Question #1.

Solution 1. *For part [A]:*

For (i):

We want to prove equation:

$$|\mathbf{S}| = (s_{11}s_{22}\cdots s_{pp})|\mathbf{R}|$$

where

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{12} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1p} & s_{2p} & \cdots & s_{pp} \end{bmatrix}$$

with

$$s_{ik} = \frac{1}{n}\sum_{j=1}^{n}(x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)$$

and

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{12} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 1 & \frac{s_{12}}{\sqrt{s_{11}s_{22}}} & \cdots & \frac{s_{1p}}{\sqrt{s_{11}s_{pp}}} \\ \frac{s_{12}}{\sqrt{s_{11}s_{22}}} & 1 & \cdots & \frac{s_{2p}}{\sqrt{s_{22}s_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{s_{1p}}{\sqrt{s_{11}s_{pp}}} & \frac{s_{2p}}{\sqrt{s_{22}s_{pp}}} & \cdots & 1 \end{bmatrix}$$

with

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}s_{kk}}} \qquad r_{ii} = \frac{s_{ii}}{\sqrt{s_{ii}s_{ii}}} = 1$$

We prove this by using the Lebniz formula (also called Laplace formula) for determinant.

First of all, we denote by S_p the set of permuations of integer set $\{1, 2, \ldots, p\}$. (we did not use bold font for S_p, so it should not be confused with the sample variance matrix $\mathbf{S}$).

So an element $\sigma \in S_p$ represents a particular permutation for $\{1, 2, \ldots, p\}$.

for example, if σ make the following permutation:

$$\{1, 2, \ldots, p\} \xrightarrow{\sigma} \{2, 1, 3, 4, \ldots, p\}$$

then we denote $\sigma_1 = 2, \sigma_2 = 1, \sigma_3 = 3, \sigma_4 = 4, \ldots, \sigma_p = p$.

Also, we use sgn(σ) to denote the signature of σ, a value that is $+1$ whenever the permutation given by σ can be achieved by successively interchanging two entries an even number of times, and -1 whenever it can be achieved by an odd number of such interchanges.

Then the Leibniz formula (or Laplace formula) says:

$$|\mathbf{S}| = \Sigma_{\sigma \in S_p}\Big(sgn(\sigma)\prod_{i=1}^{p} s_{i,\sigma_i}\Big)$$

and

$$|\mathbf{R}| = \Sigma_{\sigma \in S_p}\Big(sgn(\sigma)\prod_{i=1}^{p} r_{i,\sigma_i}\Big)$$

Notice that for each $\sigma \in S_p$, we have the following relationship:

$$r_{i,\sigma_i} = \frac{s_{i,\sigma_i}}{\sqrt{s_{ii}s_{\sigma_i,\sigma_i}}} \text{ or } s_{i,\sigma_i} = r_{i,\sigma_i} \cdot (\sqrt{s_{ii}s_{\sigma_i,\sigma_i}})$$

So

$$\prod_{i=1}^{p} s_{i,\sigma_i} = \prod_{i=1}^{p}\Big[r_{i,\sigma_i} \cdot (\sqrt{s_{ii}s_{\sigma_i,\sigma_i}})\Big] = \Big(\prod_{i=1}^{p}\sqrt{s_{ii}}\Big)\Big(\prod_{i=1}^{p}\sqrt{s_{\sigma_i,\sigma_i}}\Big)\Big(\prod_{i=1}^{p} r_{i,\sigma_i}\Big)$$

Notice that since σ is a permutation of $\{1, 2, \ldots, p\}$, so

$$\prod_{i=1}^{p}\sqrt{s_{\sigma_i,\sigma_i}} = \prod_{i=1}^{p}\sqrt{s_{ii}}$$

So

$$\begin{aligned}
\prod_{i=1}^{p} s_{i,\sigma_i} &= \prod_{i=1}^{p}\Big[r_{i,\sigma_i} \cdot (\sqrt{s_{ii}s_{\sigma_i,\sigma_i}})\Big] \\
&= \Big(\prod_{i=1}^{p}\sqrt{s_{ii}}\Big)\Big(\prod_{i=1}^{p}\sqrt{s_{ii}}\Big)\Big(\prod_{i=1}^{p} r_{i,\sigma_i}\Big) \\
&= (s_{11}s_{22}\cdots s_{pp})\prod_{i=1}^{p} r_{i,\sigma_i}
\end{aligned}$$

Since this relationship holds true for any permutation $\sigma \in S_p$, thus we have:

$$\begin{aligned}
|\mathbf{S}| &= \Sigma_{\sigma \in S_p}\Big(sgn(\sigma)\prod_{i=1}^{p} s_{i,\sigma_i}\Big) \\
&= \Sigma_{\sigma \in S_p}\Big(sgn(\sigma)s_{11}s_{22}\cdots s_{pp}\prod_{i=1}^{p} r_{i,\sigma_i}\Big) \\
&= (s_{11}s_{22}\ldots s_{pp})\Sigma_{\sigma \in S_p}\Big(sgn(\sigma)\prod_{i=1}^{p} r_{i,\sigma_i}\Big) \\
&= (s_{11}s_{22}\ldots s_{pp})|\mathbf{R}|
\end{aligned}$$

Thus completed the proof of equation (3.21).

As for part (*ii*)

we know that for any 2×2 *matrix, say* $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2]$ *as in our question (so* $\mathbf{U}$ *is a* 2×2 *matrix whose first column is vector* $\mathbf{u}_1$ *and second column is vector* $\mathbf{u}_2$*), the area generated by* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *is the absolute value of the determinant of* $\mathbf{U}$*, which is* $\Big|det(\mathbf{U})\Big|$.

Similarly, given $\mathbf{v}_1 = \mathbf{A}\mathbf{u}_1$ *and* $\mathbf{v}_2 = \mathbf{A}\mathbf{u}_2$*, we have:*

$$\begin{aligned}
Area(\mathbf{v}_1, \mathbf{v}_2) &= \Big|det(\Big[\mathbf{v}_1, \mathbf{v}_2\Big])\Big| = \Big|det(\Big[\mathbf{A}\mathbf{u}_1, \mathbf{A}\mathbf{u}_2\Big])\Big| \\
&= \Big|det(\mathbf{A} \cdot \Big[\mathbf{u}_1, \mathbf{u}_2\Big])\Big| \\
&= \Big|det(A) \cdot det(\Big[\mathbf{u}_1, \mathbf{u}_2\Big])\Big| \\
&= \Big|det(A)\Big| \cdot \Big|det(\Big[\mathbf{u}_1, \mathbf{u}_2\Big])\Big| \\
&= \Big|det(A)\Big| \cdot Area(\mathbf{u}_1, \mathbf{u}_2)
\end{aligned}$$

Thus completed the proof for this part.

For part [*B*]*:*

For part (*i*)*: So we have a data with* $n = 5$ *observatios and* $p = 3$ *variables(covariates).*

Since $\mathbf{x}_1 = (3, 6, 4, 7, 5)'$, $\mathbf{x}_2 = (1, 4, 2, 0, 3)'$ *and* $\mathbf{x}_3 = (0, 6, 2, 3, 4)$*, we have the sample means:*

$$\begin{aligned}
\bar{\mathbf{x}}_1 &= \frac{3+6+4+7+5}{5} = 5 \\
\bar{\mathbf{x}}_2 &= \frac{1+4+2+0+3}{5} = 2 \\
\bar{\mathbf{x}}_3 &= \frac{0+6+2+3+4}{5} = 3
\end{aligned}$$

Hence we have deviation vectors:

$$\begin{aligned}
\mathbf{d}_1 &= \mathbf{x}_1 - \bar{\mathbf{x}}_1\mathbf{1} = (3, 6, 4, 7, 5)' - (5, 5, 5, 5, 5)' = (-2, 1, -1, 2, 0)' \\
\mathbf{d}_2 &= \mathbf{x}_2 - \bar{\mathbf{x}}_2\mathbf{1} = (1, 4, 2, 0, 3)' - (2, 2, 2, 2, 2)' = (-1, 2, 0, -2, 1)' \\
\mathbf{d}_3 &= \mathbf{x}_3 - \bar{\mathbf{x}}_3\mathbf{1} = (0, 6, 2, 3, 4)' - (3, 3, 3, 3, 3)' = (-3, 3, -1, 0, 1)'
\end{aligned}$$

So the deviation matrix is:

$$\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3] = \begin{bmatrix} -2 & -1 & -3 \\ 1 & 2 & 3 \\ -1 & 0 & -1 \\ 2 & -2 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

it is obvious to see that $\mathbf{d}_1 + \mathbf{d}_2 = \mathbf{d}_3$ *so with coefficient vector* $\mathbf{a}^T = [1, 1, -1]$ *we establish the dependence among* $\mathbf{d}_1, \mathbf{d}_2$ *and* $\mathbf{d}_3$.

For part (ii):

We are looking for sample covariance matrix $\mathbf{S}$ *and verify the generalized variance is* 0.

