Chaper 7 Computing Techniques

Sec 7.1 Intro

In GLM, the first important step is to find the point estimator of β and σ^2, then derive a test statistic or confidence interval. The basic information contained in the data matrix X and $\underline{y}$ vector are very important for our estimation process, whether we use maximum likelihood estimation or least squares estimation.

Some methods already exist for the computing required in GLMs, which method to use depends on several factors:

(1) the amount of data (the values of n and p);
(2) how "ill-conditioned" the matrix X is;
(3) what statistics are needed?
(4) the type of computers available (less of a concern nowadays as computer technolodge ~~has~~ advanced a lot lately).

The problem of computing for GLM is not a statistical problem, but a numerical analysis problem in applied math. Statisticians still need to know important aspects of computing in GLM.

In normal equations or OL estimation process, we need to solve $\hat{\beta}$ from:

$$\underbrace{X'X}_{S,\ p.d.}\hat{\underline{\beta}} = X'\underline{y} \qquad (7.1.1)$$

$$\Longrightarrow \quad S\hat{\underline{\beta}} = X'\underline{y} = \underline{s}$$

Instead of using S^{-1} (may be ill-conditioned) to solve for $\hat{\underline{\beta}}$, its very popular to use upper- and lower-triangular matrix decomposition of S. (factorization)

That is: $S = AB$

A: lower triangular of full rank

B: upper triangular of full rank

So
$$AB\hat{\underline{\beta}} = \underline{s}$$
$$\Longrightarrow \quad B\hat{\underline{\beta}} = A^{-1}\underline{s}$$
$$\Longrightarrow \quad \hat{\underline{\beta}} = B^{-1}A^{-1}\underline{s}$$

Three Popular factorization methods for S:

1. Gaussian elimination
2. Doolittle;
3. square root (Cholesky)

If not computing the normal equations, ~~to get~~ $\hat{\underline{\beta}}$ is obtained through X^{-}, and two computing procedures are popular:

1. Gram-Schmidt orthogonalization,
2. orthogonal Householder transformation.

We will talk about the square root method for factoring S.

Sec 7.2 Square Root Method of Factoring a Positive Definite Matrix.

Thm 7.2.1 Let $S = X'X$ be a $p \times p$ positive definite matrix. There exists an upper triangular matrix T of rank p s.t.

$$S = T'T \qquad (7.2.1)$$

and such that $t_{ii} > 0$ for $i = 1, \cdots, p$. Also, T is unique.

student read skip

Proof: By mathematical induction.

First, let $p = 1$; then $S = [S_{ii}^2]$, clearly $t_{11} = |S_{11}| > 0$ and is unique. Thm is true for $p = 1$.

Next, assume that the theorem is true for $p = k$. That is, for any $k \times k$ p.d. S_{11}, there exists a unique upper triangular (real) matrix T_{11}, with $t_{ii} > 0$ for $i = 1, \cdots, k$ s.t.:

$$S_{11} = T_{11}' T_{11}$$

Now let S be any known $(k+1) \times (k+1)$ p.d. matrix. Since S is p.d., we can write S as.

$$S = \begin{bmatrix} S_{11} & \underset{\sim}{S}_{12} \\ \underset{\sim}{S}_{21} & S_{22} \end{bmatrix}$$

$\uparrow$ scaler

where S_{11} is $k \times k$ p.d. and $S_{22} > 0$, $\underset{\sim}{S}_{12}' = \underset{\sim}{S}_{21}$.

Using the induction hypothesis, we can writ

$$S_{11} = T_{11}' T_{11},$$ where T_{11} is a unique $k \times k$ upper triangular matrix with $t_{ii} > 0$ for $i = 1, \cdots, k$.

Therefore, S becomes:

$$S = \begin{bmatrix} T_{11}' T_{11} & \underline{S}_{12} \\ \underline{S}_{21} & S_{22} \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} T_{11}' & \underline{0} \\ \underset{\underline{S}_{21}}{\underset{\|}{\underline{S}_{12}'}} T_{11}^{-1} & b \end{bmatrix}}_{T'} \underbrace{\begin{bmatrix} T_{11} & T_{11}'^{-1}\underline{S}_{12} \\ \underline{0} & b \end{bmatrix}}_{T} = T'T$$

where $S_{22} = \underline{S}_{21} T_{11}^{-1} T_{11}'^{-1} \underline{S}_{12} + b^2 = \underline{S}_{21} (T_{11}' T_{11})^{-1} S_{12} + b^2$

$= \underline{S}_{21} (S_{11}^{-1}) \underline{S}_{12} + b^2 > 0$

$\cancel{\Rightarrow b = (S_{22} - \underline{S}_{21} S_{11}^{-1} \underline{S}_{12})^{1/2}}$

As $\quad \underset{0}{\underset{\vee}{|S|}} = \underset{0}{\underset{\vee}{|S_{11}|}} \; |S_{22} - \underset{\uparrow \text{scalar}}{\underline{S}_{21}} S_{11}^{-1} \underline{S}_{12}|$ $\qquad |\cdot| = \text{determinant}$

$\Rightarrow \; |S_{22} - \underline{S}_{21} S_{11}^{-1} S_{12}| > 0$

$\Rightarrow \; S_{22} - \underline{S}_{21} S_{11}^{-1} S_{12} > 0$ as the determinant of a scalar is it self.

$\Rightarrow \; b = (S_{22} - \underline{S}_{21} S_{11}^{-1} \underline{S}_{12})^{\frac{1}{2}}$ is unique

$\Rightarrow$ T is unique.

$\Rightarrow$ When $p = k+1$, the theorem also holds. //

In real computation, T is obtained by the Cholesky decomposition or factorization.

