3.2 Given the data matrix
\[ X = \begin{bmatrix} 3 & 4 \\ 6 & -2 \\ 3 & 1 \end{bmatrix} , \]

(a) Graph the scatter plot in \( p = 2 \) dimension and locate the sample mean on your diagram.

(b) Sketch the \( n = 3 \) space representation of the data, and plot the deviation vectors \( y_1 - \bar{x}_1 \) and \( y_1 - \bar{x}_1 \).

(c) Sketch the deviation vectors in (b) emanating from the origin. Calculate their lengths and the cosine of the angle between them. Relate these quantities to \( S_n \) and \( R \).

\[
\text{Sln: (a)} \quad \bar{x} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 3+4+1 \\ 4+2+1 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix} .
\]

\[
\text{(b)} \quad y_1 - \bar{x}_1 = \begin{bmatrix} 3-4 \\ 6-4 \\ 3-4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} , \quad y_2 - \bar{x}_2 = \begin{bmatrix} 4-1 \\ -2-1 \\ 1-1 \end{bmatrix} = \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix} .
\]

\[
\text{(c)} \quad \|d_i\| = \sqrt{\frac{3+4+1}{2} = \sqrt{6} , \|d_i\| = \sqrt{\frac{4+2+1}{2} = \sqrt{2} . \cos \theta_{12} = \frac{d_1^T d_2}{\|d_1\|\|d_2\|} = \frac{-\frac{3}{2}}{2} . \text{Since we have} \|d_i\|^2 = n S_{ii} , (i = 1, 2) \text{ and } d_i^T d_k = n S_{ik} (i, k = 1, 2; i \neq k) , \text{ we calculate the } S_n \text{ as:}
\]

\[
S_n = \frac{1}{2} \begin{bmatrix} \|d_1\|^2 & d_1^T d_2 \\ d_2^T d_1 & \|d_2\|^2 \end{bmatrix} = \begin{bmatrix} 3 & -\frac{9}{2} \\ -\frac{9}{2} & 9 \end{bmatrix} ; \text{Similarly, } \cos \theta_{12} = \frac{d_1^T d_2}{\|d_1\|\|d_2\|} = \frac{S_{12}}{\sqrt{S_{11} S_{22}}} , \text{ we calculate}
\]

\[
R \text{ as: } R = \frac{1}{2} \begin{bmatrix} \cos \theta_{12} & 1 \\ 1 & \cos \theta_{21} \end{bmatrix} = \begin{bmatrix} 1 & -\frac{\sqrt{6}}{2} \\ -\frac{\sqrt{6}}{2} & 1 \end{bmatrix} .
\]

3.3 Perform the decomposition of \( y_1 \) into \( \bar{x}_1 \) and \( y_1 - \bar{x}_1 \) using the first column of the data matrix in Example 3.9.

\[
\text{Sln: } y_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix} , \text{ the projection of } y_1 \text{ into } \bar{x}_1 \text{ is } y_1^T \cdot \bar{x}_1 = \frac{1+4+4}{3} = \frac{3}{3} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} . \]

\[
y_1 - \bar{x}_1 = \begin{bmatrix} 1-3 \\ 4-3 \\ 4-3 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} .
\]
Solutions Stat755 Homework 3

3.6 Consider the data matrix \( X = \begin{bmatrix} -1 & 3 & -2 \\ 2 & 4 & 2 \\ 5 & 2 & 3 \end{bmatrix} \). (a) Calculate the matrix of deviations (residuals). \( X - \bar{X}^T \). Is this matrix of full rank? Explain.