I would like to point out that, although does not affect the result for generalized variance, the book from Richard and Dean has not been consistent with the definition of $\mathbf{S}$. *On page* 118 *example* 3.4, *it is computing sample variance and covariance with formulas:*

$$s_{ii} = \frac{1}{n}\mathbf{d}_i'\mathbf{d}_i = \frac{1}{n}\sum_{j=1}^{n}(x_{ji} - \bar{\mathbf{x}}_i)^2$$
$$s_{ik} = \frac{1}{n}\mathbf{d}_i'\mathbf{d}_k = \frac{1}{n}\sum_{j=1}^{n}(x_{ji} - \bar{\mathbf{x}}_i)(x_{jk} - \bar{\mathbf{x}}_k)$$

with $n = 3$ *in the example.*

However on page 123 *section 3.4, it is computing* s_{ik} *as*

$$s_{ik} = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)$$

I would follow the second way here because it is consistent with the sample variance definition I am familiar with.

In our example, $n = 5$ *so* $n - 1 = 4$, *thus:*

$$\begin{aligned}
s_{11} &= \frac{1}{4}\mathbf{d}_1'\mathbf{d}_1 = \frac{1}{4}\Big((-2)^2 + 1^2 + (-1)^2 + 2^2\Big) = \frac{10}{4} = \frac{5}{2} \\
s_{12} &= \frac{1}{4}\mathbf{d}_1'\mathbf{d}_2 = \frac{1}{4}\Big((-2)\times(-1) + 1\times 2 + (-1)\times 0 + (2\times(-2)) + 0\times 1\Big) = 0 \\
s_{13} &= \frac{1}{4}\mathbf{d}_1'\mathbf{d}_3 = \frac{1}{4}\Big((-2)\times(-3) + (1\times 3) + (-1)\times(-1) + 2\times 0 + 0\times 1\Big) = \frac{5}{2} \\
s_{22} &= \frac{1}{4}\mathbf{d}_2'\mathbf{d}_2 = \frac{1}{4}\Big((-1)^2 + 2^2 + 0^2 + (-2)^2 + 1^2\Big) = \frac{5}{2} \\
s_{23} &= \frac{1}{4}\mathbf{d}_2'\mathbf{d}_3 = \frac{1}{4}\Big((-1)\times(-3) + 2\times 3 + 0\times(-1) + (-2)\times 0 + 1\times 1\Big) = \frac{5}{2} \\
s_{33} &= \frac{1}{4}\mathbf{d}_3'\mathbf{d}_3 = \frac{1}{4}\Big((-3)^2 + 3^2 + (-1)^2 + 0^2 + 1^2\Big) = 5
\end{aligned}$$

So the sample covariance matrix is:

$$\mathbf{S} = \begin{bmatrix} \frac{5}{2} & 0 & \frac{5}{2} \\ 0 & \frac{5}{2} & \frac{5}{2} \\ \frac{5}{2} & \frac{5}{2} & 5 \end{bmatrix}$$

Notice that the for S, *the third column is the sum of first and second column, so* $\mathbf{S}$ *is not of full rank(or in other words,* $\mathbf{S}$ *is singular), and hence the generalized variance is:*

$$\Big|\mathbf{S}\Big| = 0$$

For part (iii)

To show that the columns of the data matrix are linearly independent, it suffice to show that the rank of the data matrix is 3*, and since we have three columns, that complete the proof.*

To show that, we can do Gauss elimination to get the echelon matrix, and we will see the echelon matrix has rank 3*.*

$$\begin{bmatrix} 3 & 1 & 0 \\ 6 & 4 & 6 \\ 4 & 2 & 2 \\ 7 & 0 & 3 \\ 5 & 3 & 4 \end{bmatrix} \Longrightarrow \begin{bmatrix} 3 & 1 & 0 \\ 0 & 2 & 6 \\ 0 & \frac{2}{3} & 2 \\ 0 & -\frac{7}{3} & 3 \\ 0 & \frac{4}{3} & 4 \end{bmatrix} \Longrightarrow \begin{bmatrix} 3 & 1 & 0 \\ 0 & 1 & 3 \\ 0 & \frac{2}{3} & 2 \\ 0 & -\frac{7}{3} & 3 \\ 0 & \frac{4}{3} & 4 \end{bmatrix} \Longrightarrow \begin{bmatrix} 3 & 1 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 10 \\ 0 & 0 & 0 \end{bmatrix}$$

The elimination above is all standard row elimination process, and as we can see there are three rows left so the rank of data matrix is 3*, and hence the three columns are linearly independent.*

Question #2.

Solution 2. *For part* $[A]$*:*

For part (i)*:*

We want to prove the following equation by using Theorem 2.9*c.*

$$\begin{vmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{vmatrix} = |\mathbf{A}_{11}| \cdot |\mathbf{A}_{22}|$$

Notice that

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}$$

So

$$\left| \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \right| = \left| \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \right| \cdot \left| \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \right|$$

We just need to show that

$$\left| \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \right| = |\mathbf{A}_{11}| \qquad \textit{and} \qquad \left| \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \right| = |\mathbf{A}_{22}|$$

We could not directly claim the above is true without using Theorem 2.9b. The question here requires only using Theorem 2.9c, so that needs justification.

It would suffice to show the first one because the proof of second one is very similar.

Assume that $\mathbf{A}_{11}$ *is an* $r \times r$ *matrix, and denote by* σ *the permutation of* $\{1, 2, \ldots, n\}$ *and* S_n *the set of permutations, also denote sgn*(σ) *the signature of* σ *as we did in part* $[A]$

Also, in the proof below we abuse the notation a little bit and use a_{ij} *to denote the* (i, j) *entry of*

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

and specifically, when $1 \leq i \leq r$ and $1 \leq j \leq r$, a_{ij} is the (i,j) entry of $\mathbf{A}_{11}$. Then we have by Leibniz formula:

$$\left| \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \right| = \sum_{\sigma \in S_n} \left(sgn(\sigma) \prod_{i=1}^{n} a_{i,\sigma_i} \right)$$

However since the lower right corner is identity matrix $\mathbf{I}$, so $\prod_{i=1}^{n} a_{i,\sigma_i} \neq 0$ if and only if $a_{i,\sigma_i} = 1$ for $i = r+1, r+2, \ldots, n$, or equivalently, if and only if $\sigma_i = i$ for $i = r+1, r+2, \ldots, n$. Thus σ is really just a permutation of the first r integers $\{1, 2, \ldots, r\}$, so $\sigma \in S_r$, and we have:

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \sum_{\sigma \in S_r} \left(sgn(\sigma) \prod_{i=1}^{r} a_{i,\sigma_i} \right) = |\mathbf{A}_{11}|$$

Similar proof follows for the other matrix. Thus we have proved that

$$\begin{vmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{vmatrix} = |\mathbf{A}_{11}| \cdot |\mathbf{A}_{22}|$$

For part [B]:

Notice that:

$$\begin{aligned}
&\begin{bmatrix} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0}^T & \mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}
\end{aligned}$$

By the results from both [A] and [B], we have:

$$\begin{aligned}
&\left| \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0}^T & \mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{bmatrix} \right| \\
&= \left| \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0}^T & \mathbf{I} \end{bmatrix} \right| \cdot \left| \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \right| \cdot \left| \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{bmatrix} \right| \\
&= 1 \cdot \left| \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \right| \cdot 1 \\
&= \left| \begin{bmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \right| \\
&= |\mathbf{A}_{22}| \cdot |\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}|
\end{aligned}$$

Similarly, notice that:

$$\begin{aligned}
&\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix}
\end{aligned}$$

So we have:

$$
\begin{aligned}
&\left|\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\right| \\
&= \left|\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix}\right| \cdot \left|\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}\right| \cdot \left|\begin{bmatrix} \mathbf{I} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\right| \\
&= 1 \cdot \left|\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}\right| \cdot 1 \\
&= \left|\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{bmatrix}\right| \\
&= |\mathbf{A}_{11}| \cdot |\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}|
\end{aligned}
$$

Thus completed the proof of equation of both (2.71) *and* (2.72).

For part [C]*:*

We want to show that:

$$
\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{bmatrix} \begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{A}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0}^T & \mathbf{I} \end{bmatrix}
$$

We just need to show that the right hand side multiply by $\mathbf{A}$ *gives identity matrix.*

We have:

$$
\begin{aligned}
&\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{bmatrix} \begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{A}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0}^T & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \\
&= \left(\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{bmatrix} \begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & \mathbf{0} \\ \mathbf{0}^T & \mathbf{A}_{22}^{-1} \end{bmatrix}\right)\left(\begin{bmatrix} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0}^T & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}\right) \\
&= \begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & \mathbf{A}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{0} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}
\end{aligned}
$$

thus completed the proof.

Question #3.

Solution 3. *For part* [A]*:*

with

$$
\mu = \begin{bmatrix} 3 \\ 4 \\ -1 \\ -2 \end{bmatrix} \qquad \textit{and } \Sigma = \begin{bmatrix} 3 & -4 & 2 & 1 \\ -4 & 4 & -3 & -1 \\ 2 & -3 & 2 & 6 \\ 1 & -1 & 6 & 5 \end{bmatrix}
$$

We work on the following problems:

part (a):

The joint marginal distribution of y_1 and y_3:

It is straightforward to see that:

$$\begin{bmatrix} y_1 \\ y_3 \end{bmatrix} \text{ is } N_2\left(\begin{bmatrix} 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix}\right)$$

part (b):

The marginal distribution of y_2:

The marginal distribution of y_2 can simply be read from the joint distribution of $\mathbf{y}$, and y_2 is $N(4,4)$

part (c):

The distribution of $z = y_1 + 2y_2 - y_3 + 3y_4$

Notice that $z = \mathbf{a}'\mathbf{y}$ where $\mathbf{a} = (1,2,-1,3)'$.