See (7.2.3), the procedure of square root method

Using this factorization of S,

Solving: $S\hat{\underline{\beta}} = \underline{s}$, $\underline{s} = X'\underline{y}$

becomes solve $\hat{\beta}$ from: $T'T\hat{\underline{\beta}} = \underline{s}$

$\Rightarrow T\hat{\underline{\beta}} = T'^{-1}\underline{s} = \underline{t}$

$\Rightarrow$ easy to solve without any T^{-1}

Sec 7.3 Computing Point Estimates, Test Statistics and Confidence Intervals

For optimal computing purpose, we establish some computational ~~strategy~~ strategies for obtaining the point estimates, $\hat{\beta}$, $\ell'\hat{\beta}$, $\hat{\sigma}^2 = \frac{1}{n-p}[\underline{y}'\underline{y} - \hat{\beta}X'\underline{y}]$, $\ell_i'\hat{\beta}$, $\ell_i'(X'X)^{-1}\ell_i$, $i = 1, \cdots, q$, $H\hat{\beta} - \underline{h}$, $[H(X'X)^{-1}H']^{-1}$, etc.

normal equations

proof: $X'X = T'T$
$T'T\hat{\beta} = X'Y = \underline{s}$
$T\hat{\beta} = T'^{-1}\underline{s} = \underline{t}$

1. $\hat{\beta}$ is obtained using $T\hat{\beta} = \underline{t}$ $\Rightarrow$ $\hat{\beta} = T^{-1}\underline{t}$, $T'^{-1}\underline{s} = \underline{t}$
as in sec 7.2 $\quad \underline{s} = T'\underline{t}$

2. $\hat{\sigma}^2 = \frac{1}{n-p}[\underline{y}'\underline{y} - \underline{t}'\underline{t}]$, $\underline{t} = T'^{-1}\underline{s}$.
[This way, it means $\hat{\sigma}^2$ can be computed without $\hat{\beta}$].

Proof: From the normal equations $[S \mid \underline{s}]$, we reduce S to the triangular matrix T:

move to next page

$$
\begin{aligned}
[T \mid \underline{t}] &= T'^{-1}[T'T \mid T'\underline{t}] \\
&= T'^{-1}[S \mid \underline{s}] \\
&= T'^{-1}[X'X \mid X'\underline{y}] \qquad \cdots \quad (1)
\end{aligned}
$$

move to next page

Note that:

$$
\begin{aligned}
\underline{t}'\underline{t} &= (T'^{-1}\underline{s})'(T'^{-1}\underline{s}) = \underline{s}'T^{-1}T'^{-1}\underline{s} \\
&= (X'\underline{y})'(T'T)^{-1}(X'\underline{y}) \\
&= \underline{y}'X(S)X'\underline{y} = \underline{y}'X\underbrace{(X'X)^{-1}}X'\underline{y} \\
&= \underline{y}'X\underbrace{X^{-}\underline{y}}_{\hat{\beta}} = \underbrace{\underline{y}'X\hat{\beta}}_{\text{scalar}} = \hat{\beta}'X'\underline{y}
\end{aligned}
$$

$$
\begin{aligned}
\Longrightarrow \hat{\sigma}^2 &= \frac{1}{n-p}[\underline{y}'\underline{y} - \hat{\beta}X'\underline{y}] \\
&= \frac{1}{n-p}[\underline{y}'\underline{y} - \underline{t}'\underline{t}]
\end{aligned}
$$

3. $\underline{l}_i'\hat{\underline{\beta}}$ and $\underline{l}_i'(X'X)^{-1}\underline{l}_i$, $i=1,\cdots,q$ can be computed by:

$$\underline{l}_i'\hat{\underline{\beta}} = \underline{a}_i'\underline{t}, \quad \underline{a}_i = T'^{-1}\underline{l}_i,$$

$$\underline{l}_i'(X'X)^{-1}\underline{l}_i = \underline{a}_i'\underline{a}_i.$$

$S\hat{\underline{\beta}} = \underline{s}$

Proof: From the normal equations $[S \mid \underline{s}]$, we reduce S to the triangular matrix T as:

$$[T \mid \underline{t}] = T'^{-1}[T'T \mid T'\underline{t}]$$
$$= T'^{-1}[S \mid \underline{s}]$$
$$= T'[X'X \mid X'\underline{y}]$$

$\underline{s} = T'\underline{t} = X'\underline{y}$

$\underline{t} = T'^{-1}\underline{s}$

In practice how to carry out the operation

Apply the above process to $[S \mid \underline{s} \mid \underline{l}_1\ \underline{l}_2 \cdots \underline{l}_q]$

$$\Longrightarrow \quad T'^{-1}[S \mid \underline{s} \mid \underline{l}_1\ \underline{l}_2 \cdots \underline{l}_q]$$
$$= T'^{-1}[T'T \mid T'\underline{t} \mid \underline{l}_1\ \underline{l}_2 \cdots \underline{l}_q]$$
$$= [T \mid \underline{t} \mid T'^{-1}\underline{l}_1\ \ T'^{-1}\underline{l}_2 \cdots T'^{-1}\underline{l}_q]$$
$$= [T \mid \underline{t} \mid \underline{a}_1\ \underline{a}_2 \cdots \underline{a}_q]$$

with $\underline{a}_i = T'^{-1}\underline{l}_i$, $i=1,\cdots,q$, $\underline{t} = T'^{-1}\underline{s} = T'^{-1}X'\underline{y}$

$$\Longrightarrow \quad \underline{l}_i'\hat{\underline{\beta}} = \underline{l}_i'X^{-}\underline{y} = \underline{l}_i'(X'X)^{-1}X'\underline{y} = \underline{l}_i'(S)^{-1}X'\underline{y}$$
$$= \underline{l}_i'(T'T)^{-1}X'\underline{y} = \underbrace{\underline{l}_i'T^{-1}}_{\underline{a}_i'}\underbrace{T'^{-1}X'\underline{y}}_{\underline{t}} = \underline{a}_i'\underline{t}$$

$$\Longrightarrow \quad \underline{l}_i'(X'X)^{-1}\underline{l}_i = \underline{l}_i'(S)^{-1}\underline{l}_i = \underline{l}_i'(T'T)^{-1}\underline{l}_i$$
$$= \underbrace{\underline{l}_i'T^{-1}}_{\underline{a}_i'}\ \underbrace{T'^{-1}\underline{l}_i}_{\underline{a}_i} = \underline{a}_i'\underline{a}_i$$

