Question #1.

Solution 1. *Exercise* 8.5*:*

We want to show:

$$E(SSR/k) = \sigma^2 + (1/k)\beta_1'\mathbf{X}_c'\mathbf{X}_c\beta_1$$

(a): approach the proof by using:

$$E[\mathbf{y}'\mathbf{A}\mathbf{y}] = tr(\mathbf{A}\boldsymbol{\Sigma}) + \mu'\mathbf{A}\mu$$

for symmetric matrix $\mathbf{A}$*.*

We have:

$$\begin{gathered} SSR = \mathbf{y}'\mathbf{X}_c(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'\mathbf{y} = \mathbf{y}\mathbf{H}_c\mathbf{y} \\ \textit{with} \\ \mathbf{H}_c = \mathbf{X}_c(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c' \end{gathered}$$

Since we know from Theorem 8.1*(a) that* $\mathbf{H}$ *is idempotent with rank* k*, thus*

$$tr(\mathbf{H}_c\Sigma) = tr(\sigma^2\mathbf{H}_c) = k\sigma^2$$

and we claim before hand that $\mathbf{j}'\mathbf{X}_c = \mathbf{0}'$ *due to the fact that the column sums of* $\mathbf{X}_c$ *are all* 0*. We will use this soon. Also we have:*

$$E[\mathbf{y}] = \mathbf{X}\beta = (\mathbf{j}, \mathbf{X}_c)\begin{pmatrix}\alpha \\ \beta_1\end{pmatrix} = \alpha\mathbf{j} + \mathbf{X}_c\beta_1$$

so we have the following computation:

$$\begin{aligned} E[SSR] &= E[\mathbf{y}'\mathbf{H}_c\mathbf{y}] = E[\mathbf{y}'\mathbf{X}_c(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'\mathbf{y}] \\ &= tr(\sigma^2\mathbf{H}_c) + \Big(\mathbf{X}\beta\Big)'\mathbf{X}_c\Big(\mathbf{X}_c'\mathbf{X}_c\Big)^{-1}\mathbf{X}_c'\Big(\mathbf{X}\beta\Big) \\ &= k\sigma^2 + \Big(\alpha\mathbf{j}' + \beta_1'\mathbf{X}_c'\Big)\mathbf{X}_c\Big(\mathbf{X}_c'\mathbf{X}_c\Big)^{-1}\mathbf{X}_c'\Big(\alpha\mathbf{j} + \mathbf{X}_c\beta_1\Big) \\ &= k\sigma^2 + \alpha^2\mathbf{j}'\mathbf{X}_c\Big(\mathbf{X}_c'\mathbf{X}_c\Big)^{-1}\mathbf{X}_c\mathbf{j} + 2\alpha\mathbf{j}'\mathbf{X}_c\Big(\mathbf{X}_c'\mathbf{X}_c\Big)^{-1}\mathbf{X}_c'\mathbf{X}_c\beta_1 \\ &\qquad + \beta_1'\mathbf{X}_c'\mathbf{X}_c\Big(\mathbf{X}_c'\mathbf{X}_c\Big)^{-1}\mathbf{X}_c'\mathbf{X}_c\beta_1 \\ &= k\sigma^2 + 0 + 0 + \beta_1'\mathbf{X}_c'\mathbf{X}_c\beta_1 \end{aligned}$$

So

$$E(SSR/k) = \sigma^2 + (1/k)\beta_1'\mathbf{X}_c'\mathbf{X}_c\beta_1$$

(b): approach the proof by using the noncentral chi square distribution result:

By Theorem 8.1*b we know that* SSR/σ^2 *is* $\chi^2(k, \lambda_1)$ *with*

$$\lambda_1 = \frac{1}{2\sigma^2}\beta_1'\mathbf{X}_c'\mathbf{X}_c\beta_1$$

So by theorem 5.23(b) there is:

$$E[SSR/\sigma^2] = k + 2\lambda = k + \frac{1}{\sigma^2}\beta_1'\mathbf{X}_c'\mathbf{X}_c\beta_1$$

Thus

$$\begin{aligned} E[SSR/k] &= \frac{\sigma^2}{k}\Big(k + \frac{1}{\sigma^2}\beta_1'\mathbf{X}_c'\mathbf{X}_c\beta_1\Big) \\ &= \sigma^2 + \frac{1}{k}\beta_1'\mathbf{X}_c'\mathbf{X}_c\beta_1 \end{aligned}$$

Exercise 8.9:

Given the set up $\mathbf{y}$ *is* $N_n(\mathbf{X}\beta, \sigma^2\mathbf{I})$, *we know that* $\mathbf{y}'(\mathbf{H}-\mathbf{H}_1)\mathbf{y}/\sigma^2$ *is* $\chi^2(h, \lambda_1)$, *with* $\lambda_1 = \frac{1}{2\sigma^2}\mu'\Big(\mathbf{H}-\mathbf{H}_1\Big)\mu$ *according to corollary* 2 *of Theorem* 5.5.

We keep in mind that:

$$\begin{aligned} \mathbf{X} &= \mathbf{H}\mathbf{X} \\ \mathbf{X}_1 &= \mathbf{H}\mathbf{X}_1 \\ \mathbf{X}_1 &= \mathbf{H}_1\mathbf{X}_1 \\ \mathbf{X}_2 &= \mathbf{H}\mathbf{X}_2 \end{aligned}$$

Then we have the following computation:

$$\begin{aligned} &\mu'(\mathbf{H}-\mathbf{H}_1)\mu \\ &= (\mathbf{X}\beta)'\Big(\mathbf{H}-\mathbf{H}_1\Big)(\mathbf{X}\beta) \\ &= \Big[(\mathbf{X}_1, \mathbf{X}_2)\begin{pmatrix}\beta_1 \\ \beta_2\end{pmatrix}\Big]'\Big(\mathbf{H}-\mathbf{H}_1\Big)\Big[(\mathbf{X}_1, \mathbf{X}_2)\begin{pmatrix}\beta_1 \\ \beta_2\end{pmatrix}\Big] \\ &= (\beta_1'\mathbf{X}_1' + \beta_2'\mathbf{X}_2')\Big(\mathbf{H}-\mathbf{H}_1\Big)(\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2) \\ &= \Big[\beta_1'\mathbf{X}_1'\mathbf{H} - \beta_1'\mathbf{X}_1'\mathbf{H}_1 + \beta_2'\mathbf{X}_2'\mathbf{H} - \beta_2'\mathbf{X}_2'\mathbf{H}_1\Big](\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2) \\ &= \Big(\beta_1'\mathbf{X}_1' - \beta_1'\mathbf{X}_1' + \beta_2'\mathbf{X}_2' - \beta_2'\mathbf{X}_2'\mathbf{H}_1\Big)(\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2) \\ &= \Big(\beta_2'\mathbf{X}_2' - \beta_2'\mathbf{X}_2'\mathbf{H}_1\Big)(\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2) \\ &= \cancel{\beta_2'\mathbf{X}_2'\mathbf{X}_1\beta_1} + \beta_2'\mathbf{X}_2'\mathbf{X}_2\beta_2 - \cancel{\beta_2'\mathbf{X}_2'\mathbf{X}_1\beta_1} - \beta_2'\mathbf{X}_2'\mathbf{H}_1\mathbf{X}_2\beta_2 \\ &= \beta_2'\Big(\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{H}_1\mathbf{X}_2\Big)\beta_2 \\ &= \beta_2'\Big(\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\Big)\beta_2 \end{aligned}$$