So we know that:

$$z \text{ is } N(\mathbf{a}'\mu, \mathbf{a}'\Sigma\mathbf{a})$$

In this case, we have:

$$\mathbf{a}'\mu = (1,2,-1,3)\begin{bmatrix} 3 \\ 4 \\ -1 \\ -2 \end{bmatrix} = 6$$

and

$$\begin{aligned}\mathbf{a}'\Sigma\mathbf{a} &= (1,2,-1,3)\begin{bmatrix} 3 & -4 & 2 & 1 \\ -4 & 4 & -3 & -1 \\ 2 & -3 & 2 & 6 \\ 1 & -1 & 6 & 5 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ -1 \\ 3 \end{bmatrix} \\ &= (-4,4,12,8)\begin{bmatrix} 1 \\ 2 \\ -1 \\ 3 \end{bmatrix} = 16\end{aligned}$$

So z is $N(6,16)$.

part (d):

The joint distribution of $z_1 = y_1 + y_2 - y_3 - y_4$ and $z_2 = -3y_1 + y_2 + 2y_3 - 2y_4$:

So we have:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ -3 & 1 & 2 & -2 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

If we denote:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ -3 & 1 & 2 & -2 \end{bmatrix}$$

Then

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \textit{ is } N_2(\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}')$$

We have:

$$\mathbf{A}\mu = \begin{bmatrix} 1 & 1 & -1 & -1 \\ -3 & 1 & 2 & -2 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ -1 \\ -2 \end{bmatrix} = \begin{bmatrix} 10 \\ -3 \end{bmatrix}$$

and

$$\begin{aligned} \mathbf{A}\Sigma\mathbf{A}' &= \begin{bmatrix} 1 & 1 & -1 & -1 \\ -3 & 1 & 2 & -2 \end{bmatrix} \begin{bmatrix} 3 & -4 & 2 & 1 \\ -4 & 4 & -3 & -1 \\ 2 & -3 & 2 & 6 \\ 1 & -1 & 6 & 5 \end{bmatrix} \begin{bmatrix} 1 & -3 \\ 1 & 1 \\ -1 & 2 \\ -1 & -2 \end{bmatrix} \\ &= \begin{bmatrix} -4 & 4 & -9 & -11 \\ -11 & 12 & -17 & -2 \end{bmatrix} \begin{bmatrix} 1 & -3 \\ 1 & 1 \\ -1 & 2 \\ -1 & -2 \end{bmatrix} \\ &= \begin{bmatrix} 20 & 20 \\ 20 & 15 \end{bmatrix} \end{aligned}$$

So we have:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \textit{ is } N_2(\begin{bmatrix} 10 \\ -3 \end{bmatrix}, \begin{bmatrix} 20 & 20 \\ 20 & 15 \end{bmatrix})$$

part (e):

we are looking for $f(y_1, y_2 | y_3, y_4)$

Let's denote by:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} y_3 \\ y_4 \end{bmatrix}$$

Then we have:

$$\mu_y = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$
$$\mu_x = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$
$$\mathbf{\Sigma}_{yy} = \begin{bmatrix} 3 & -4 \\ -4 & 4 \end{bmatrix}$$
$$\mathbf{\Sigma}_{yx} = \begin{bmatrix} 2 & 1 \\ -3 & -1 \end{bmatrix}$$
$$\mathbf{\Sigma}_{xy} = \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix}$$
$$\mathbf{\Sigma}_{xx} = \begin{bmatrix} 2 & 6 \\ 6 & 5 \end{bmatrix}$$

So $f(y_1, y_2|y_3, y_4)$ *is multivariate normal with mean vector and covariance matrix:*

$$\begin{aligned}
E[\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} | \begin{bmatrix} y_3 \\ y_4 \end{bmatrix}] &= \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \mu_x) \\
&= \begin{bmatrix} 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ -3 & -1 \end{bmatrix} \begin{bmatrix} -\frac{5}{26} & \frac{3}{13} \\ \frac{3}{13} & -\frac{1}{13} \end{bmatrix} (\begin{bmatrix} y_3 \\ y_4 \end{bmatrix} - \begin{bmatrix} -1 \\ -2 \end{bmatrix}) \\
&= \begin{bmatrix} 3 \\ 4 \end{bmatrix} + \begin{bmatrix} -\frac{2}{13} & \frac{5}{13} \\ \frac{9}{26} & -\frac{8}{13} \end{bmatrix} (\begin{bmatrix} y_3 \\ y_4 \end{bmatrix} - \begin{bmatrix} -1 \\ -2 \end{bmatrix}) \\
&= \begin{bmatrix} -\frac{2}{13}y_3 + \frac{5}{13}y_4 + \frac{47}{13} \\ \frac{9}{26}y_3 - \frac{8}{13}y_4 + \frac{81}{26} \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
cov(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} | \begin{bmatrix} y_3 \\ y_4 \end{bmatrix}) &= \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \\
&= \begin{bmatrix} 3 & -4 \\ -4 & 4 \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ -3 & -1 \end{bmatrix} \begin{bmatrix} 2 & 6 \\ 6 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} \\
&= \begin{bmatrix} 3 & -4 \\ -4 & 4 \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ -3 & -1 \end{bmatrix} \begin{bmatrix} -\frac{5}{26} & \frac{3}{13} \\ \frac{3}{13} & -\frac{1}{13} \end{bmatrix} \begin{bmatrix} 2 & -3 \\ 1 & -1 \end{bmatrix} \\
&= \begin{bmatrix} 3 & -4 \\ -4 & 4 \end{bmatrix} - \begin{bmatrix} \frac{1}{13} & \frac{1}{13} \\ \frac{1}{13} & -\frac{11}{26} \end{bmatrix} \\
&= \begin{bmatrix} \frac{38}{13} & -\frac{53}{13} \\ -\frac{53}{13} & \frac{115}{26} \end{bmatrix} \\
&= \begin{bmatrix} 2.923077 & -4.076923 \\ -4.076923 & 4.423077 \end{bmatrix}
\end{aligned}$$

part (f):

we are looking for $f(y_1, y_3|y_2, y_4)$

Let's denote by

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_3 \end{bmatrix}$$
$$\mathbf{x} = \begin{bmatrix} y_2 \\ y_4 \end{bmatrix}$$

Then we have:

$$\mu_y = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$
$$\mu_x = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$$
$$\mathbf{\Sigma}_{yy} = \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix}$$
$$\mathbf{\Sigma}_{yx} = \begin{bmatrix} -4 & 1 \\ -3 & 6 \end{bmatrix}$$
$$\mathbf{\Sigma}_{xy} = \begin{bmatrix} -4 & -3 \\ 1 & 6 \end{bmatrix}$$
$$\mathbf{\Sigma}_{xx} = \begin{bmatrix} 4 & -1 \\ -1 & 5 \end{bmatrix}$$

So $f(y_1, y_3|y_2, y_4)$ is multivariate normal with mean vector and covariance matrix:

$$\begin{aligned}
E[\begin{bmatrix} y_1 \\ y_3 \end{bmatrix} | \begin{bmatrix} y_2 \\ y_4 \end{bmatrix}] &= \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \mu_x) \\
&= \begin{bmatrix} 3 \\ -1 \end{bmatrix} + \begin{bmatrix} -4 & 1 \\ -3 & 6 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ -1 & 5 \end{bmatrix}^{-1} (\begin{bmatrix} y_2 \\ y_4 \end{bmatrix} - \begin{bmatrix} 4 \\ -2 \end{bmatrix}) \\
&= \begin{bmatrix} 3 \\ -1 \end{bmatrix} + \begin{bmatrix} -4 & 1 \\ -3 & 6 \end{bmatrix} \begin{bmatrix} \frac{5}{19} & \frac{1}{19} \\ \frac{1}{19} & \frac{4}{19} \end{bmatrix} (\begin{bmatrix} y_2 \\ y_4 \end{bmatrix} - \begin{bmatrix} 4 \\ -2 \end{bmatrix}) \\
&= \begin{bmatrix} 3 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 & 0 \\ -\frac{9}{19} & \frac{21}{19} \end{bmatrix} (\begin{bmatrix} y_2 \\ y_4 \end{bmatrix} - \begin{bmatrix} 4 \\ -2 \end{bmatrix}) \\
&= \begin{bmatrix} -y_2 + 7 \\ -\frac{9}{19}y_2 + \frac{21}{19}y_4 + \frac{59}{19} \end{bmatrix}
\end{aligned}$$

$$
\begin{aligned}
cov(\begin{bmatrix} y_1 \\ y_3 \end{bmatrix} \mid \begin{bmatrix} y_2 \\ y_4 \end{bmatrix}) &= \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \\
&= \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix} - \begin{bmatrix} -4 & 1 \\ -3 & 6 \end{bmatrix} \begin{bmatrix} \frac{5}{19} & \frac{1}{19} \\ \frac{1}{19} & \frac{4}{19} \end{bmatrix} \begin{bmatrix} -4 & -3 \\ 1 & 6 \end{bmatrix} \\
&= \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix} - \begin{bmatrix} -1 & 0 \\ -\frac{9}{19} & \frac{21}{19} \end{bmatrix} \begin{bmatrix} -4 & -3 \\ 1 & 6 \end{bmatrix} \\
&= \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix} - \begin{bmatrix} 4 & 3 \\ 3 & \frac{153}{19} \end{bmatrix} \\
&= \begin{bmatrix} -1 & -1 \\ -1 & -\frac{115}{19} \end{bmatrix}
\end{aligned}
$$