So, finding T and T^{-1} is the key in the computational approaches presented here.

where $G' = T'^{-1}H'$ or $G = HT^{-1}$

Proof starts here

Note: $GG' = HT^{-1}T'^{-1}H' = H(T'T)^{-1}H'$

$= H(S)^{-1}H' = H(X'X)^{-1}H'$ $q\times q$ p.d.

$\Longrightarrow \exists\, T_0$, $q\times q$ nonsingular upper triangular matrix

s.t. $GG' = H(X'X)^{-1}H' = T_0'T_0$.

Let $\underline{g} = G\underline{t} - \underline{h}$, then:

$$H\hat{\underline{\beta}} = H(X'X)^{-1}X'\underline{y} = H(T'T)^{-1}X'\underline{y}$$

$$= HT^{-1}\underbrace{T'^{-1}X'\underline{y}}_{\underline{t}} = HT^{-1}\underline{t}$$

$$= G\underline{t} = \underline{g} + \underline{h}$$ (i) proved.

also, $\underline{g} = H\hat{\underline{\beta}} - \underline{h}$

For $\underline{t}_0 = T_0'^{-1}\underline{g}$, we have:

$$\underline{t}_0'\underline{t}_0 = \underline{g}'T_0^{-1}T_0'^{-1}\underline{g} = \underline{g}'(T_0'T_0)^{-1}\underline{g}$$

$$= \underline{g}'[H(X'X)^{-1}H']^{-1}\underline{g}$$

$$= \underline{g}'(GG')^{-1}\underline{g} = \underline{g}'G'^{-1}G^{-1}\underline{g}$$

$$= (H\hat{\underline{\beta}} - \underline{h})'(GG')^{-1}(H\hat{\underline{\beta}} - \underline{h})$$

$$= (H\hat{\underline{\beta}} - \underline{h})'[H(X'X)^{-1}H']^{-1}(H\hat{\underline{\beta}} - \underline{h}).$$ (2) proved

Note that:

$$W = \frac{(H\hat{\underline{\beta}} - \underline{h})'[H(X'X)^{-1}H']^{-1}(H\hat{\underline{\beta}} - \underline{h})}{\underline{y}'\underline{y} - \hat{\underline{\beta}}X'\underline{y}} \cdot \left(\frac{n-p}{q}\right)$$

$$= \frac{\underline{t}_0'\underline{t}_0}{\underline{Y}'\underline{Y} - \hat{\underline{\beta}}X'\underline{y}}\left(\frac{n-p}{q}\right)$$

$$\overset{d.}{=} \frac{\underline{t}_0'\underline{t}}{\underline{Y}'\underline{Y} - \underline{t}'\underline{t}} \quad \frac{n-p}{q}$$

(3) is proved.

Computing the inverse of $X'X \overset{\text{reduces}}{\Longrightarrow}$ computing T and T'^{-1}.

eg 7.3.1 We have

GLM: $Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i$,

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, $i = 1, \cdots, 36$.

Find point estimates of $\beta_1, \beta_2, \beta_3$ and β_4, find point estimate of σ^2 and 95% individual CI for $2\beta_1 + \beta_2 + 3\beta_3 - \beta_4$ and for $2\beta_1 + 2\beta_2 + \beta_3$.

With $\underline{y}'\underline{y} = 581$, $X'X = \begin{bmatrix} 4 & 2 & 2 & 2 \\ 2 & 2 & 2 & 3 \\ 2 & 2 & 3 & 4 \\ 2 & 3 & 4 & 10 \end{bmatrix}$.

Sol.

See book.

4. ① $H\hat{\beta} = G\underline{t} = \underline{g} + \underline{h}$ where $G' = T'^{-1}H'$, $\underline{t} = T'^{-1}X'\underline{y}$ (as before) $G = HT^{-1}$ $\underline{g} = H\hat{\beta} - \underline{h}$

② $(H\hat{\beta} - \underline{h})'[H(X'X)^{-1}H']^{-1}(H\hat{\beta} - \underline{h}) = \underline{g}'(GG')^{-1}\underline{g}$

$= \underline{g}' T_0^{-1} T_0'^{-1} \underline{g} = \underline{t}_0' \underline{t}_0$ with $\underline{t}_0 = T_0'^{-1}\underline{g}$, T_0 is the Cholesky factorization of $H(X'X)^{-1}H'$ ($q \times q$ pd), and

$\underline{g} = H\hat{\beta} - \underline{h}$

③ $W = \dfrac{\underline{t}_0'\underline{t}_0}{\underline{y}'\underline{y} - \underline{t}'\underline{t}} \left(\dfrac{n-p}{q}\right)$.

Proof: Consider the augmented matrix: $[\overset{S}{\overset{\shortparallel}{X'X}} \mid X'\underline{y} \mid H']$:

$[X'X \mid X'\underline{y} \mid H'] = [S \mid X'\underline{y} \mid H'] = [T'T \mid X'\underline{y} \mid H']$.