divide both sides by $2\sigma^2$ *then we finished the proof.*

Exercise 8.11*:*

Continue from above, since by Theorem 8.2*b we know that:*

$$SS(\beta_2\ \beta_1)/\sigma^2 = \mathbf{y}'(\mathbf{H}-\mathbf{H}_1)\mathbf{y}/\sigma^2 \text{ is } \chi^2(h, \lambda_1)$$

with

$$\lambda_1 = \beta_2'\Big(\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\Big)\beta_2/2\sigma^2$$

as we have just verified in exercise 8.9.

Then according to Theorem 5.3*(b), we have:*

$$\begin{aligned} E[SS(\beta_2|\beta_1)]/\sigma^2 &= h + 2\lambda_1 \\ &= h + \beta_2'\Big(\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\Big)\beta_2/\sigma^2 \end{aligned}$$

Multiply both sides of the equation above by $\frac{\sigma^2}{h}$ *then we completed the proof.*

Now for the general linear hypothesis, we have from Theorem 8.4*(a):*

$$\begin{aligned} &\frac{SSH}{\sigma^2} \sim \chi^2(q,\lambda) \\ &\textit{with } \lambda = (\mathbf{C}\beta)'\Big[\mathbf{C}\Big(\mathbf{X}'\mathbf{X}\Big)^{-1}\mathbf{C}'\Big]^{-1}\mathbf{C}\beta/2\sigma^2 \end{aligned}$$

So we have:

$$E[\frac{SSH}{\sigma^2}] = q + 2\lambda = q + (\mathbf{C}\beta)'\Big[\mathbf{C}\Big(\mathbf{X}'\mathbf{X}\Big)^{-1}\mathbf{C}'\Big]^{-1}\mathbf{C}\beta/\sigma^2$$

Then multiply both sides by $\frac{\sigma^2}{q}$*, we get equation* (8.28).

Question 2.

Solution 2. *For part [A]:*

(i)-(v) we use the general linear hypothesis $H_0 : \mathbf{C}\beta = 0$ *(we could use Ful-Reduced model for (i) but it is just easier to do all with the same).*

We have:

$$\begin{aligned} \mathbf{C}_1 &= (0,1,0,0,0) \\ \mathbf{C}_2 &= \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \\ \mathbf{C}_3 &= \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \\ \mathbf{C}_4 &= (0,1,0,-2,-2) \\ \mathbf{C}_5 &= (0,-3,1,-3,-3) \textit{ with } t = 0.25 \end{aligned}$$

For $(i) - (iv)$ *our test statistic is:*

$$F = \frac{SSH/q}{SSE/(n-k-1)} = \frac{(\mathbf{C}\hat{\beta})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\hat{\beta})/q}{\mathbf{y}'(\mathbf{I}-\mathbf{H})\mathbf{y}/(n-k-1)} \sim F(q, n-k-1) \textit{ under null}$$

For (v) our test statistic is:

$$F = \frac{SSH/q}{SSE/(n-k-1)} = \frac{(\mathbf{C}\hat{\beta} - \mathbf{t})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\hat{\beta} - \mathbf{t})/q}{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}/(n-k-1)} \sim F(q, n-k-1) \text{ under null}$$

We reject for large value of F.

We intput the data and compute $\hat{\beta}$ under proc iml:

```
proc iml;

/*use the senic data we imported*/
use data.senic;
/*read in the average length of stay into Y column*/
read all var {Length_of_stay} into Y;
/*read the 4 required variables into X matrix*/
read all var{age infection_risk available_facilities_and_service
           routine_chest_x_ray_ratio} into X;

/*set the number of parametsrs*/;
k = 4;

/*find the number of observations*/
n = nrow(X[, 1]);
print n;

/*create column of 1s and insert into X*/
intercept = j(n,1,1);
print intercept;
X_b = X;
X_a = intercept||X;

/*check dimension of design matrix*/
d_b = dimension(X_b);
d_a = dimension(X_a);
print d_b d_a;

/*compute beta hat*/
beta_hat = inv(t(X_a)*X_a)*t(X_a)*Y;
print beta_hat;
```

We then set up the matrix $\mathbf{C}$, compute SSE and SSH for $(i)-(v)$:

```
/*make the C matrix*/
c_1 = {0 1 0 0 0};
q_1 = 1;
c_2 = {0 1 0 0 0,
        0 0 0 1 0 };
q_2 = 2;
c_3 = {0 0 1 0 0,
        0 0 0 1 0,
       0 0 0 0 1};
q_3 = 3;
c_4 = {0 1 0 -2 -2};
q_4 = 1;
c_5 = {0 -3 1 -3 -3};
q_5 = 1;
t = 0.25;

/*compute SSE and SSH*/

/*create identity matrix*/
I = i(n);
print I;
/*compute H matrix*/
H = X_a*inv(t(X_a)*X_a)*t(X_a);
/*compute SSE*/
SSE = t(Y)*(I - H)*Y;
print SSE;
/*compute SSH*/
SSH_1 = t(c_1*beta_hat)*inv((c_1*inv(t(X_a)*X_a)*t(c_1)))*(c_1*beta_hat);
SSH_2 = t(c_2*beta_hat)*inv((c_2*inv(t(X_a)*X_a)*t(c_2)))*(c_2*beta_hat);
SSH_3 = t(c_3*beta_hat)*inv((c_3*inv(t(X_a)*X_a)*t(c_3)))*(c_3*beta_hat);
SSH_4 = t(c_4*beta_hat)*inv((c_4*inv(t(X_a)*X_a)*t(c_4)))*(c_4*beta_hat);
SSH_5 = t(c_5*beta_hat - t)*inv((c_5*inv(t(X_a)*X_a)*t(c_5)))*(c_5*beta_hat - t);
```