Sln: We have observation matrix \( X \), so \( \bar{X} = \frac{1}{n} X^T \). \( 1 = \frac{1}{3} X^T 1 = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix} = \begin{bmatrix} -1 + 2 + 5 \\ 3 + 4 + 2 \\ -1 + 2 + 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} \). Since \( d_i = y_i - \bar{x}_i \cdot 1 \), the deviations matrix is:
\[
\begin{bmatrix}
-1 & 3 & -2 \\
2 & 4 & 2 \\
5 & 2 & 3
\end{bmatrix} - \begin{bmatrix}
-1 & 3 & -2 \\
2 & 4 & 2 \\
5 & 2 & 3
\end{bmatrix} = \begin{bmatrix}
-3 & 0 & -3 \\
0 & 1 & 1 \\
3 & -1 & 2
\end{bmatrix}.
\]

The matrix rank is 2. Because the third column is a linear combination of the first and second column. +1 \[
\begin{bmatrix}
-3 \\
0 \\
3
\end{bmatrix} + 1 \begin{bmatrix}
0 \\
1 \\
-1
\end{bmatrix} = \begin{bmatrix}
-3 \\
1 \\
2
\end{bmatrix}.
\]

3.12 Show that \( |S| = (s_{11} s_{22} \cdots s_{pp}) |R| \).

Sln: \( S = D^{1/2} R D^{1/2} \). By taking the determinant gives \( |S| = |D|^{1/2} |R| |D|^{1/2} \). Since \( |D|^{1/2} = \sqrt{s_{11}} \sqrt{s_{22}} \cdots \sqrt{s_{pp}} \), we have \( |D|^{1/2} |D|^{1/2} = s_{11} s_{22} \cdots s_{pp} \). Therefore, \( |S| = (s_{11} s_{22} \cdots s_{pp}) |R| \).

3.13 Given a data matrix \( X \) and the resulting sample correlation matrix \( R \). Consider the standardized observations \( \left( \frac{x_{jk} - \bar{x}_k}{\sqrt{s_{kk}}} \right) \), \( k = 1, 2, \cdots, \), \( p; j = 1, 2, \cdots, n \). Show that these standardized quantities have sample covariance matrix \( R \).

Sln: let \( Z = \{ z_{jk} \} = \left( \frac{x_{jk} - \bar{x}_k}{\sqrt{s_{kk}}} \right), k = 1, 2, \cdots, p; j = 1, 2, \cdots, n \). be the standardized matrix, then \( z_{kk} = \frac{1}{n} \sum_{j=1}^{n} z_{jk} \), \( k = 1, 2, \cdots, p \). Therefore, the covariance matrix \( R \) of the \( Z \) is calculated as:
\[
s_{kl} = \frac{1}{n} \sum_{j=1}^{n} (z_{jk} - \bar{z}_k)(z_{jl} - \bar{z}_l) = \frac{1}{n} \sum_{j=1}^{n} \left( \frac{x_{jk} - \bar{x}_k}{\sqrt{s_{kk}}} \right) \left( \frac{x_{jl} - \bar{x}_l}{\sqrt{s_{ll}}} \right) = \frac{1}{n} \frac{1}{\sqrt{s_{kk}s_{ll}}} \sum_{j=1}^{n} (x_{jk} - \bar{x}_k)(x_{jl} - \bar{x}_l) = r_{kl}.
\]
3.14 Consider the data matrix $X$ in Exercise 3.1. We have $n=3$ observations on $p=2$ variables $X_1$ and $X_2$. Form the linear combinations:

$$
c'X = [-1 \ 2] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = -X_1 + 2X_2, \quad b'X = [2 \ 3] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = 2X_1 + 3X_2.
$$

(a) Evaluate the sample means, variances, and covariance of $b'X$ and $c'X$ from first principles. That is, calculate the observed values of $b'X$ and $c'X$, and then use the sample mean, variances and covariance formulas.

(b) Calculate the sample means, variance, and covariance of $b'X$ and $c'X$ using (3-36). Compare the results.