$\Longrightarrow T'^{-1}[X'X \mid X'\underline{y} \mid H'] = T'^{-1}[T'T \mid X'\underline{y} \mid H']$

$= [T \mid \underbrace{T'^{-1}X'\underline{y}}_{\underline{t}} \mid \underbrace{T'^{-1}H'}_{\text{let } = G'}]$

$= [T \mid \underline{t} \mid G']$

Sec 7.4 Analysis of Variance

For GLMs, a convinient $\overset{\text{computational}}{\text{procedure}}$ is called the "<u>analysis of variance</u>" or ANOVA. The idea of ANOVA comes from partitioning the total sum of squares of the observations, $\underline{y}'\underline{y}$, into a sum of k quadratic forms (hope all are independent and are corresponding to the specified sub-models).

If the data matrix X is partioned into

$$\underset{n\times p}{X} = [\underset{n\times(p-q)}{X_1} \quad \underset{n\times q}{X_2}]$$

[illegible] to same time

Then, we can write $\underline{y}'\underline{y}$ as:

$$\underline{y}'\underline{y} = \underline{y}'\underline{y} + \underline{y}'(X_1X_1^-)\underline{y} - \underline{y}'(X_1X_1^-)\underline{y} + \underline{y}'XX^-\underline{y} - \underline{y}'XX^-\underline{y}$$

$$= \underline{y}'(X_1X_1^-)\underline{y} + \underline{y}'(XX^- - X_1X_1^-)\underline{y} + \underline{y}'(I - XX^-)\underline{y}$$

quadratic forms

For example, for GLM: $\underline{Y} = X\beta + \underline{\varepsilon}$, $\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I)$

to test $H_0: H\beta = \underset{\underline{h}}{\underline{0}}$ vs $H_1: H\beta \neq \underline{0}$ (Case 1 in chapter 6 with $\underline{h} = \underline{0}$)

$\underline{\gamma} = G\beta$

G is $(p-q)\times p$ of rank $p-q$

We have had: full model $\underline{Y} = X\beta + \underline{\varepsilon}$, $\Rightarrow X'X\hat{\beta} = X'\underline{y}$

reduced model $\underline{Y} = B\underline{\gamma} + \underline{\varepsilon}$, $\Rightarrow B'B\hat{\underline{\gamma}} = B'\underline{Y}$

So, $B'B\hat{\underline{\gamma}} = B'\underline{Y} = B'\underline{z}$, $\hat{\gamma} = (B'B)^- B'\underline{Y} = B^-\underline{Y}$, $B = XG^-$, $\underline{z} = \underline{y} - XH^-\underset{\underset{0}{\|}}{\underline{h}} = \underline{y}$

In this case, we can write

$$\underline{y}'\underline{y} = \underbrace{\underline{y}'BB^-\underline{y}}_{\hat{\gamma}'B'\underline{y}} + \underbrace{\underline{y}'(XX^- - BB^-)\underline{y}}_{= \hat{\beta}'X'\underline{y} - \hat{\gamma}'B'\underline{y}} + \underbrace{\underline{y}'(I - XX^-)\underline{y}}_{= \underline{Y}'\underline{Y} - \hat{\beta}'X'\underline{Y}}$$

$$\Rightarrow \cancel{\underline{y}'\underline{y} - \underline{y}'BB^-\underline{y} = \underline{y}'(XX^- - BB^-)\underline{y} + \underline{y}'(I - XX^-)\underline{y}}$$

$$\Longrightarrow \underline{y}'\underline{y} = \hat{\underline{\gamma}}'B'\underline{y} + \left(\hat{\underline{\beta}}'X'\underline{y} - \hat{\underline{\gamma}}'B'\underline{y}\right) + \left(\underline{y}'\underline{y} - \hat{\underline{\beta}}'X'\underline{y}\right) \ldots$$

Interesting interpretation of above:

If $\underline{\beta}$ is not in the model, the model is just $\underline{y} = \underline{\varepsilon}$, then $\frac{1}{\sigma^2}\underline{y}'\underline{y}$ is a chi-square random variable, and $\frac{1}{n}\underline{y}'\underline{y}$ is an unbiased estimator for σ^2.

If $\underline{\beta}$ is in the model, that is, $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$, then $\frac{1}{n-p}(\underline{y}'\underline{y} - \hat{\underline{\beta}}'X'\underline{y})$ is an unbiased estimator for σ^2

$\Longrightarrow$ The reduction in the sum of squares for estimating σ^2 due to $\underline{\beta}$ in the model is $\hat{\underline{\beta}}X'\underline{y}$, in short, we say: "reduction due to $\underline{\beta}$" is $\hat{\underline{\beta}}X'\underline{y} \triangleq R(\underline{\beta})$

If $\underline{\gamma}$ is in the model, that is, $\underline{Y} = B\underline{\gamma} + \underline{\varepsilon}$, then the "reduction due to $\underline{\gamma}$" is $\hat{\underline{\gamma}}B'\underline{Y} \triangleq R(\underline{\gamma})$.
$= \hat{\underline{\gamma}}B'\underline{z}$

If $H_0: H\underline{\beta} = \underline{0}$ is true, the variance in the reduced parameter space is estimated by $\hat{\sigma}^2_\omega = \underline{z}'(I - BB^-)\underline{z}$

$$\hat{\sigma}^2_\omega = \hat{\sigma}^2_\Omega + \frac{1}{n}(H\hat{\underline{\beta}})'[H(X'X)^{-1}H']^{-1}H\hat{\underline{\beta}}$$

$\Longrightarrow$ reduction due to H_0 is: due to the difference in the $R(\underline{\beta})$ and $R(\underline{\gamma})$ or say:

reduction due to H_0 is $R(H_0) = R(\underline{\beta}) - R(\underline{\gamma}) = \hat{\underline{\beta}}X'\underline{y} - \hat{\underline{\gamma}}B'\underline{Y}$

① $\Longrightarrow$ becomes

$$\underbrace{\underline{y}'\underline{y}}_{SST} = \underbrace{\overset{R(\underline{\gamma})}{\cancel{\hat{\gamma}B'\underline{y}}}}_{SSR} + \underbrace{R(\underline{\beta}) - R(\underline{\gamma})}_{SSH_0} + \underbrace{\underline{y}'\underline{y} - R(\underline{\beta})}_{SSE}$$

$$\Longrightarrow \quad SST = \underbrace{SSR}_{R(\underline{\gamma})} + \underbrace{SSH_0}_{R(\underline{\beta}|\underline{\gamma})} + SSE$$

where we define $R(\beta|\gamma) = R(\beta) - R(\gamma) = R(H_0)$

$\Longrightarrow$ reduction due to β after reduction in γ has been accounted for.