We then compute the F statistics and p value for each test:

```
/*compute F statistic and p-value*/;
F_1 = SSH_1/q_1/(SSE/(n - k - 1));
p_1 = 1 - CDF('F', F_1, q_1, n-k-1);
print F_1 p_1;

F_2 = SSH_2/q_2/(SSE/(n - k - 1));
p_2 = 1 - CDF('F', F_2, q_2, n-k-1);
print F_2 p_2;

F_3 = SSH_3/q_3/(SSE/(n - k - 1));
p_3 = 1 - CDF('F', F_3, q_3, n-k-1);
print F_3 p_3;

F_4 = SSH_4/q_4/(SSE/(n - k - 1));
p_4 = 1 - CDF('F', F_4, q_4, n-k-1);
print F_4 p_4;

F_5 = SSH_5/q_5/(SSE/(n - k - 1));
p_5 = 1 - CDF('F', F_5, q_5, n-k-1);
print F_5 p_5;
```

The output is the following:

F_1	p_1
6.9086163	0.0098277

F_2	p_2
5.782492	0.004114

F_3	p_3
19.655422	3.059E-10

F_4	p_4
0.0019765	0.9646213

F_5	p_5
0.3915301	0.5328163

For $\alpha = 0.05$, we would reject that $H_0 : \beta_1 = 0$, reject $H_0 : \beta_1 = \beta_3 = 0$, reject $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$, fail to reject $H_0 : \beta_1 = 2(\beta_3 + \beta_4)$ and fail to reject $H_0 : \beta_2 - 3(\beta_1 + \beta_3 + \beta_4) = 0.25$.

Now for part (vi):

We are doing simultaenoues tests for $H_{0i} : \mathbf{a}_i\beta = 0$ with

$$\begin{aligned}
\mathbf{a}_1 &= (0, 1, 0, 0, 0) \\
\mathbf{a}_2 &= (0, 0, 1, 0, 0) \\
\mathbf{a}_3 &= (0, 0, 0, 1, 0) \\
\mathbf{a}_4 &= (0, 0, 0, 0, 1)
\end{aligned}$$

Our F statistic is:

$$F_i = \frac{(\mathbf{a}_i\hat{\beta})'\left[\mathbf{a}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}_i'\right]^{-1}\mathbf{a}_i\hat{\beta}}{SSE/(n-k-1)} \sim F(1, n-k-1) \text{ under null}$$

The following code copute the F statistic and the Bonferronni p-values:

```
/*Simultaneous Tests*/

/*Bonferroni and */
a_1 = {0 1 0 0 0 };
a_2 = {0 0 1 0 0 };
a_3 = {0 0 0 1 0 };
a_4 = {0 0 0 0 1 };
/*compute F statistic and p values*/
F_critical_bon = finv((1 - 0.05/4), 1, n - k - 1);
F_a1 = t(a_1*beta_hat)*inv((a_1*inv(t(X_a)*X_a)*t(a_1)))*(a_1*beta_hat)/(SSE/(n - k - 1));
p_a1 = 1 - CDF('F', F_a1, 1, n - k - 1);
F_a2 = t(a_2*beta_hat)*inv((a_2*inv(t(X_a)*X_a)*t(a_2)))*(a_2*beta_hat)/(SSE/(n - k - 1));
p_a2 = 1 - CDF('F', F_a2, 1, n - k - 1);
F_a3 = t(a_3*beta_hat)*inv((a_3*inv(t(X_a)*X_a)*t(a_3)))*(a_3*beta_hat)/(SSE/(n - k - 1));
p_a3 = 1 - CDF('F', F_a3, 1, n - k - 1);
F_a4 = t(a_4*beta_hat)*inv((a_4*inv(t(X_a)*X_a)*t(a_4)))*(a_4*beta_hat)/(SSE/(n - k - 1));
p_a4 = 1 - CDF('F', F_a4, 1, n - k - 1);
print F_critical_bon;
print F_a1 p_a1 F_a2 p_a2 F_a3 p_a3 F_A4 p_a4;
```

Output is:

F_critical_bon
6.4527594

F_a1	p_a1	F_a2	p_a2	F_a3	p_a3	F_a4	p_a4
6.9086163	0.0098277	15.1433	0.0001727	5.1953445	0.0246131	5.4494574	0.0214239

Keep in mind here we have the family wise sigificance level $\alpha_f = 0.05$ and comparions wise significan level $\alpha_c = \alpha_f/d = 0.05/4 = 0.0125$.

So from above output we will reject $H_{01} : \beta_1 = 0$, reject $H_{02} : \beta_2 = 0$, fail to reject $H_{03} : \beta_3 = 0$ and fail to reject $H_{04} : \beta_4 = 0$.

For the same F statistic, we also have:

$$\max_{1 \leq i \leq 4} F_i \sim (k+1)F(k+1, n-k-1)$$

To use Scheffe, we reject $H_{0i} : \beta_i = 0$ when $F_i \geq (k+1)F_{\alpha,k+1,n-k-1}$ The following code compare the left and right hand side of the inequality above:

```
/*Scheffe*/
/*compute (k + 1)F(alpha, k + 1,n - k - 1)*/
F_critical_s = (k + 1)*finv(1 - 0.05, k + 1, n- k - 1);
print F_a1 F_a2 F_a3 F_a4 F_critical_s;
```

The output is:

F_a1	F_a2	F_a3	F_a4	F_critical_s
6.9086163	15.1433	5.1953445	5.4494574	11.492154

so we will fail to reject $H_{01}: \beta_1 = 0$, reject $H_{02}: \beta_2 = 0$, fail to reject $H_{03}: \beta_3 = 0$ and fail to reject $H_{04}: \beta_4 = 0$ under scheffe method.

Thus completed the solution for part [A].