Sln: (a) $X = \begin{bmatrix} 5 \\ 3 \\ 1 \end{bmatrix}$, $c'X = -X_1 + 2X_2 = \begin{bmatrix} -7 \\ 1 \\ 3 \end{bmatrix}$, $b'X = 2X_1 + 3X_2 = \begin{bmatrix} 21 \\ 8 \end{bmatrix}$.

$$
\bar{c'X} = \frac{-7+1+3}{3} = -1; \quad \bar{b'X} = \frac{21+19+8}{3} = 16.
$$

$$
Var(c'X) = \frac{(-7+1)^2 + (1+1)^2 + (3+1)^2}{3-1} = \frac{64+4+16}{2} = 28.
$$

$$
Var(b'X) = \frac{(21-16)^2 + (19-16)^2 + (8-16)^2}{3-1} = \frac{25+9+64}{2} = 49.
$$

$$
Cov(c'X, b'X) = \frac{(-7+1)(21-16) + (1+1)(19-16) + (3+1)(8-16)}{3-1} = -28.
$$

(b) The covariance matrix of $X$ is $S = \begin{bmatrix} 16 & -2 \\ -2 & 1 \end{bmatrix}$, mean is $\bar{X} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$. Therefore,

$$
\bar{c'X} = c'\bar{X} = [-1 \ 2] \begin{bmatrix} 5 \\ 2 \end{bmatrix} = -1; b'\bar{X} = b'\bar{X} = [2 \ 3] \begin{bmatrix} 5 \\ 2 \end{bmatrix} = 16.
$$

$$
Var(c'X) = c'Sc = [-1 \ 2] \begin{bmatrix} 16 & -2 \\ -2 & 1 \end{bmatrix} [-1 \\ -2] = 28; Var(b'X) = b'Sb = [2 \ 3] \begin{bmatrix} 16 & -2 \\ -2 & 1 \end{bmatrix} [2 \\ 3] = 49.
$$

$$
Cov(c'X, b'X) = c'Xb = [-1 \ 2] \begin{bmatrix} 16 & -2 \\ -2 & 1 \end{bmatrix} [2 \\ 3] = -28.
$$
3.15 Repeat Exercise 3.14 using the data matrix \( X = \begin{bmatrix} 1 & 4 & 3 \\ 6 & 2 & 6 \\ 8 & 3 & 3 \end{bmatrix} \) and linear combinations \( b'X = \begin{bmatrix} 1 & 1 & 1 \\ X_1 \\ X_2 \\ X_3 \end{bmatrix} \) and \( c'X = \begin{bmatrix} 1 & 2 & -3 \\ X_1 \\ X_2 \\ X_3 \end{bmatrix} \).

Sln: The covariance matrix of \( X \) is \( S = \begin{bmatrix} 13 & \frac{3}{2} & \frac{3}{2} \\ \frac{3}{2} & 1 & \frac{3}{2} \\ \frac{3}{2} & \frac{3}{2} & 3 \end{bmatrix} \), mean is \( \mu = \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix} \). Therefore,

\[
\begin{align*}
\bar{c'}X &= c'\mu = \begin{bmatrix} 5 & 3 & 4 \end{bmatrix} = 12; \\
\bar{b'}X &= b'\mu = \begin{bmatrix} 5 & 3 \end{bmatrix} = -1.
\end{align*}
\]

\( \text{Var}(c'X) = c'Sc = \begin{bmatrix} 13 & \frac{3}{2} & \frac{3}{2} \\ \frac{3}{2} & 1 & \frac{3}{2} \\ \frac{3}{2} & \frac{3}{2} & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 12; \)

\( \text{Var}(b'X) = b'Sb = \begin{bmatrix} 13 & \frac{3}{2} & \frac{3}{2} \\ \frac{3}{2} & 1 & \frac{3}{2} \\ \frac{3}{2} & \frac{3}{2} & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = 43. \)

\( \text{Cov}(c'X, b'X) = c'Xb = \begin{bmatrix} 13 & \frac{3}{2} & \frac{3}{2} \\ \frac{3}{2} & 1 & \frac{3}{2} \\ \frac{3}{2} & \frac{3}{2} & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = -3. \)