Then, we have the following ANOVA table for GLM: $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ under $H_0: H\underline{\beta} = \underline{0}$:

Source of Variation	Sum of Squares	d.f.	MSE
total	$\underline{y}'\underline{y} = SST$	n	
* Reduction due to β	$\hat{\underline{\beta}}'X'\underline{y} = R(\underline{\beta})$	p	
Reduction due to γ	$\hat{\underline{\gamma}}'B'\underline{y} = \underline{\gamma}'B'\underline{z} = R(\underline{\gamma})$	$p-q$	
Reduction due to H_0	$R(\underline{\beta}) - R(\underline{\gamma}) = R(\underline{\beta}\|\underline{\gamma})$	q	$R(\underline{\beta}\|\underline{\gamma})/q$
Error variance	$\underline{y}'\underline{y} - \underbrace{\hat{\underline{\beta}}'X'\underline{y}}_{R(\underline{\beta})} = SSE$	$n-p$	$SSE/(n-p)$

Note: $SSE + R(\underline{\gamma}) + R(\underline{\beta}|\underline{\gamma}) = SST$.

as a side note.

$$W = \frac{R(\underline{\beta}|\underline{\gamma})/q}{SSE/(n-p)}.$$

In general, we have the following two theorems.

HW

Thm 7.4.1 For GLM $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$, if we write it as $\underline{Y} = X_1\underline{\beta_1} + X_2\underline{\beta_2} + X_3\underline{\beta_3} + \underline{\varepsilon}$, then

$$R(\underline{\beta_2}, \underline{\beta_3} | \underline{\beta_1}) - R(\underline{\beta_2}|\underline{\beta_1}) = R(\underline{\beta_3}|\underline{\beta_1}, \underline{\beta_2}).$$

or $R(\underline{\beta_2}, \underline{\beta_3}|\underline{\beta_1}) = R(\underline{\beta_2}|\underline{\beta_1}) + R(\underline{\beta_3}|\underline{\beta_1}, \underline{\beta_2})$

Proof: Note that $R(\underline{\beta_2}, \underline{\beta_3}|\underline{\beta_1})$ is for testing the full model $\underline{Y} = X_1\underline{\beta_1} + X_2\underline{\beta_2} + X_3\underline{\beta_3} + \underline{\varepsilon}$ $\cdots$ (1) against the reduced model: $\underline{Y} = X_1\underline{\beta_1} + \underline{\varepsilon}$

That is, to test $H_0: \underline{\beta}_2 = \underline{0}$ and $\underline{\beta}_3 = \underline{0}$

$$\Longrightarrow \quad R(\underline{\beta}_2, \underline{\beta}_3 | \underline{\beta}_1) = \underbrace{R(\underline{\beta}_1, \underline{\beta}_2, \underline{\beta}_3)}_{\text{also as } R(\underline{\beta})} - R(\underline{\beta}_1)$$

Similarly, $R(\underline{\beta}_3 | \underline{\beta}_1, \underline{\beta}_2)$ is for testing the full model (i) against the reduced model

$$\underline{Y} = X_1 \underline{\beta}_1 + X_2 \underline{\beta}_2 + \underline{\varepsilon}$$

That is, to test $H_0: \underline{\beta}_3 = \underline{0}$

$$\Longrightarrow \quad R(\underline{\beta}_3 | \underline{\beta}_1, \underline{\beta}_2) = R(\underline{\beta}_1, \underline{\beta}_2, \underline{\beta}_3) - R(\underline{\beta}_1, \underline{\beta}_2$$

Again similarly, $R(\underline{\beta}_2 | \underline{\beta}_1) = R(\underline{\beta}_1, \underline{\beta}_2) - R(\underline{\beta}_1)$

$$\Longrightarrow \quad R(\underline{\beta}_2, \underline{\beta}_3 | \underline{\beta}_1) - R(\underline{\beta}_2 | \underline{\beta}_1) = R(\underline{\beta}_3 | \underline{\beta}_1, \underline{\beta}_2) \quad //$$

HW

Thm 7.4.2 Consider GLM $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ which is written as $\underline{Y} = X_1\underline{\beta}_1 + X_2\underline{\beta}_2 + \underline{\varepsilon}$. Then:

$$R(\underline{\beta}_1 | \underline{\beta}_2) = R(\underline{\beta}_1) \text{ iff } X_1' X_2 = \underline{0}.$$

Proof: Note that

$$\begin{aligned} R(\underline{\beta}_1 | \underline{\beta}_2) &= R(\underline{\beta}_1; \underline{\beta}_2) - R(\underline{\beta}_2) \\ &= R(\underline{\beta}) - R(\underline{\beta}_2) \\ &= \hat{\underline{\beta}}' X' \underline{y} - \hat{\underline{\beta}}_2' X_2' \underline{y} \quad \cdots \text{(I)} \end{aligned}$$

where $\hat{\underline{\beta}} = (X'X)^{-1} X' \underline{y}$ and $\hat{\underline{\beta}}_2 = (X_2' X_2)^{-1} X_2' \underline{y}$.