For part [B]:

For part i: The individual confidence intervals for β_j assume the following form:

$$\hat{\beta}_j \pm t_{\alpha/2,n-k-1} s \sqrt{g_{jj}}$$

In our previous code we already have $\hat{\beta}_j$ and SSE, so $s = SSE/(n-k-1)$, and we only need to extract

$$g_{jj} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}_{jj}, j = 1, 2, 3, 4$$

Our SAS code:

```
/*individual confidence intervals for betas*/
g = inv(t(X_a)*X_a);
print g;
g11 = g[2, 2];
g22 = g[3, 3];
g33 = g[4, 4];
g44 = g[5, 5];
print g11 g22 g33 g44;
s = sqrt(SSE/(n - k - 1));
lower_1 = beta_hat[2] - tinv(0.975, n - k - 1)*s*sqrt(g11);
upper_1 = beta_hat[2] + tinv(0.975, n - k - 1)*s*sqrt(g11);
lower_2 = beta_hat[3] - tinv(0.975, n - k - 1)*s*sqrt(g22);
upper_2 = beta_hat[3] + tinv(0.975, n - k - 1)*s*sqrt(g22);
lower_3 = beta_hat[4] - tinv(0.975, n - k - 1)*s*sqrt(g33);
upper_3 = beta_hat[4] + tinv(0.975, n - k - 1)*s*sqrt(g33);
lower_4 = beta_hat[5] - tinv(0.975, n - k - 1)*s*sqrt(g44);
upper_4 = beta_hat[5] + tinv(0.975, n - k - 1)*s*sqrt(g44);
print lower_1 upper_1 lower_2 upper_2 lower_3 upper_3 lower_4 upper_4;
```

The matrix $(\mathbf{X}'\mathbf{X})^{-1}$:

g

1.5288257	-0.024251	-0.001829	-0.001625	-0.001849
-0.024251	0.0004497	-0.000054	6.857E-6	3.0332E-6
-0.001829	-0.000054	0.0075093	-0.000243	-0.000215
-0.001625	6.857E-6	-0.000243	0.0000471	3.5339E-6
-0.001849	3.0332E-6	-0.000215	3.5339E-6	0.0000303

The estiamte $\hat{\beta}$:

beta_hat
0.1801384
0.0856936
0.5184167
0.0240421
0.0197392

Confidence interval:

lower_1	upper_1	lower_2	upper_2	lower_3	upper_3	lower_4	upper_4
0.0210695	0.1503176	0.254352	0.7824815	0.0031344	0.0449498	0.0029784	0.0365

or explicitly, the 95% *confidence interval for the individual parameters are:*

$$\begin{aligned}
&\textit{for } \beta_1 : (0.0210695, 0.1503176)\\
&\textit{for } \beta_2 : (0.254352, 0.7824815)\\
&\textit{for } \beta_3 : (0.0031344, 0.0449498)\\
&\textit{for } \beta_4 : (0.0029784, 0.0365)
\end{aligned}$$

For part (*ii*)*:*

the general set up for the confidence interval for $\mathbf{a}'\beta$ *is:*

$$\mathbf{a}'\hat{\beta} \pm t_{\alpha/2,n-k-1} s \sqrt{\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}}$$

here we want to estimate the confidence interval for $\beta_1 - 2(\beta_3 + \beta_4)$ *hence*

$$\mathbf{a} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -2 \\ -2 \end{bmatrix}$$

Our code is:

```
/*confidence interval for beta_1 - 2beta_3- 2beta_4*/
a = t({0 1 0 -2 -2});

lower_a = t(a)*beta_hat - tinv(0.975, n -  k - 1)*s*sqrt(t(a)*g*a);
upper_a = t(a)*beta_hat + tinv(0.975, n -  k - 1)*s*sqrt(t(a)*g*a);
print lower_a upper_a;
```

Output is:

```
lower_a    upper_a

-0.085198 0.0814596
```

So our confidence interval for $\beta_1 - 2\beta_3 - 2\beta_4$ *is:*

$$(-0.085198, 0.0814596)$$

For part (iii)*:*

We have the general formula for $100(1-\alpha)\%$ *confidence interval of* σ^2 *as:*

$$\frac{(n-k-1)s^2}{\chi^2_{\alpha/2,n-k-1}} \leq \sigma^2 \leq \frac{(n-k-1)s^2}{\chi^2_{1-\alpha/2,n-k-1}}$$

Our code is:

```
/*confidence interval for sigma^2*/
s_square = SSE/(n- k - 1);
lower_sigma = (n - k - 1)*s_square/cinv(0.975, n - k - 1);
upper_sigma = (n - k - 1)*s_square/cinv(0.025, n - k - 1);
print lower_sigma upper_sigma;
```

Output:

```
lower_sigma upper_sigma

1.8409298    3.146023
```

So the 95% *confidence interval for* σ^2 *is:*

$$(1.8409298, 3.146023)$$

For part [C]:

The prediction interval for $\mathbf{y}_0 = \mathbf{x}_0\beta + \epsilon$ *is given by the following formula:*

$$\mathbf{x}_0'\hat{\beta} \pm t_{\alpha/2,n-k-1}s\sqrt{1 + \mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x_0}}$$

here since we are only using age as predictor, we have $k = 1$*, and also we need to re-compute* $\hat{\beta}$ *and SSE.*

We also have

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 67 \end{bmatrix}$$

We have the folowing code:

```
/*for part [C]*/

/*use only age X_1 as the predictor variable*/
read all var{age} into X_1;
X_1 = intercept||X_1;
beta_hat_1 = inv(t(X_1)*X_1)*t(X_1)*Y;
print beta_hat_1;
k_1 = 1;
H_1 = X_1*inv(t(X_1)*X_1)*t(X_1);
SSE_1 = t(Y)*(I - H_1)*Y;
print k_1;
s_1 = sqrt(SSE_1/(n - k_1 - 1));
/*given patient 67 years old*/
x_01 = {1 67};
/*compute prediction interval*/
lower_x_1 = x_01 * beta_hat_1 - tinv(0.975, n - k_1 - 1)*s_1*sqrt(1 + x_01*inv(t(X_1)*X_1)*t(x_01));
upper_x_1 = x_01 * beta_hat_1 + tinv(0.975, n - k_1 - 1)*s_1*sqrt(1 + x_01*inv(t(X_1)*X_1)*t(x_01));
print lower_x_1 upper_x_1;
```

The output is:

lower_x_1	upper_x_1
6.8550178	14.670272

So the prediction interval for a patient of age 67 *is:*

$$(6.8550178, 14.670272)$$

For part [*D*]*:*

The maximum likelihood estimate of β *and* σ^2 *under the null* $H_0 : \mathbf{C}\beta = 0$ *is given by:*

$$\hat{\beta}_0 = \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\Big[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\Big]^{-1}\mathbf{C}\hat{\beta}$$
$$\hat{\sigma}_0^2 = \hat{\sigma}^2 + \frac{1}{n}(\mathbf{C}\hat{\beta})'\Big[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\Big]^{-1}\mathbf{C}\hat{\beta}$$

where

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$
$$\hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})/n$$

are the MLE under alternative $H_1 : \mathbf{C}\beta \neq 0$*. We then have:*

$$\begin{aligned}
LR &= \frac{\max_{H_0} L(\beta, \sigma^2)}{\max_{H_1} L(\beta, \sigma^2)} \\
&= \Big[\frac{SSE}{SSE + (\mathbf{C}\hat{\beta})'[C(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\beta}}\Big]^{\frac{n}{2}} \\
&= \Big[\frac{1}{1 + SSH/SSE}\Big]^{n/2} \\
&= \Big[\frac{1}{1 + qF/(n-k-1)}\Big]^{n/2}
\end{aligned}$$

We use χ^2 approximation for $-2\log LR$ and show that it gives the same p-value as in the F test we did in part [A](iv).