I found it is a bit odd that there is negative value on the diagonal entries of conditional covariance matrix. At first I thought I made a calculation mistake somewhere, but I could not find any after double check. Then I checked the updated variance-covariance matrix given by the question and found that Σ has two negative eigenvalues -0.4613693 and -3.1477987 thus fail to be positive definite.

part (g):

We have:

$$
\begin{aligned}
s_{11} &= 3 \\
s_{33} &= 2 \\
s_{13} &= 2
\end{aligned}
$$

and hence

$$\rho_{13} = \frac{s_{13}}{\sqrt{s_{11}}\sqrt{s_{33}}} = \frac{2}{\sqrt{3}\sqrt{2}} = \frac{2}{\sqrt{6}} = 0.8164966$$

part (h):

we have from part (f) that:

$$s_{13\cdot 24} = -1 \qquad s_{11\cdot 24} = -1 \qquad s_{33\cdot 24} = \frac{-115}{19}$$

So

$$
\begin{aligned}
\rho_{13\cdot 24} &= \frac{s_{13\cdot 24}}{\sqrt{s_{11\cdot 24}}\sqrt{s_{33\cdot 24}}} \\
&= \frac{-1}{\sqrt{-1}\cdot\sqrt{-\frac{115}{19}}} \qquad \textit{Oops!}
\end{aligned}
$$

For part (i):

denote by:

$$\mathbf{y} = y_1, \qquad \mathbf{x} = \begin{bmatrix} y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

We have

$$\mu_y = 3, \qquad \mu_x = \begin{bmatrix} 4 \\ -1 \\ -2 \end{bmatrix}$$

and

$$\Sigma_{yy} = 3, \qquad \Sigma_{yx} = (-4, 2, 1) \qquad \Sigma_{xx} = \begin{bmatrix} 4 & -3 & -1 \\ -3 & 2 & 6 \\ -1 & 6 & 5 \end{bmatrix}$$

Thus $f(y_1|y_2, y_3, y_4) = f(y|\mathbf{x})$ *is multivariate normal with the following mean and covariance matrix:*

$$\begin{aligned}
E[\mathbf{y}|\mathbf{x}] &= \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x} - \mu_x) \\
&= 3 + (-4, 2, 1) \begin{bmatrix} 4 & -3 & -1 \\ -3 & 2 & 6 \\ -1 & 6 & 5 \end{bmatrix}^{-1} \left(\begin{bmatrix} y_2 \\ y_3 \\ y_4 \end{bmatrix} - \begin{bmatrix} 4 \\ -1 \\ -2 \end{bmatrix} \right) \\
&= 3 + (-4, 2, 1) \begin{bmatrix} \frac{26}{115} & -\frac{9}{115} & \frac{16}{115} \\ -\frac{9}{115} & -\frac{19}{115} & \frac{21}{115} \\ \frac{16}{115} & \frac{21}{115} & \frac{1}{115} \end{bmatrix} \left(\begin{bmatrix} y_2 \\ y_3 \\ y_4 \end{bmatrix} - \begin{bmatrix} 4 \\ -1 \\ -2 \end{bmatrix} \right) \\
&= 3 + (-\frac{106}{115}, \frac{19}{115}, -\frac{21}{115}) \left(\begin{bmatrix} y_2 \\ y_3 \\ y_4 \end{bmatrix} - \begin{bmatrix} 4 \\ -1 \\ -2 \end{bmatrix} \right) \\
&= 3 + \left(-\frac{106}{115}y_2 + \frac{19}{115}y_3 - \frac{21}{115}y_4 \right) + \frac{401}{115} \\
&= -\frac{106}{115}y_2 + \frac{19}{115}y_3 - \frac{21}{115}y_4 + \frac{746}{115}
\end{aligned}$$

and for covariance matrix, we have:

$$\begin{aligned}
cov[\mathbf{y}|\mathbf{x}] &= \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \\
&= 3 - (-4, 2, 1) \begin{bmatrix} 4 & -3 & -1 \\ -3 & 2 & 6 \\ -1 & 6 & 5 \end{bmatrix}^{-1} \begin{bmatrix} -4 \\ 2 \\ 1 \end{bmatrix} \\
&= 3 - (-4, 2, 1) \begin{bmatrix} \frac{26}{115} & -\frac{9}{115} & \frac{16}{115} \\ -\frac{9}{115} & -\frac{19}{115} & \frac{21}{115} \\ \frac{16}{115} & \frac{21}{115} & \frac{1}{115} \end{bmatrix} \begin{bmatrix} -4 \\ 2 \\ 1 \end{bmatrix} \\
&= 3 - (-\frac{106}{115}, \frac{19}{115}, -\frac{21}{115}) \begin{bmatrix} -4 \\ 2 \\ 1 \end{bmatrix} \\
&= 3 - \frac{441}{115} \\
&= -\frac{96}{115}
\end{aligned}$$

Thus completed the solution for part [A].

For part [B]:

To find the MLE, let's first compute a few necessary components:

Given sample:

$$\mathbf{X} = \begin{bmatrix} 3 & 6 \\ 4 & 4 \\ 5 & 7 \\ 4 & 7 \end{bmatrix}$$

we have:

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \bar{\mathbf{x}} = \begin{bmatrix} \frac{3+4+5+4}{4} \\ \frac{6+4+7+7}{4} \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

and also:

$$s_{11} = \frac{1}{4}\sum_{i=1}^{4}(x_{i1} - \bar{x}_1)^2 = \frac{1}{2}$$
$$s_{12} = \frac{1}{4}\sum_{i=1}^{4}(x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) = \frac{1}{4}$$
$$s_{21} = s_{12} = \frac{1}{4}$$
$$s_{22} = \frac{1}{4}\sum_{i=1}^{4}(x_{i2} - \bar{x}_2)^2 = \frac{3}{2}$$

So the MLE for μ and Σ is:

$$\hat{\mu} = \bar{\mathbf{x}} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$
$$\hat{\Sigma} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{2} \end{bmatrix}$$

For part [C]:

The first method for testing bivariate normality:

Mardia's multivariate normality test:

Mardia (1970) proposed a multivariate normality test which is based on multivariate extensions of skewness ($\hat{\gamma}_{1,p}$) and kurtosis ($\hat{\gamma}_{2,p}$) measures as follows:

$$\hat{\gamma}_{1,p} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}^3 \qquad \textit{and } \hat{\gamma}_{2,p} = \frac{1}{n}\sum_{i=1}^{n} m_{ii}^2$$

Here

$$m_{ij} = (\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}), 1 \le i, j \le p$$

and p is the number of variables. For bivariate case, simply put $p = 2$.

m_{ij} *is called the squared Mahalanobis distance, and the test statistic for skewness is* $\frac{n}{6}\hat{\gamma}_{1,p}$, *which follows approximately* χ^2 *distribution with degree of freedom* $p(p+1)(p+2)/6$ *(so it is* 4 *for bivariate case). Also, the test statistic for kurtosis,* $\hat{\gamma}_{2,p}$ *is approximately normally distributed with mean* $p(p+2)$ *and variance* $8p(p+2)/n$.

To test normality, both p values of skewness and kurtosis statistics should be greater than 0.05 *(null hypothesis is normality).*

For small samples, the power and type I error could be violated. Therefore, Mardia introduced a correction term into skewness test statistic, usually when $n < 20$, *in order to control type I error. The corrected skewness statistic for small samples is* $\frac{nk}{6}\hat{\gamma}_{1,p}$, *where* $k = (p+1)(n+1)(n+3)/(n(n+1)(p+1)-6)$. *This statistic is also distributed as* χ^2 *with degrees of freedom* $p(p+1)(p+2)/6$.