$$\begin{aligned} R(\underline{\beta}) &= \underline{y}' X (X'X)^{-1} X' \underline{y} \\ &= \underline{y}' [X_1 \ \ X_2] \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} \underline{y} \end{aligned}$$

When $X_1'X_2 = 0 \Rightarrow X_2'X_1 = 0$

$$\Rightarrow R(\underline{\beta}) = \underline{y}'[X_1 \quad X_2]\begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{bmatrix}^{-1}\begin{bmatrix} X_1' \\ X_2' \end{bmatrix}\underline{y}$$

$$= \underline{y}'[X_1 \quad X_2]\begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{bmatrix}\begin{bmatrix} X_1' \\ X_2' \end{bmatrix}\underline{y}$$

$$= \underline{y}'[X_1 \quad X_2]\begin{bmatrix} (X_1'X_1)^{-1}X_1' \\ (X_2'X_2)^{-1}X_2' \end{bmatrix}\underline{y}$$

$$= \underline{y}'\left(X_1(X_1'X_1)^{-1}X_1' + X_2(X_2'X_2)^{-1}X_2'\right)\underline{y}$$

$$= \hat{\underline{\beta}}_1'X_1'\underline{y} + \hat{\underline{\beta}}_2'X_2'\underline{y} = R(\underline{\beta}_1) + R(\underline{\beta}_2) \quad (*)$$

plug ⓧ in ①

$$\Rightarrow R(\underline{\beta}_1|\underline{\beta}_2) = \hat{\underline{\beta}}_1'X_1'\underline{y} = R(\underline{\beta}_1).$$

When $R(\underline{\beta}_1|\underline{\beta}_2) = R(\underline{\beta}_1)$

$$\Rightarrow R(\underline{\beta}) = R(\underline{\beta}_1|\underline{\beta}_2) + R(\underline{\beta}_2) = R(\underline{\beta}_1) + R(\underline{\beta}_2)$$

(*) can be reversed all the way up

$\Rightarrow X_1'X_2 = 0$

//

Sec 7.5 The Normal Equations Using Deviations from Means

When $\underset{\sim}{\beta}$ in the GLM has a constant term β_0, the GLM can be written as:

$$Y = [\underset{\sim}{1} \;\; X_2] \begin{bmatrix} \beta_0 \\ \underset{\sim}{\beta}_2 \end{bmatrix} + \underset{\sim}{\varepsilon} = \beta_0 \underset{\sim}{1} + X_2 \underset{\sim}{\beta}_2 + \underset{\sim}{\varepsilon}$$

where $\underset{\sim}{\beta}_2 = [\beta_1, \beta_2, \cdots, \beta_{p-1}]$ and X_2 is an $n \times (p-1)$ ma

In this case, the normal equations $X'X\hat{\underset{\sim}{\beta}} = X'\underset{\sim}{y}$ becomes.

$$\begin{bmatrix} \underset{\sim}{1}' \\ X_2' \end{bmatrix} [\underset{\sim}{1} \;\; X_2] \begin{bmatrix} \hat{\beta}_0 \\ \hat{\underset{\sim}{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \underset{\sim}{1}' \\ X_2' \end{bmatrix} \underset{\sim}{y}$$

$$\Longrightarrow \begin{bmatrix} n & \underset{\sim}{1}'X_2 \\ X_2'\underset{\sim}{1} & X_2'X_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\underset{\sim}{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \underset{\sim}{1}'\underset{\sim}{y} \\ X_2'\underset{\sim}{y} \end{bmatrix} \qquad (7.5.1)$$

$X_2 = [\underbrace{\underset{\sim}{x}_1 \;\; \underset{\sim}{x}_2 \cdots \underset{\sim}{x}_{p-1}}_{n\times 1}]$

$\underset{\sim}{1}'X_2 = [n\bar{x}_1 \;\; n\bar{x}_2 \;\; n\bar{x}_{p-1}] = n\bar{\underset{\sim}{x}}'$

$$\Longrightarrow X'X \longrightarrow \begin{bmatrix} n & n\bar{\underset{\sim}{x}}' \\ n\bar{\underset{\sim}{x}} & X_2'X_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\underset{\sim}{\beta}}_2 \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ X_2'\underset{\sim}{y} \end{bmatrix}$$

multiplying $\begin{bmatrix} 1 & \underset{\sim}{0} \\ -\bar{\underset{\sim}{x}} & I \end{bmatrix}$

$$\Longrightarrow \begin{bmatrix} 1 & \underset{\sim}{0} \\ -\bar{\underset{\sim}{x}} & I \end{bmatrix} \begin{bmatrix} n & n\bar{\underset{\sim}{x}}' \\ n\bar{\underset{\sim}{x}} & X_2'X_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\underset{\sim}{\beta}}_2 \end{bmatrix} = \begin{bmatrix} 1 & \underset{\sim}{0} \\ -\bar{\underset{\sim}{x}} & I \end{bmatrix} \begin{bmatrix} n\bar{y} \\ X_2'\underset{\sim}{y} \end{bmatrix}$$

$$\Longrightarrow \begin{bmatrix} n & n\bar{\underset{\sim}{x}}' \\ \underset{\sim}{0} & X_2'X_2 - n\bar{\underset{\sim}{x}}\bar{\underset{\sim}{x}}' \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\underset{\sim}{\beta}}_2 \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ X_2'\underset{\sim}{y} - n\bar{y}\bar{\underset{\sim}{x}} \end{bmatrix}$$

$$\Longrightarrow \begin{bmatrix} n\hat{\beta}_0 + n\bar{\underset{\sim}{x}}'\hat{\underset{\sim}{\beta}}_2 \\ (X_2'X_2 - n\bar{\underset{\sim}{x}}\bar{\underset{\sim}{x}}')\hat{\underset{\sim}{\beta}}_2 \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ X_2'\underset{\sim}{y} - n\bar{y}\bar{\underset{\sim}{x}} \end{bmatrix}$$