We have $-2logLR \sim \chi^2(1)$. We have the following code:

```
/*for part [D]*/

/*compute likelihood ratio*/
LR = (1/(1 + q_4*F_4/(n - k - 1)))**(n/2);

test = -2*log(LR);
print test;
/*use chi square approximation for -2 log LR to obtain p value*/;
p_LR = 1 - probchi(-2*log(LR), 1);
print p_LR;
```

The output for $-2\log LR$(chi square test statistic) and p-value is:

```
test

0.002068


p_LR

0.9637284
```

So we have $-2\log LR = 0.002068$ and p value 0.9637284, which is about identical to the p value we have earlier with F test(0.9646). The difference is due to two factors. One is that our chi-square distribution is only approximate, and the other factor is rounding error. But we can see they are pretty close.

Thus finished Problem 2.

Problem 3.

Solution 3. *To detect outliers, we consider the following plots:*

We plot (i)studentized residuals against fitted value, (ii)deleted residuals against fitted value, (iii)ordinary residual against deleted residual:

We need to be careful that since $var(\hat{\epsilon}_i) = \sigma^2(1 - h_{ii})$ is not constant, we should scale it into studentized residual or deleted residual before we make a plot agasint fitted value:

Our code:

Input the data:

```
/*Question 3*/

/*create data*/
data Table7_5;
    input y x_1 x_2 x_3 @@;
    datalines;
    18.38   15.50   17.25   0.24    20.00   22.29   18.51   0.20
    11.50   12.36   11.13   0.12    25.00   31.84    5.54   0.12
    52.50   83.90    5.44   0.04    82.50   72.25   20.37   0.05
    25.00   27.14   31.20   0.27    30.67   40.41    4.29   0.10
    12.00   12.42    8.69   0.41    61.25   69.42    6.63   0.04
    60.00   48.46   27.40   0.12    57.50   69.00   31.23   0.08
    31.00   26.09   28.50   0.21    60.00   62.83   29.98   0.17
    72.50   77.06   13.59   0.05    60.33   58.83   45.46   0.16
    49.75   59.48   35.90   0.32     8.50    9.00    8.89   0.08
    36.50   20.64   23.81   0.24    60.00   81.40    4.54   0.05
    16.25   18.92   29.62   0.72    50.00   50.32   21.36   0.19
    11.50   21.33    1.53   0.10    35.00   46.85    5.42   0.08
    75.00   65.94   22.10   0.09    31.56   38.68   14.55   0.17
    48.50   51.19    7.59   0.13    77.50   59.42   49.86   0.13
    21.67   24.64   11.46   0.21    19.75   26.94    2.48   0.10
    56.00   46.20   31.62   0.26    25.00   26.86   53.73   0.43
    40.00   20.00   40.18   0.56    56.67   62.52   15.89   0.05
    ;
run;

proc print data= Table7_5;
run;
```

Import the data into proc iml:

```
proc iml;

/*use the senic data we imported*/
use work.Table7_5;
read all var {y} into Y;
read all var{x_1 x_2 x_3} into X;

/*set the number of parametsrs*/;
k = 3;

/*find the number of observations*/
n = nrow(X[, 1]);
print n;

/*create column of 1s and insert into X*/
intercept = j(n,1,1);
print intercept;
X = intercept||X;
print X;

/*check dimension of design matrix*/
d = dimension(X);
print d;
```

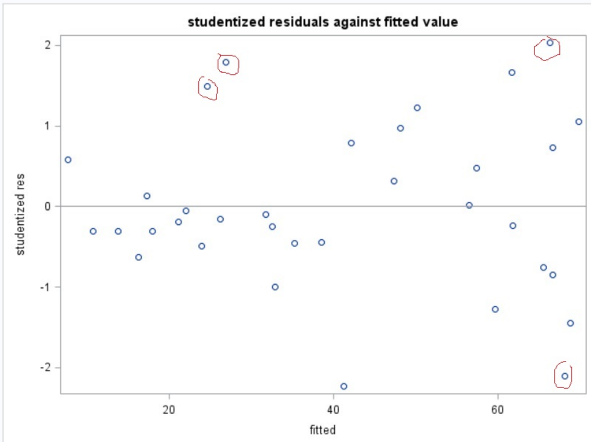
Compute necessary quantites: $\hat{\beta}$, SSE, $\hat{y}$, $\hat{e}_i$(residual),r_i(studentized residual) and $\hat{e}_{(i)}$(deleted residual):

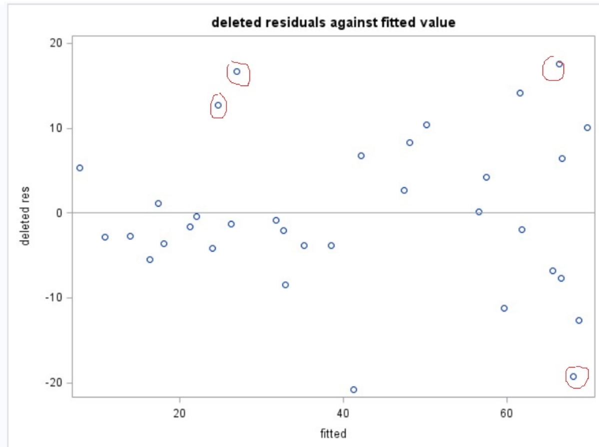
```
/*compute beta hat*/
beta_hat = inv(t(X)*X)*t(X)*Y;
print beta_hat;

/*create identity matrix*/
I = i(n);
print I;
/*compute H matrix*/
H = X*inv(t(X)*X)*t(X);
/*compute SSE*/
SSE = t(Y)*(I - H)*Y;
print SSE;
/*compute s*/
s = sqrt(SSE/(n - k - 1));

/*compute fitted values*/
y_hat = H*y;

/*compute the residuals*/

/*regular residuals*/
residual = (I - H)*Y;

/*studentized residuals*/
residual_stu = (inv(I - diag(H)))##(0.5)*residual/s;

/*deleted residuals*/
residual_del = inv(I - diag(H))*residual;
print residual_del;
```

