Y}) + a_0$

$= \underline{\ell}'\underbrace{X^- X}_{I}\beta + \underline{b}'X\beta + a_0$

$= \underline{\ell}'\beta + \underline{b}'X\beta + a_0$

Note: $\{E(\underline{\varepsilon}) = 0,\ Cov(\underline{\varepsilon}) = \sigma^2 I\}$
$\Longrightarrow E(\underline{Y}) = X\beta$
$Var(\underline{Y}) = \sigma^2 I$

$\overset{\text{Want}}{\underset{\text{unbiasedness}}{=}}\ \underline{\ell}'\beta$

$\Longrightarrow \quad \underline{b}'X\beta + a_0 = 0 \quad$ for all β in E_p

$\Longrightarrow \quad \underline{b}'X = \underline{0}$ and $a_0 = 0$.

Also $X'\underline{b} = \underline{0}$

$\cancel{Var[\underline{a}'\underline{Y} + a_0]} = \cancel{Var[\underline{a}'\underline{Y}]} = \overset{Var[T]}{Var}[(\underline{\ell}'X^- + \underline{b}')\underline{Y}]$

$= (\underline{\ell}'X^- + \underline{b}')Cov(\underline{Y})(\underline{\ell}'X^- + \underline{b}')'$

$= (\underline{\ell}'X^- + \underline{b}')\sigma^2 I(X^{-\prime}\underline{\ell} + \underline{b})$

$= \sigma^2\underline{\ell}'X^-X^{-\prime}\underline{\ell} + \sigma^2\underline{b}'X^{-\prime}\underline{\ell} + \sigma^2\underline{\ell}'X^-\underline{b} + \sigma^2\underline{b}'\underline{b}$

$= \sigma^2\underline{\ell}'(X'X)^{-1}X'X(X'X)^{-1}\underline{\ell} + \sigma^2\underbrace{\underline{b}'X}_{=0}(X'X)^{-1}\underline{\ell}$

$+ \sigma^2\underline{\ell}'(X'X)^{-1}\underbrace{X'\underline{b}}_{=0} + \sigma^2\underline{b}'\underline{b}$

$= \sigma^2\underline{\ell}'(X'X)^{-1}\underline{\ell} + \sigma^2\underline{b}'\underline{b}$ ~~$\sigma^2 b'b$~~

constant & p.d. (cor 1.4.2(3))

> 0 if $\underline{\ell} \neq \underline{0}$

$\Rightarrow$ The minimum variance $Var(T)$ is attained if

$\sigma^2 \underline{b}'\underline{b} \geq 0$ is minimal $\Longrightarrow$ $\underline{b}'\underline{b} = 0 \Rightarrow \underline{b} = 0$ (also consistent with $\underline{b}'X = \underline{0}$ requirement for unbiased-ness of T)

Therefore, the minimum variance of T, $Var(T)$, subject to $E[T] = \underline{\ell}'\beta$, is $\sigma^2 \underline{\ell}'(X'X)^{-1}\underline{\ell}$ and is attained by $Var(T)$ when: $\underline{b} = \underline{0}$ and $a_0 = 0$.

$\Longrightarrow$ The BLU estimator of $\underline{\ell}'\beta$ is $T = \underline{\ell}'X^{-}\underline{y} = \underline{\ell}'\hat{\beta}$ //.

Remarks: What about the LS estimator $\hat{\sigma}^2$ of σ^2?

It ($\hat{\sigma}^2$) is called the best quadratic unbiased estimator of σ^2.

(See Defn 6.10.2 & Thm 6.10.3)