The second method for testing bivariate normality*:*

Henze-Zirkler's multivariate normality test:

The Henze-Zirkler's test is based on a non-negative functional distance that measures the distance between two distribution functions. If data are distributed as ultivariate normal, the test statistic is approximately log-normally distributed. First, the mean, variance and smoothness parameter are calculated. Then the mean and the variance are log-normalized and the p-value is estimated. The test statistic of Henze-Zirkler's multivariate normality test is given in the following equation:

$$HZ = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} e^{-\frac{\beta^2}{2}D_{ij}} - 2(1+\beta^2)^{-\frac{p}{2}}\sum_{i=1}^{n} e^{-\frac{\beta^2}{2(1+\beta^2)}D_i} + n(1+2\beta^2)^{-\frac{p}{2}}$$

where

$$\begin{aligned} p &: \text{ number of variables} \\ \beta &= \frac{1}{\sqrt{2}}\Big(\frac{n(2p+1)}{4}\Big)^{\frac{1}{p+4}} \\ D_{ij} &= (\mathbf{x}_i - \bar{x}_j)'\mathbf{S}^{-1}(\mathbf{x}_i - \mathbf{x}_j) \\ D_i &= (\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) = m_{ii} \end{aligned}$$

From the above equation, D_i *gives the squared Mahalanobis distance of* i^{th} *observation to the centroid and* D_{ij} *gives the Mahalanobis distance between* i^{th} *and* j^{th} *observations. If data are multivariate normal, the test statistic* (HZ) *is approximately log-normally distributed with mean* μ *and variance* σ^2 *as given below:*

$$\begin{aligned} \mu &= 1 - \frac{a^{-\frac{p}{2}}\Big(1 + p\beta^{\frac{2}{a}} + p(p+2)\beta^4\Big)}{2a^2} \\ \sigma^2 &= 2(1+4\beta^2)^{-\frac{p}{2}} + \frac{2a^{-p}(1+2p\beta^4)}{a^2} + \frac{3p(p+2)\beta^8}{4a^4} - 4w_\beta^{-\frac{p}{2}}\Big(1 + \frac{3p\beta^4}{2w_\beta} + \frac{p(p+2)\beta^8}{2w_\beta^2}\Big) \end{aligned}$$

where $a = 1 + 2\beta^2$ and $w_\beta = (1+\beta^2)(1+3\beta^2)$. Hence the log-normalized mean and variance of the HZ statistic can be defined as follows:

$$\log(\mu) = \log\left(\sqrt{\frac{\mu^4}{\sigma^2+\mu^2}}\right) \qquad \text{and } \log(\sigma^2) = \log\left(\frac{\sigma^2+\mu^2}{\sigma^2}\right)$$

By using the log-normal distribution parameters, μ and σ, we can test the significance of multivariate (or bivariate when $p = 2$)normality. The Wald test statistic for multivariate normlality is given as following:

$$z = \frac{\log(HZ) - \log(\mu)}{\log(\sigma)}$$

Question 4.

Solution 4. *Part [A]*

part (i)

We read in the data and make the scatter plot from the plots option of proc corr:

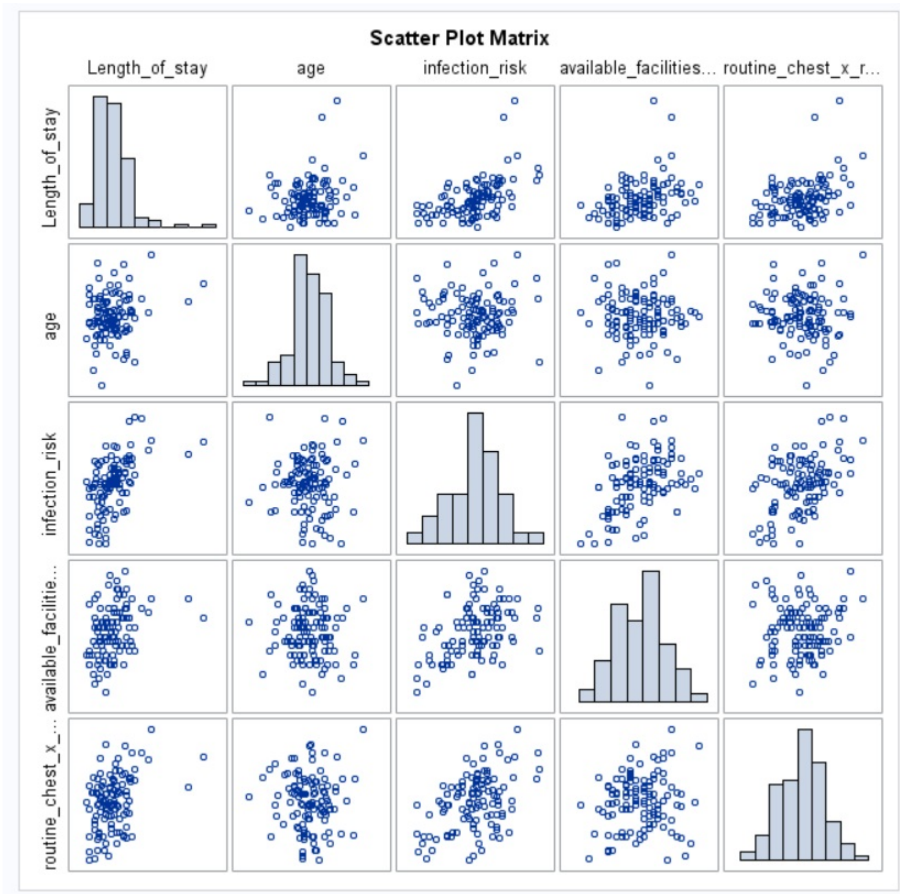
```
libname data  "C:\akira\data";

/*import the data*/
proc import datafile = "C:\akira\data\senic.csv"
    dbms = dlm
    /*out = data.senic*/
    out = senic
    replace;
    delimiter = ',';
    getnames = yes;
run;


proc print data = senic;
run;


/*to make scatter plots*/
ods graphics on;
ods html;
title 'senic data analysis';
proc corr data=senic nomiss plots=matrix(histogram);
   var length_of_stay age infection_risk available_facilities_and_service routine_chest_x_ray_ratio;
run;
ods graphics off;
ods html close;
```

The plot output is:



From the plot there seem to be association between Y and the rest of X variables.

For (ii)

We are going to estimate b by the following formula using proc IML:

$$\hat{\mathbf{b}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

We build X and Y matrix and compute $\hat{\beta}$ in proc IML with the following code:

```
proc iml;
/*use the senic data we imported*/
use work.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;

/*find the number of observations*/
d1 = nrow(X[, 1]);
print d1;

/*create column of 1s and insert into X*/
intercept = j(d1,1,1);
print intercept;
X = intercept||X;

/*create labels for matrices*/
cY = {"Y"};
cX = {"1" "X1" "X2" "X3" "X4"};
mattrib Y colname=cY X colname=cX;

/*optional steps, print to check Y and X*/
print Y;
print X;

/*OLS for beta*/
est_beta =inv(t(X)*X)*t(X)*Y;
print est_beta;
```

and the output is:

```
est_beta

0.1801384
0.0856936
0.5184167
0.0240421
0.0197392
```

For (iii):

In the model, β_0 as the intercept does not give much meaning. If there is no patient data, then there should be no length of hospital stay. In fact our estimate above for β_0 is very close to 0.

β_1 means given fixed X_2, X_3, X_4, on average how long the hospital stay will change when the age increase by 1.

β_2 means given fixed X_1, X_3, X_4, on average how long the hospital stay will change when the infection risk increase by 1.

β_3 means given fixed X_1, X_2, X_4, on average how long the hospital stay will change when the available faciilites and services increase by 1.

β_4 means given fixed X_1, X_2, X_3, on average how long the hospital stay will change when the routine chest X-ray ratio increase by 1.

For (iv):

We compute the eigenvalues and eigenvecotrs of $(\mathbf{X}'\mathbf{X})^{-1}$ in proc IML as following:

```
/*eigen values and eigen vectors of (X'X)^{-1}*/
a = inv(t(X)*X);
call eigen( e, u, a);
print,,, a,,, 'Eigenvalues and eigenvectors', e u;
```

The output is the following: (the column e is the 5 eigen values, and each column of u represents one eigenvector.)

```
                         a

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

               Eigenvalues and eigenvectors
        e         u

1.5292165 0.9998722 0.0009573 0.0128818 0.0022527 0.0091475
0.0075224 -0.015861 -0.010893 0.8640027 0.1241409 0.4875636
0.0000849 -0.001201 0.9989865 -0.012515 0.0139522 0.0409051
0.0000375 -0.001063 -0.032753 -0.350156  0.844055 0.4048307
7.5893E-7 -0.001209 -0.028873 -0.361341 -0.521498 0.7724228
```

Notice that all eigen values are positive, and hence $(\mathbf{X}'\mathbf{X})^{-1}$ is positive definite.

So the data is of full rank.

Remark: I suspected first that it may not be of full rank since one of the eigen value is on the scale of 10^{-7} and may be actually 0 due to rounding error. So I went back and checked the eigen values on $(\mathbf{X}'\mathbf{X})$ and it came back with all 5 eigen values positive.

For (v):

Since $(\mathbf{X}'\mathbf{X})$ has the same rank as $\mathbf{X}$, and it is easier to handle the rank problem for a square matrix, so we will compute the rank of $\mathbf{X}'\mathbf{X}$ instead.