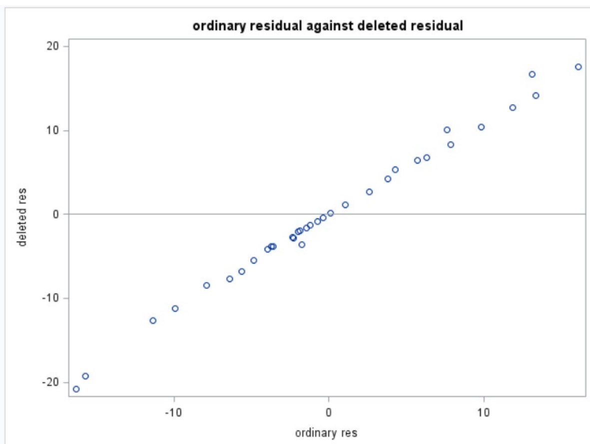
Make a scatter plot of studentized residuals against fitted value, and deleted residuals against fitted value, as well as deleted residual against ordinary residuals

```
/*scatter plot of residual against estimated mean*/
ods html;
ods graphics on;
title "studentized residuals against fitted value";
run Scatter(y_hat, residual_stu)
    /*add reference line*/
    other = "refline 0/axis = y"
    label={"fitted" "studentized res"}
    ;
title "deleted residuals against fitted value";
run Scatter(y_hat, residual_del)
    /*add refernece line*/
    other = "refline 0/axis = y"
    label={"fitted" "deleted res"}
    ;
title "ordinary residual against deleted residual";
run Scatter(residual, residual_del)
    /*add refernece line*/
    other = "refline 0/axis = y"
    label={"ordinary res" "deleted res"}
    ;
ods graphics close;
ods html close;
```

The output is:







As we can see that although on different scale, both residuals against fitted value scatter plots display the same pattern. And we marked with red circle for potential outliers.

On the other hand, the ordinary residual against deleted residual plot does not indicate any obvious candidate for outliers.

The data does not come with input index i, so we could not plot residuals against i here.

For influential observations, we compute the following things (will show code and output):

1. *Residuals: we compute studentized, studentized external, and deleted residuals.*
2. *PRESS: prediction sum of square*
3. *Cooks distance*

Our code is as following:

```
/*compute the residuals*/

/*regular residuals*/
residual = (I - H)*Y;

/*studentized residuals*/
residual_stu = (inv(I - diag(H)))##(0.5)*residual/s;

/*deleted residuals*/
residual_del = inv(I - diag(H))*residual;
print residual_del;

/*external studentized residuals*/

/*compute SSEs without a single observation*/
SSE_ext = SSE*j(n, 1, 1)  - inv(I - diag(H))*(residual##2);
/*compute s*/
s_ext = (SSE_ext/(n - k - 2))##(0.5);
/*compute esternal studentized residuals*/
residual_ext = (inv(I - diag(H))##(0.5))*(residual # (s_ext##(-1)));


/*comopute PRESS(prediction sum of square)*/
PRESS = t(residual_del)*residual_del;
print PRESS;

/*compute leverage*/
leverage = diag(H)*j(n, 1, 1);
high_leverage = 2*(k + 1)/n;

/*compute cook distance*/
A = inv(I - diag(H))*diag(H);
D = A*(residual_stu##2)/(k + 1);

/*creating obervation number*/
obs= t(1:n);

/*creating table*/
table = obs||y||y_hat||residual||leverage||residual_stu||residual_ext||D;
/*create labels for our table*/
cTable = {"obs" "response" "fitted" "residual" "leverage(h_ii)" "r_i"
        "t_i" "Cook"};
mattrib table colname=cTable;
print table high_leverage PRESS SSE;

quit;
```

also, as suggested by Hoaglin and Welsch(1978), the high leverage point is

$$\frac{2(k+1)}{n} = \frac{2 \times (3+1)}{34} = 0.2352941$$

We give the SAS output as a table similar to Table 9.1 as the example from the book:

table

	obs	response	fitted	residual	leverage(h_ii)	r_i	t_i	Cook
ROW1	1	18.38	17.331725	1.0482747	0.0779182	0.1323872	0.1302001	0.0003703
ROW2	2	20	23.947706	-3.947706	0.0625186	-0.494446	-0.488129	0.0040759
ROW3	3	11.5	13.85461	-2.35461	0.1408004	-0.308054	-0.303357	0.0038878
ROW4	4	25	26.241849	-1.241849	0.0699208	-0.156158	-0.153596	0.0004583
ROW5	5	52.5	68.180463	-15.68046	0.1858282	-2.107446	-2.244844	0.2534243
ROW6	6	82.5	66.430884	16.069116	0.0833024	2.0353282	2.1554637	0.094111
ROW7	7	25	32.919595	-7.919595	0.066605	-0.99409	-0.993888	0.0176292
ROW8	8	30.67	32.641876	-1.971876	0.067617	-0.24765	-0.243737	0.0011119
ROW9	9	12	7.7147424	4.2852576	0.1871236	0.576395	0.5698713	0.0191198
ROW10	10	61.25	57.480852	3.7691476	0.1028701	0.4825819	0.4763231	0.006676
ROW11	11	60	50.208153	9.7918466	0.0580871	1.2235304	1.2341521	0.0230802
ROW12	12	57.5	68.845878	-11.34588	0.1002079	-1.450516	-1.47894	0.0585794
ROW13	13	31	31.767901	-0.767901	0.0759472	-0.096875	-0.095262	0.0001928
ROW14	14	60	61.863684	-1.863684	0.0666054	-0.233935	-0.230213	0.0009763
ROW15	15	72.5	66.772835	5.7271652	0.1089489	0.7357731	0.7300231	0.0165481
ROW16	16	60.33	66.701899	-6.371899	0.1679011	-0.847104	-0.84301	0.0361987
ROW17	17	49.75	59.663209	-9.913209	0.1145003	-1.277543	-1.291698	0.0527606
ROW18	18	8.5	10.789921	-2.289921	0.192343	-0.309003	-0.304294	0.0056848
ROW19	19	36.5	24.642785	11.857215	0.06763	1.4891693	1.5214525	0.0402142
ROW20	20	60	65.605976	-5.605976	0.1807147	-0.751085	-0.745504	0.0311083
ROW21	21	16.25	18.015727	-1.765727	0.5046366	-0.304241	-0.299589	0.0235738
ROW22	22	50	47.423596	2.5764041	0.0346163	0.3179944	0.3131779	0.0009065
ROW23	23	11.5	16.365695	-4.865695	0.1181674	-0.628358	-0.621903	0.0132271
ROW24	24	35	38.577351	-3.577351	0.0636324	-0.448326	-0.442275	0.0034148