Letting $\underbrace{(X_2'X_2 - n\bar{\underline{x}}\,\bar{\underline{x}}')}_{X_D'X_D}\hat{\beta}_2 = \underbrace{X_2'\underline{Y} - n\bar{Y}\bar{\underline{x}}}_{X_D'Y_D}$

$$\Longrightarrow \quad X_D'X_D\hat{\beta}_2 = X_D'Y_D$$

where $X_D = (I - \frac{1}{n}J)X_2$; $\underline{Y_D} = (I - \frac{1}{n}J)\underline{Y}$

are called matrices with deviations from the means.

Let $S_D = X_D'X_D$,

$$\Longrightarrow \quad S_D\hat{\beta}_2 = \underline{s}_D \qquad (7.5.4)$$

In fact,

$$S_D = \begin{bmatrix} \Sigma(x_{i1}-\bar{x}_1)^2 & \Sigma(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2) & \cdots & \Sigma(x_{i1}-\bar{x}_1)(x_{i,p-1}-\bar{x}_{p-1}) \\ \Sigma(x_{i2}-\bar{x}_1)(x_{i1}-\bar{x}_1) & \Sigma(x_{i2}-\bar{x}_2)^2 & \cdots & \Sigma(x_{i2}-\bar{x}_2)(x_{i,p-1}-\bar{x}_{p-1}) \\ \cdots & \cdots & \cdots & \cdots \\ \Sigma(x_{i,p-1}-\bar{x}_{p-1})(x_{i1}-\bar{x}_1) & \cdots & \cdots & \Sigma(x_{i,p-1}-\bar{x}_{p-1})^2 \end{bmatrix}$$

$$\underline{s}_D = \begin{bmatrix} \Sigma(x_{i1}-\bar{x}_1)(y_i-\bar{y}) \\ \Sigma(x_{i2}-\bar{x}_2)(y_i-\bar{y}) \\ \vdots \\ \Sigma(x_{i,p-1}-\bar{x}_{p-1})(y_i-\bar{y}) \end{bmatrix} \qquad (7.5.5)$$

After solving for $\hat{\beta}_2$ from (7.5.4), ~~we get~~ from

$$n\hat{\beta}_0 + n\bar{\underline{x}}'\hat{\beta}_2 = n\bar{y}$$

$$\Longrightarrow \quad \hat{\beta}_0 = \bar{y} - \bar{\underline{x}}'\hat{\beta}_2 = \bar{y} - \hat{\beta}_2'\bar{\underline{x}}$$

Note: (7.5.4) are called the deviation normal equations and $Y_D'Y_D$ are needed to test $H_0: H_2\beta_2 = \underline{h}_2$ or finding confidence intervals on $\ell_2'\beta_2$.

Thm 7.5.1. In the GLM: $\underline{Y} = \underline{1}\beta_0 + X_2\underline{\beta}_2 + \underline{\varepsilon}$, with $\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I)$, to test $H_0: H_2\underline{\beta}_2 = \underline{h}_2$ vs $H_1: H_2\underline{\beta}_2 \neq$ where H_2 is a $q\times(p-1)$ known matrix of rank q, $0<q<p$, the test statistic is:

$$W = \frac{(H_2\hat{\underline{\beta}}_2 - \underline{h}_2)'(H_2 S_D^{-1} H_2')^{-1}(H_2\hat{\underline{\beta}}_2 - \underline{h}_2)}{\underline{Y}_D'\underline{Y}_D - \hat{\underline{\beta}}_2' X_D' \underline{Y}_D}\left(\frac{n-p}{q}\right) \quad (7.5.7)$$

where $\hat{\underline{\beta}}_2$ is obtained from the deviation normal equations in (7.5.4).

Proof: The problem is equivalent to the testing problem in $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ for $H_0: H\underline{\beta} = \underline{h}$ with specifically, $H = [\underline{0}\ \ H_2]$, $\underline{h}_2 = \underline{h}$

and ~~where~~

$$\hat{\underline{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\underline{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ S_D^{-1} X_D' \underline{Y}_D \end{bmatrix}$$

So, we just need to find W given by (6.3.9) with the specifics in this case.

Now, $H\hat{\underline{\beta}} - \underline{h} = [\underline{0}\ \ H_2]\begin{bmatrix} \hat{\beta}_0 \\ \hat{\underline{\beta}}_2 \end{bmatrix} - \underline{h}_2 = H_2\hat{\underline{\beta}}_2 - \underline{h}_2$

$$H(X'X)^{-1}H' = [\underline{0}\ \ H_2]\begin{bmatrix} n & n\bar{\underline{x}}' \\ n\bar{\underline{x}} & X_2'X_2 \end{bmatrix}^{-1}\begin{bmatrix} \underline{0}' \\ H_2' \end{bmatrix}$$

$$= [\underline{0}\ \ H_2]\begin{bmatrix} * & ** \\ **' & \left(X_2'X_2 - n\bar{\underline{x}}(\tfrac{1}{n})n\bar{\underline{x}}'\right)^{-1} \end{bmatrix}\begin{bmatrix} \underline{0}' \\ H_2' \end{bmatrix}$$

$$= H_2(X_2'X_2 - n\bar{\underline{x}}\bar{\underline{x}}')^{-1}H_2'$$

$$= H_2 S_D^{-1} H_2'$$

Further,

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

$$= \frac{1}{n-p}(\underline{Y}_D'\underline{Y}_D + n\bar{y}^2 - \hat{\underline{\beta}}'X'\underline{Y})$$