We comopute the rank of $\mathbf{X}'\mathbf{X}$ in proc IML as follows:

```
/*compute the rank of X'X*/
a = t(X)*X;
ranka=round(trace(ginv(a)*a));
call eigen(e, u, a);
print a ranka;
print e u;
```

The output is then:

a					ranka
113	6015.2	492.1	4877	9224	5
6015.2	322429.74	26196.13	259304.51	490828.22	
492.1	26196.13	2344.41	22180.61	41487.8	
4877	259304.51	22180.61	236367.28	401791.21	
9224	490828.22	41487.8	401791.21	794934.88	

The 5×5 matrix on the left is our $\mathbf{X}'\mathbf{X}$ which has the same rank as $\mathbf{X}$, and the right hand side shows the rank is 5, which is consistent with our results from last question that all 5 eigenvalues of $(\mathbf{X}'\mathbf{X})^{-1}$ are positive.

For (vi):

With the estimation $\mathbf{b}$ already computed before, it is easy to compute the estimated mean $\hat{\mathbf{Y}}$ and the residuals under proc iml:

We have the following code:

```
/*compute the estimatd means est_Y*/
est_Y = X*est_beta;
/*compute the residuals*/
res = Y - est_Y;

print est_Y res;
```

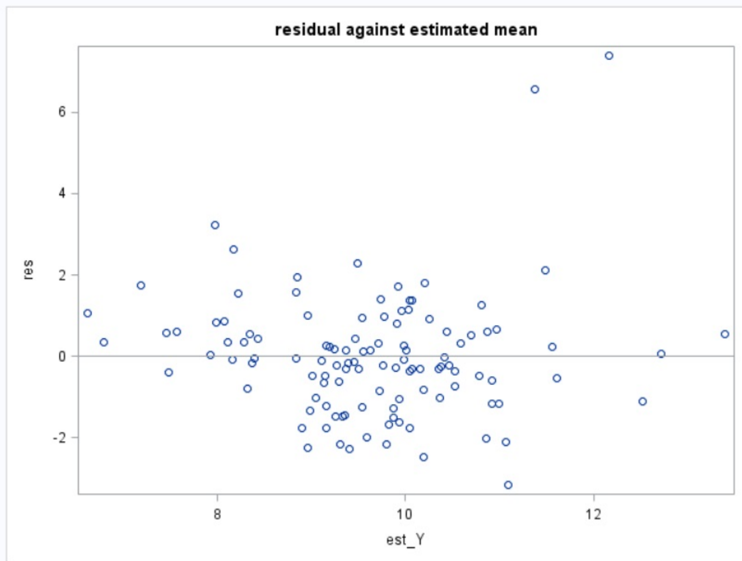
The result of $\hat{\mathbf{Y}}$ and $\mathbf{e}$ are both column vectors of length 113, so we only show a portion of the output here:

est_Y	res
9.3029753	-2.172975
7.9791702	0.8408298
8.3973695	-0.05737
11.070673	-2.120673
10.693298	0.5067022
10.062251	-0.302251
10.039023	-0.359023
10.031932	1.1480677
9.2937623	-0.623762
10.862852	-2.022852
11.610826	-0.540826

The following code gives a plot of residual against the estimated means:

```
/*scatter plot of residual against estimated mean*/
ods html;
title "residual against estimated mean";
run Scatter(est_Y, res);
ods html close;
```

The output is:



There is some outliers but most are scattered around 0 *pretty closely. This indicates that the data fits well to the assumption of equal variance.*

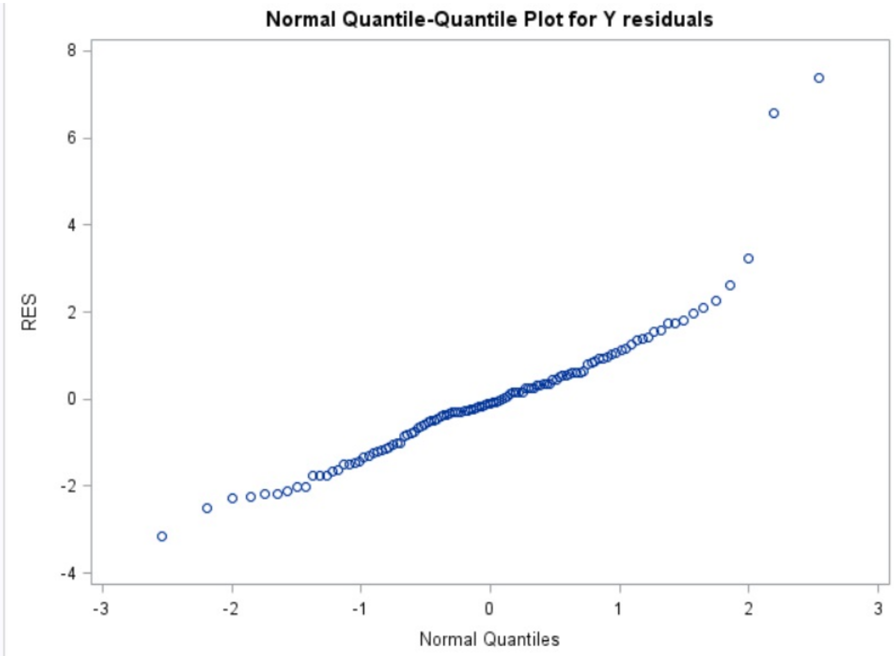
For (vii):

To create the normal qq-plot for residuals, we have the following code:

```
/*export resiual to a SAS data set called work.residual*/
/*we make qq normal plot of residual outside of proc iml later*/
create residual var{res};
append;
close residual;


/*scattered normal qq plot for residual*/
title "Normal Quantile-Quantile Plot for Y residuals";
ods graphics on;
ods html;
proc univariate data=work.residual;
    qqplot res/odstitle=title;
run;
ods graphics off;
ods html close;
```

and the output is:



The qq plot is showing a straight line trend for the residual against the normal quantile, so we conclude that the normal assumption on the data is valid.

For part [B]:

For (i)

we only need to do minor adjustification to our previous SAS code. I added "_a", "_b" to some of the variable names to differ between those for part [A] and part [B]. To estiamte the parameter under the new model without intercept, we have the following code (see my comment in between codes which is for part A which is for part B):

```
/*OLS for beta*/
/*variables with _a is for part A model*/
/*variables with _b is for part B model*/
est_beta_a =inv(t(X_a)*X_a)*t(X_a)*Y;
est_beta_b = inv(t(X_b)*X_b)*t(X_b)*Y;
print est_beta_a;
print est_beta_b;
```

the output for the estimate parameter is

est_beta_b
0.088551
0.5186322
0.0242336
0.019957

This model is different from model 1 because we did not assume intercept in the first place, which should actually make more sense as I explained in part [A], that given no patient information, there should be no length of hospital stay.

For part (ii):

Similar to the process as in part [A], the rank of $\mathbf{X}$ *is the same as the rank of* $\mathbf{X}'\mathbf{X}$*, and we compute it with the following code:*

```
/*compute the rank of X'X*/
a = t(X_a)*X_a;
b = t(X_b)*X_b;
ranka=round(trace(ginv(a)*a));
rankb=round(trace(ginv(b)*b));
call eigen(e_a, u_a, a);
call eigen(e_b, u_b, b);
print a ranka;
print e_a u_a;
print b rankb;
print e_b u_b;
```

and the output is:

b				rankb
322429.74	26196.13	259304.51	490828.22	4
26196.13	2344.41	22180.61	41487.8	
259304.51	22180.61	236367.28	401791.21	
490828.22	41487.8	401791.21	794934.88	

e_b	u_b			
1317529	0.487583	0.1241316	0.8641388	-0.010878
26638.264	0.0409068	0.0139534	-0.012511	0.9989872
11776.097	0.4048476	0.8440737	-0.350093	-0.032752
132.93686	0.7724557	-0.521475	-0.361306	-0.028872

as we can see the rank of design matrix $\mathbf{X}$ *is* 4 *and hence it is full rank.*

For part (iii):

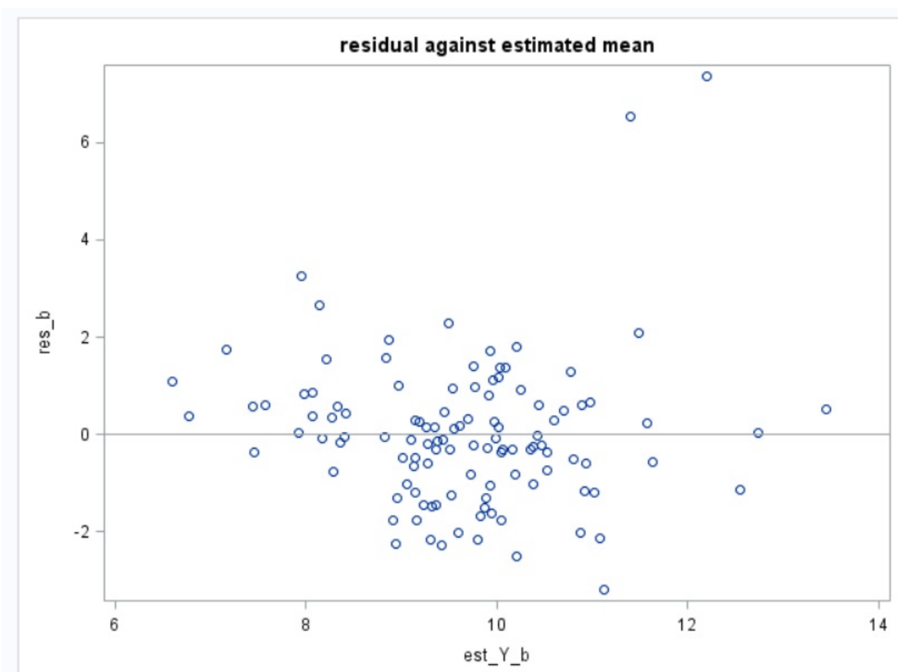
The code for computing residual and plotting scatter plot of residual against estimated mean:

```
/*compute the estimatd means est_Y*/
est_Y_a = X_a*est_beta_a;
est_Y_b = X_b*est_beta_b;
/*compute the residuals*/
res_a = Y - est_Y_a;
res_b = Y - est_Y_b;

print est_Y_a res_a;
print est_Y_b res_b;


/*scatter plot of residual against estimated mean*/
ods html;
title "residual against estimated mean";
run Scatter(est_Y_a, res_a)
    /*add reference line*/
    other = "refline 0/axis = y"
    ;
run Scatter(est_Y_b, res_b)
    /*add refernece line*/
    other = "refline 0/axis = y"
    ;
ods html close;
```