table

	obs	response	fitted	residual	leverage(h_ii)	r_i	t_i	Cook
ROW25	25	75	61.693762	13.306238	0.0627637	1.6668106	1.7203915	0.0465127
ROW26	26	31.56	35.257	-3.697	0.0351124	-0.456422	-0.450317	0.0018952
ROW27	27	48.5	42.200485	6.2995146	0.0629836	0.7892035	0.7841211	0.0104664
ROW28	28	77.5	69.888994	7.6110061	0.2414672	1.0597671	1.0620236	0.0893809
ROW29	29	21.67	22.063345	-0.393345	0.0601332	-0.049204	-0.048378	0.0000387
ROW30	30	19.75	21.220979	-1.470979	0.0963542	-0.187656	-0.18461	0.0009387
ROW31	31	56	48.173872	7.8261275	0.0511969	0.9743487	0.973499	0.0128067
ROW32	32	25	41.300371	-16.30037	0.2172292	-2.23427	-2.405996	0.346334
ROW33	33	40	26.907405	13.092595	0.2142543	1.7911841	1.863543	0.2187096
ROW34	34	56.67	56.584871	0.0851295	0.0600625	0.0106484	0.0104695	1.8114E-6

high_leverage	PRESS	SSE
0.2352941	2751.1794	2039.9062

The observation number here are what we manually added, just make it easier to point out which observation we are talking about. But it is not the real observation index since it is not given in the data.

As we can see that observation 21 *and* 28 *has a larger leverage than the suggested high leverage point. Also observation* 5, 32 *and* 33 *have relatively large leverage as well. From leverage aspect, these points can be potentially influential to the model.*

Among these points, we can see that observation 5, 32, 33 *also have relatively large Cook's distance, relatively large studentized residual(r_i) and relatively large studentized external residuals(t_i),* *their absolute values either larger than or close to* 2*. So we believe than they are potentially very influential to the model.*

Under the current model we have PRESS value as 2751.1794*. If we decide to fit other models by deleting a few observations, we can also compare different PRESS values (the model with smaller PRESS value may be preferred).*

Question 4.

Solution 4. *For part* (*i*)*:*

To verify $\mathbf{G}$ *is the generalized inverse of* $\mathbf{X}'\mathbf{X}$*, we just need to check by definition that we have:*

$$\mathbf{X}'\mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$$

The following code compute both left hand side and right hand side and print them out, and we will see from the output that they are equal.

```
/*Question 4*/

proc iml;

/*create design matrix X*/

X = {1 1 0 0 , 1 1 0 0 ,
     1 0 1 0 , 1 0 1 0 ,
     1 0 1 0 , 1 0 1 0 ,
     1 0 0 1,  1 0 0 1};

/*create matrix G*/
G = {0 0 0 0 , 0 0.5 0 0,
     0 0 0.25 0 , 0 0 0 0.5};
print G;

/*Verify G is a generalized inverse of X'X*/
LHS = (t(X)*X)*G*(t(X)*X);
RHS = t(X)*X;
print LHS RHS;
```

```
LHS                                 RHS

8     2     4     2                 8     2     4     2
2     2     0     0                 2     2     0     0
4     0     4     0                 4     0     4     0
2     0     0     2                 2     0     0     2
```

So $\mathbf{G}$ *is the generalized inverse of* $\mathbf{X}'\mathbf{X}$*.*

For part (*ii*)*:*

A solution to the equation

$$(\mathbf{X}'\mathbf{X})\mathbf{b} = \mathbf{X}'\mathbf{Y}$$

assumes the form of

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{(-)}\mathbf{X}'\mathbf{Y}$$

From part (a) we have checked that $\mathbf{G} = (\mathbf{X}'\mathbf{X})^{(-)}$*, So we just need to plug it into the equation above to get a version of the solution.*

The following code compute $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{(-)}\mathbf{X}' = \mathbf{G}\mathbf{X}'$ *and print it out:*

```
/*get a solution for the MLE equation*/
A = G*t(X);
print A;
```

A

0	0	0	0	0	0	0	0
0.5	0.5	0	0	0	0	0	0
0	0	0.25	0.25	0.25	0.25	0	0
0	0	0	0	0	0	0.5	0.5

So a solution would be:

$$\hat{\mathbf{b}} = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \end{bmatrix} = \mathbf{AY} = \begin{bmatrix} 0 \\ \frac{1}{2}Y_{11} + \frac{1}{2}Y_{12} \\ \frac{1}{4}(Y_{21} + Y_{22} + Y_{23} + Y_{24}) \\ \frac{1}{2}(Y_{31} + Y_{32}) \end{bmatrix}$$

For part (iii):

Let's form a matrix from $\mathbf{c}_1^T, \mathbf{c}_2^T, \mathbf{c}_3^T$*:*

$$\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

We know for all matrices its row rank is equal to its column rank. If we look at the rows for matrix $\mathbf{C}$*, the 2nd, 3rd and 4th rows are the natural basis of* $\mathbb{R}^3$*, and hence independent. Also, the 2rd, 3rd, and 4th rows sum up equal to the first row, so the row rank of* $\mathbf{C}$ *is* 3*, then the column rank of* $\mathbf{C}$ *is also* 3*, hence* $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3$ *are independent, so are their transpose.*

For (iv):

We have already proved in part (iii) that $\mathbf{c}_1^T, \mathbf{c}_2^T$ *and* $\mathbf{c}_3^T$ *are linearly independent. Now we just need to show that every single row in* $\mathbf{X}$ *can be linearly spanned from* $\mathbf{c}_1^T, \mathbf{c}_2^T$ *and* $\mathbf{c}_3^T$*.*

$$\begin{aligned}
\textit{row } 1 \ \textit{and row } 2 & \\
& (1,1,0,0) = c_1^T \\
& \textit{row } 3-6 \\
& (1,0,1,0) = c_2^T \\
& \textit{row } 7-8 \\
& (1,0,0,1) = c_3^T
\end{aligned}$$

So the vector space spanned by the rows of $\mathbf{X}$ *are essentially the space spanned by* $\mathbf{c}_1, \mathbf{c}_2$ *and* $\mathbf{c}_3$*.*