and

$$\hat{\underline{\beta}}'X'\underline{Y} = [\hat{\beta}_0 \;\; \hat{\underline{\beta}}_2'] \begin{bmatrix} \underline{1}' \\ X_2' \end{bmatrix} \underline{Y}$$

$$= [\hat{\beta}_0 \;\; \hat{\underline{\beta}}_2'] \begin{bmatrix} \underline{1}'\underline{Y} \\ X_2'\underline{Y} \end{bmatrix}$$

$$= [\hat{\beta}_0 \;\; \hat{\underline{\beta}}_2'] \begin{bmatrix} n\bar{y} \\ X_2'\underline{Y} \end{bmatrix}$$

$$= n\hat{\beta}_0\bar{y} + \hat{\underline{\beta}}_2'X_2'\underline{Y} \quad \cdots\cdots (*)$$

$$= n(\bar{y} - \hat{\underline{\beta}}_2'\bar{\underline{x}})\bar{y} + \hat{\underline{\beta}}_2'X_2'\underline{Y}$$

$$= n\bar{y}^2 - n\hat{\underline{\beta}}_2'\bar{\underline{x}}\bar{y} + \hat{\underline{\beta}}_2'X_2'\underline{Y}$$

$$= n\bar{y}^2 + \hat{\underline{\beta}}_2'(X_2'\underline{Y} - n\bar{\underline{x}}\bar{y})$$

$$= n\bar{y}^2 + \hat{\underline{\beta}}_2'X_D'\underline{Y}_D \quad \cdots \quad (**)$$

$$\underline{Y}_D = (I - \tfrac{1}{n}J)\underline{Y}$$

$$\underline{Y}_D'\underline{Y}_D = \underline{Y}'(I - \tfrac{1}{n}J')(I - \tfrac{1}{n}$$

$$= \underline{Y}'(I - \tfrac{1}{n}J - \tfrac{1}{n}J + \tfrac{1}{n\,n}J$$

$$\cancel{= \underline{Y}'\underline{Y} - \underline{Y}'\bar{y} - \bar{y}\underline{Y}}$$

$$= \underline{Y}'(I - \tfrac{2}{n}J + \tfrac{1}{n}J)\underline{Y}$$

$$= \underline{Y}'(I - \tfrac{1}{n}J)\underline{Y}$$

$$= \underline{Y}'\underline{Y} - \tfrac{1}{n}\underline{Y}'J\underline{Y}$$

$$= \underline{Y}'\underline{Y} - \tfrac{1}{n}\underline{Y}'n\bar{y}\underline{1}$$

$$= \underline{Y}'\underline{Y} - \tfrac{n}{n}\bar{y}\,n\bar{y}$$

$$= \underline{Y}'\underline{Y} - n\bar{y}^2$$

$$\Longrightarrow \quad \hat{\sigma}^2 = \frac{1}{n-p}(\underline{Y}_D'\underline{Y}_D + n\bar{y}^2 - n\bar{y}^2 - \hat{\underline{\beta}}_2'X_D'\underline{Y}_D)$$

$$= \frac{1}{n-p}(\underline{Y}_D'\underline{Y}_D - \hat{\underline{\beta}}_2'X_D'\underline{Y}_D)$$

$$\Longrightarrow \quad W = \frac{(H\hat{\underline{\beta}} - \underline{h})'[H(X'X)^{-1}H']^{-1}(H\hat{\underline{\beta}} - \underline{h})}{\hat{\sigma}^2} \; \frac{n-p}{q}$$

$$= \frac{(H_2\hat{\underline{\beta}}_2 - \underline{h}_2)'(H_2 S_D^{-1} H_2')^{-1}(H_2\hat{\underline{\beta}}_2 - \underline{h}_2)}{\underline{Y}_D'\underline{Y}_D - \hat{\underline{\beta}}_2'X_D'\underline{Y}_D} \; \frac{n-p}{q}.$$

//.

Thm 7.5.2 (read).

- To use the R( ) notation for this GLM, we have the following:

Thm 7.5.3

Consider the GLM $\underline{Y} = \underline{1}\beta_0 + X_2\underline{\beta}_2 + \underline{\varepsilon}$, $\underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I)$. To test $H_0: \underline{\beta}_2 = \underline{0}$ vs $H_1: \underline{\beta}_2 \neq \underline{0}$, the quantity $R(\underline{\beta}_2|$ the numerator in W statistic can be written as:

$$R(\underline{\beta}_2|\beta_0) = R(\beta_0, \underline{\beta}_2) - R(\beta_0) = R(\underline{\beta}) - R(\beta_0)$$
$$= \underline{Y}_D' X_D X_D^- \underline{Y}_D \quad , \text{ where } R(\beta_0) \triangleq n\bar{y}^2$$

Proof: $R(\underline{\beta}_2|\beta_0) = R(\underline{\beta}) - R(\beta_0)$

$$= \hat{\underline{\beta}}' X' \underline{y} - \hat{\beta}_0 \underline{1}' \underline{y}$$
$$= \hat{\underline{\beta}}' X' \underline{y} - \hat{\beta}_0 \sum y_i$$
$$= \hat{\underline{\beta}}' X' \underline{y} - n\bar{y}^2$$

From (**) on page 7-17:

$$R(\underline{\beta}_2|\beta_0) = \hat{\underline{\beta}}_2' X_D' \underline{Y}_D$$
$$= [(X_D' X_D)^{-1} X_D' \underline{Y}_D]' X_D' \underline{Y}_D$$
$$= \underline{Y}_D' X_D (X_D' X_D)^{-1} X_D' \underline{Y}_D$$
$$= \underline{Y}_D' X_D X_D^- \underline{Y}_D$$

//.