The scatter plot:



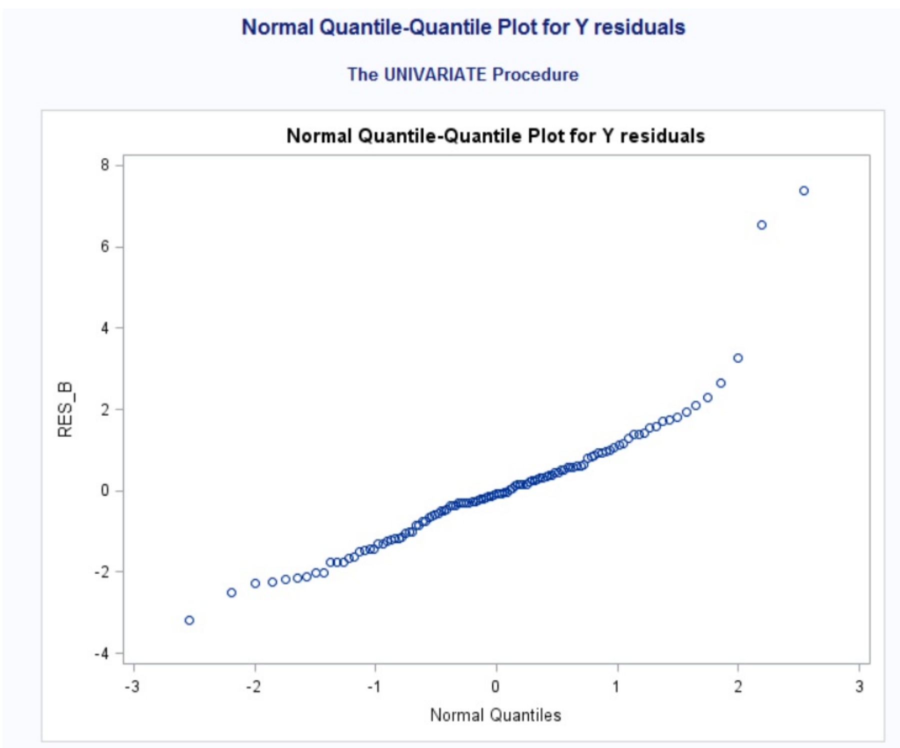
The residuals are pretty evenly scattered around 0 *except a few outliers, and most stay within rage between* -2 *and* 2*, so we consider the independence and equal variance assumption holds.*

The code for normal quantile plot:

```
create residual_b var{res_b};
append;
close residual_b;


/*scattered normal qq plot for residual*/
title "Normal Quantile-Quantile Plot for Y residuals";
ods graphics on;
ods html;
proc univariate data=work.residual;
    qqplot res/odstitle=title;
run;
proc univariate data=work.residual_b;
    qqplot res_b/odstitle=title;
run;
ods graphics off;
ods html close;
```

Output:



As we can see that the q-q plot display a straight line trend, and hence we consider the normality assumption holds.

The plots we have here are actually pretty similar compared to part [A]. I printed out the residuals and actually find that in both model 1 and and model 2 the residuals are pretty similar too.

part (iv):

The definition of estimability says that, for any given linear function of β, say $\mathbf{C}\beta$, it is said to be estimatble if there is a linear function of response $\mathbf{y}$, denoted as $\mathbf{Ay}$, such that $\mathbf{Ay}$ is an unbiased estimator of $\mathbf{C}\beta$.

In our case, we have $\mathbf{C} = \mathbf{X}$ and $\mathbf{A} = \mathbf{I}$, and since $E[\mathbf{Ay}] = E[\mathbf{y}] = \mathbf{X}\beta$, hence $\mathbf{y}$ is an unbiased estimator of $\mathbf{X}\beta$ and hence $\mathbf{X}\beta$ is estimable.

Of course the above conclusion holds only when the model is appropriate, which we have verified by checking those model assumptions.

Question #5.

Solution 5. *For part [A]:*

For part (i):

We have:

$$
\begin{aligned}
&(\mathbf{I} - \mathbf{XX}^{(-)})(\mathbf{I} - \mathbf{XX}^{(-)}) \\
&= \mathbf{I} - 2\mathbf{XX}^{(-)} + \mathbf{XX}^{(-)}\mathbf{XX}^{(-)} \\
&= \mathbf{I} - 2\mathbf{XX}^{(-)} + \mathbf{XX}^{(-)} \\
&= \mathbf{I} - \mathbf{XX}^{(-)}
\end{aligned}
$$

So $\mathbf{I} - \mathbf{XX}^{(-)}$ is idempotent.

On the other hand, since rank of $\mathbf{X}$ is k, so rank of $\mathbf{XX}^{(-)}$ is also k. Since it is symmetric, there exists orthogonal matrix $\mathbf{C}$ such that:

$$\mathbf{XX}^{(-)} = \mathbf{CDC}^T$$

where $\mathbf{D}$ is a diagonal matrix whose first k diagonal entries are the k eigenvalues of $\mathbf{D}$ and the rest $n-k$diagonal entries are 0.

So

$$
\begin{aligned}
&\mathbf{I} - \mathbf{XX}^{(-)} = \mathbf{CC}^T - \mathbf{CDC}^T \\
&= \mathbf{C}\Big(\mathbf{I} - \mathbf{D}\Big)\mathbf{C}^T
\end{aligned}
$$

since the rank of $\mathbf{I} - \mathbf{D}$ is $n-k$, so the rank of $\mathbf{I} - \mathbf{XX}^{(-)}$ is also $n-k$.

We already showed that $\mathbf{I} - \mathbf{D}$ is idempotent, and given $\mathbf{Y}$ that is $N(\mathbf{0}, \mathbf{I})$, we know that

$$Q_1 = \mathbf{Y}^T(\mathbf{I} - \mathbf{XX}^{(-)})\mathbf{Y} \sim \chi^2_{n-k}$$

from Corollary 1 for theorem 5.5 on page 118.

For (ii):

Notice that we have:

$$
\begin{aligned}
\mathbf{XX}^{(-)}\mathbf{X} &= \mathbf{XX}^{(-)}[\mathbf{X}_1, \mathbf{X}_2] \\
&= [\mathbf{XX}^{(-)}\mathbf{X}_1, \mathbf{XX}^{(-)}\mathbf{X}_2] \\
&= \mathbf{X} \\
&= [\mathbf{X}_1, \mathbf{X}_2]
\end{aligned}
$$

This implies that

$$\mathbf{X}\mathbf{X}^{(-)}\mathbf{X}_2 = \mathbf{X}_2$$

and hence further

$$\mathbf{X}\mathbf{X}^{(-)}\mathbf{X}_2\mathbf{X}_2^{(-)} = \mathbf{X}_2\mathbf{X}_2^{(-)}$$

we are going to use this in the proof for this question and in the later question too.