For part (v)*:vskip 2mm To show that* $\mathbf{c}_1^T\beta, \mathbf{c}_2^T\beta$ *and* $\mathbf{c}_3^T\beta$ *are estimable functions of* β*, we only need to show that* $\mathbf{c}_1^T, \mathbf{c}_2^T$ *and* $\mathbf{c}_3^T$ *are in the row space of* $\mathbf{X}$*, thanks to theorem* $12.2(b)$*. This is obvious, because* $\mathbf{c}_1^T$ *is just the first and second row of* $\mathbf{X}$*,* $\mathbf{c}_2^T$ *is the same as the third to sixth row of* $\mathbf{X}$ *,and* $\mathbf{c}_3^T$ *is the last two rows of* $\mathbf{X}$*, so all are in the linear space expanded by the rows of* $\mathbf{X}$*, and hence* $\mathbf{c}_1^T\beta, \mathbf{c}_2^T\beta$ *and* $\mathbf{c}_3^T\beta$ *are estimable.*

For part (vi)*:*

By Theorem 12.2c, we know that the number of linearly independent estimable function of β *is the rank of* $\mathbf{X}$*, which in this case the rank is* 3*. and as we can see that* $\mathbf{c}_1^T, \mathbf{c}_2^T$ *and* $\mathbf{c}_3^T$ *are three independent vectors, so* $\mathbf{c}_1^T\beta, \mathbf{c}_2^T\beta$ *and* $\mathbf{c}_3^T\beta$ *already form the maximum number of linearly independent estimable functions of* β*, and any other estimable linear functions of* β *should be their linear combination (remember in part* (iv) *we proved that* $\mathbf{c}_1^T, \mathbf{c}_2^T$ *and* $\mathbf{c}_3^T$ *form basis of the row space of* $\mathbf{X}$*). Thus finished the proof.*

For part (vii)*:*

We have a design matrix $\mathbf{X}$ *whose dimension is* $n \times p$ *with* $n = 8, p = 4$ *but with rank* $k = 3 < p < n$*. We can write our model as a one-way anova model as:*

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

with $1 \le i \le 3$*, and when* $i = 1, 1 \le j \le 2$ *when* $i = 2, 1 \le j \le 4$ *and when* $i = 3, 1 \le j \le 2$*.*

It is over-parametrized as we have proved in Theorem $12.2a$ *that we could not get unique estimate for each single parameter* $\mu, \alpha_1, \alpha_2, \alpha_3$

A way to reparametrize this is to write $\mu + \alpha_i = \alpha_i^*$ *as our new parameter, so our model becomes*

$$y_{ij} = \alpha_i^* + \epsilon_{ij}$$

and the design matrix becomes

$$\mathbf{X}^* = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

which is now full rank.

Then we can obtain the estimate for our parameter as:

$$\begin{bmatrix} \hat{\alpha}_1^* \\ \hat{\alpha}_2^* \\ \hat{\alpha}_3^* \end{bmatrix} = \begin{bmatrix} \hat{\mu}+\hat{\alpha}_1 \\ \hat{\mu}+\hat{\alpha}_2 \\ \hat{\mu}+\hat{\alpha}_3 \end{bmatrix} = \Big((\mathbf{X}^*)'\mathbf{X}^*\Big)^{-1}(\mathbf{X}^*)'\mathbf{y}$$

$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & & & & & & \\ & & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & & \\ & & & & & & \frac{1}{2} & \frac{1}{2} \end{bmatrix}\mathbf{y}$$

$$= \begin{bmatrix} \frac{1}{2}Y_{11}+\frac{1}{2}Y_{12} \\ \frac{1}{4}(Y_{21}+Y_{22}+Y_{23}+Y_{24}) \\ \frac{1}{2}(Y_{31}+Y_{32}) \end{bmatrix}$$

For (viii):

Interesting enough, we can simply choose the side condition $\mu = 0$, then we will get the same estimate as in part (ii) and part (vii).

It is not hard to see the relationship between the estimate of part (ii) and part (vii). If we add a side condition $\mu = 0$ to part (vii), since part (vii) already estimated $\mu+\alpha_1, \mu+\alpha_2$ and $\mu+\alpha_3$, now we can plug in $\mu = 0$, which gives us the estimate in part (ii). But as we know that the choice of side condition is not unique, so this is just one version of the estimate for $(\mu, \alpha_1, \alpha_2, \alpha_3)$.

To show the choice of side condition is non-estimable function of β, we are looking at all those $\lambda' = (0, k_1, k_2, k_3)$ since our side condition is $\mu = 0$. Apparently for some choice of k_1, k_2 and k_3, the vector $(0, k_1, k_2, k_3)$ can not be spanned from the rows of $\mathbf{X}$. For example, take $\lambda' = (0, 1, 0, 0)$, or $\lambda' = (0, 0, 1, 0)$ or $\lambda' = (0, 0, 0, 1)$.

Question 5.

Solution 5. *Given the model, our design matrix and response are:*

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 120 \\ 140 \\ 135 \\ 110 \\ 160 \\ 164 \\ 155 \\ 175 \\ 170 \\ 165 \end{bmatrix}$$

To check for estimability, we just need to see if the following functions of β has coefficient vector that can be spanned from the row space of X.

For (i):

$\mu = (1, 0, 0)\beta$, it is not in the row space of $\mathbf{X}$ and hence not estimable.

For (ii):

$\mu+\alpha_1 = \lambda'\beta = (1,1,0)\beta$, *it is in the row space of* $\mathbf{X}$ *and hence is estimable. In this case, we have* $\lambda' = (1,1,0) = \mathbf{a}'\mathbf{X}$. *Since* $(1,1,0)$ *is just the first row of* $\mathbf{X}$, *so we have* $\mathbf{a}' = (1,0,0,0,0,0,0,0,0,0)$.

Since $(1,1,0)$ *is also the same vector for some other rows of* X, *so we can have different choice of* $\mathbf{a}$. *For example, if we see* $(1,1,0)$ *as the 3rd row of* $\mathbf{X}$, *then we can have* $\mathbf{a}' = (0,0,1,0,0,0,0,0,0,0)$

For (iii):

$$\alpha_1 - \alpha_2 = \lambda'\beta = (0,1,-1)\beta$$

Since $(0,1,-1)$ *can be viewed as the difference between row* 1 *and row* 2 *of* $\mathbf{X}$, *it is in the row space, andhence* $\alpha_1 - \alpha_2$ *is estimable.*

For $(0,1,-1) = \mathbf{a}'\mathbf{X}$, *we have* $\mathbf{a}' = (1,-1,0,0,0,0,0,0,0,0)$

For (iv):

$$\alpha_1 + \alpha_2 = \lambda'\beta = (0,1,1)\beta$$

however it is impossible to get 1 *for both second and third element and cancel out the first one, hence* $(0,1,1)$ *not in the row space of* $\mathbf{X}$, *and* $\alpha_1 + \alpha_2$ *is not estimable.*

Question 6.

Solution 6. *I confirm that I have studied the solutions of these problems.*