Now we have $\left(\mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}\right)$ *as a symmetric matrix and satisfy the following:*

$$\begin{aligned}
&\left(\mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}\right)\left(\mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}\right)\\
&= \mathbf{X}\mathbf{X}^{(-)}\mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}\mathbf{X}^{(-)}\mathbf{X}_2\mathbf{X}_2^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}\mathbf{X}\mathbf{X}^{(-)} + \mathbf{X}_2\mathbf{X}_2^{(-)}\mathbf{X}_2\mathbf{X}_2^{(-)}\\
&= \mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)} - \underbrace{\mathbf{X}\mathbf{X}^{(-)}\mathbf{X}_2\mathbf{X}_2^{(-)}}_{by\ symmetry} + \mathbf{X}_2\mathbf{X}_2^{(-)}\\
&= \mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)} + \mathbf{X}_2\mathbf{X}_2^{(-)}\\
&= \mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}
\end{aligned}$$

which says that $\mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}$ *is idempotent.*

Finally due to the idempotency, we have

$$\begin{aligned}
&rank\left(\mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}\right) = tr\left(\mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}\right)\\
&= tr\left(\mathbf{X}\mathbf{X}^{(-)}\right) - tr\left(\mathbf{X}_2\mathbf{X}_2^{(-)}\right)\\
&= rank\left(\mathbf{X}\mathbf{X}^{(-)}\right) - rank\left(\mathbf{X}_2\mathbf{X}_2^{(-)}\right)\\
&= rank(\mathbf{X}) - rank(\mathbf{X}_2)\\
&= k - m
\end{aligned}$$

So by Corollary 1 for theorem 5.5 *on page* 118*,*

$$Q_2 = \mathbf{Y}^T\left(\mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}\right)\mathbf{Y} \sim \chi^2_{k-m}$$

For part (iii):

It is easy to check that

$$\begin{aligned}
&\left(\mathbf{X}_2\mathbf{X}_2^{(-)}\right)\left(\mathbf{X}_2\mathbf{X}_2^{(-)}\right)\\
&= \left(\mathbf{X}_2\mathbf{X}_2^{(-)}\mathbf{X}_2\right)\mathbf{X}_2^{(-)}\\
&= \mathbf{X}_2\mathbf{X}_2^{(-)}
\end{aligned}$$

Hence $\mathbf{X}_2\mathbf{X}_2^{(-)}$ *is idempotent with rank* m*, so by Corollary 1 for theorem* 5.5 *on page* 118*, we have:*

$$Q_3 = \mathbf{Y}^T\Big(\mathbf{X}_2\mathbf{X}_2^{(-)}\Big)\mathbf{Y} \sim \chi^2_m$$

For part (iv):

Q_1, Q_2, Q_3 *are all quadratic forms. To prove pairwise independence, by theorem 5.6b and its corollary* 1 *on page 120, we only need to show that the multiplication of their quadratic matrix is* $\mathbf{0}$.

We have:

$$\begin{aligned}
&\Big(\mathbf{I} - \mathbf{X}\mathbf{X}^{(-)}\Big)\Big(\mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}\Big)\\
&= \mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)} - \mathbf{X}\mathbf{X}^{(-)}\mathbf{X}\mathbf{X}^{(-)} + \mathbf{X}\mathbf{X}^{(-)}\mathbf{X}_2\mathbf{X}_2^{(-)}\\
&= \mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)} - \mathbf{X}\mathbf{X}^{(-)} + \underbrace{\mathbf{X}_2\mathbf{X}_2^{(-)}}_{\textit{proved in part (ii)}}\\
&= \mathbf{0}
\end{aligned}$$

Hence Q_1 *and* Q_2 *are independent.*

We have:

$$\begin{aligned}
&\Big(\mathbf{I} - \mathbf{X}\mathbf{X}^{(-)}\Big)\mathbf{X}_2\mathbf{X}_2^{(-)}\\
&= \mathbf{X}_2\mathbf{X}_2^{(-)} - \mathbf{X}\mathbf{X}^{(-)}\mathbf{X}_2\mathbf{X}_2^{(-)}\\
&= \mathbf{X}_2\mathbf{X}_2^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}\\
&= \mathbf{0}
\end{aligned}$$

Hence Q_1 *and* Q_3 *are independent.*

Finally we have:

$$\begin{aligned}
&\Big(\mathbf{X}\mathbf{X}^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}\Big)\mathbf{X}_2\mathbf{X}_2^{(-)}\\
&= \mathbf{X}\mathbf{X}^{(-)}\mathbf{X}_2\mathbf{X}_2^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}\mathbf{X}_2\mathbf{X}_2^{(-)}\\
&= \mathbf{X}_2\mathbf{X}_2^{(-)} - \mathbf{X}_2\mathbf{X}_2^{(-)}\\
&= \mathbf{0}
\end{aligned}$$

Hence Q_2 *and* Q_3 *are independent.*

Thus completed the proof.

For part [B]*:*

Question 5.20 *from the book:*

Theorem 5.5 *says that, if* $\mathbf{y}$ *is* $N_p(\mu, \Sigma)$*, and* $\mathbf{A}$ *is symmetric with rank* r*, let* $\lambda = \frac{1}{2}\mu'\mathbf{A}\mu$*. Then* $\mathbf{y}'\mathbf{A}\mathbf{y}$ *is* $\chi^2(r, \lambda)$*, if and only if* $\mathbf{A}\boldsymbol{\Sigma}$ *is idempotent. We use the reuslt of this theorem to prove the other corollaries.*

For part (a):

Given $\mathbf{y}$ *is* $N_p(\mathbf{0}, \mathbf{\Sigma})$, *we have* $\mu = 0$, *and hence:*

$$\lambda = \frac{1}{2}\mu'\mathbf{A}\mu = 0$$

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

Hence by Theorem 5.5, $\mathbf{y}'\mathbf{A}\mathbf{y}$ *is* $\chi^2(r, 0)$ *if and only if* $\mathbf{A\Sigma}$ *is idempotent of rank* r. *But* $\chi^2(r, 0)$ *is just* $\chi^2(r)$. *Proof completed.*

For part (b):

Given $\mathbf{y}$ *as* $N_p(\mu, \sigma^2\mathbf{I})$, *we have* $\Sigma = \sigma^2\mathbf{I}$, *and if we consider* $\mathbf{y}'\mathbf{y}/\sigma^2$, *we have* $\mathbf{A} = \frac{1}{\sigma^2}\mathbf{I}$.

Thus

$$\mathbf{A\Sigma} = \frac{1}{\sigma^2}\mathbf{I}\sigma^2\mathbf{I} = \mathbf{I} \textit{ which is idempotent}$$

Hence $\mathbf{y}'\mathbf{y}/\sigma^2$ *is* $\chi^2(p, \mu'\mu/2\sigma^2)$

For part (c):

Given $\mathbf{y}$ *is* $N_p(\mu, \mathbf{I})$, *we have* $\Sigma = \mathbf{I}$, *hence the condition of* $\mathbf{A\Sigma}$ *is idempotent becomes* $\mathbf{A\Sigma} = \mathbf{AI} = \mathbf{A}$ *is idempotent with rank* r. *Proof copmleted.*

For part (d):

Given $\mathbf{y}$ *is* $N_p(\mu, \sigma^2\Sigma)$, *we have* $\frac{\mathbf{y}}{\sigma}$ *is* $N_p(\frac{\mu}{\sigma}, \Sigma)$. *We can denote* $\mathbf{y}/\sigma$ *by* $\mathbf{z}$, *then when we consider* $\mathbf{y}'\mathbf{A}\mathbf{y}/\sigma^2$, *it is equivalent to thinking about* $\mathbf{z}'\mathbf{A}\mathbf{z}$. *Since the variance matrix for* $\mathbf{z}$ *is* Σ, *and the mean is* $\frac{\mu}{\sigma}$, *so the* χ^2 *distribution's noncentral parameter is*

$$\frac{1}{2} \times \frac{\mu'}{\sigma}\mathbf{A}\frac{\mu}{\sigma} = \frac{\mu'\mathbf{A}\mu}{2\sigma^2}$$

and the necessary and sufficient condition is for $\mathbf{A\Sigma}$ *to be idempotent.*

For part (e):

We can use the result from part (d), take $\mathbf{A} = \Sigma^{-1}$, *then*

$$\mathbf{A}\Sigma = \Sigma^{-1}\Sigma = \mathbf{I} \textit{ which is idempotent with rank } p$$

Hence from part (d),

$$\mathbf{y}'\mathbf{A}\mathbf{y}/\sigma^2 = \mathbf{y}'\Sigma^{-1}\mathbf{y}/\sigma^2 \sim \chi^2(p, \mu'\Sigma^{-1}\mu/2\sigma^2)$$

Thus completed the proof.

Question #6.

Solution 6. *For part (i):*

I confirm that I have read and understood the material on Singular Value Decomposition and know that it is a generalization of the case of eigen values and eigen vectors. (I have also learned this from the numerical analysis class I took from my math degree)

For part (ii):

I confirm that I have read and understood the material on the relationship between determintants of 2×2 matrix and area of parallelogram.

Some of the proofs are more intuitve than the others but actually my favorite proof is the following (not given by the link), which I taught to undergraduate students at KU math when I used to work there as a lecturer.

Given (a, b) and (c, d) as 2d row vectors, consider they are special case of 3d vectors: $\mathbf{u} = (0, a, b)$ *and* $\mathbf{v} = (0, c, d)$. *By definition of cross product, we know that* $|det\Big(\mathbf{u} \times \mathbf{v}\Big)| = |\mathbf{u}||\mathbf{v}|\sin(\theta)$ *is the area of parallelagram formed by* $\mathbf{u}$ *and* $\mathbf{v}$ *. (Here θ is the angle between* $\mathbf{u}$ *and* $\mathbf{v}$*). We also have:*

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & a & b \\ 0 & c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \mathbf{i}$$

So

$$|det\Big(\mathbf{u} \times \mathbf{v}\Big)| = \left| \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right| = |\mathbf{u}||\mathbf{v}|\sin(\theta) = \text{ area of parallelagram formed by } \mathbf{u} \text{ and } \mathbf{v}$$