Q1. Six observations on three variables, X_1, X_2, and X_3, are measured as follows:

Observation Number	X_1	X_2	X_3
1	2	3	8
2	5	2	4
3	4	2	7
4	8	0	5
5	5	4	1
6	6	1	5

(a) Find the sample mean vector $\bar{x}$ and the sample covariance matrix **S**.

(b) Calculate a sample correlation matrix **R**.

(c) Let $Z_1 = 2X_1 + X_2 - X_3$. Using a linear combination, calculate $E(Z_1)$ and $Var(Z_1)$.

(d) Let $Z_2 = X_1 + 2X_2 - 3X_3$. Using a linear combination, calculate $E(Z_2)$ and $Var(Z_2)$.

(e) Let $Z_3' = (-X_1, 2X_3)$. Using a linear combination, calculate $E(Z_3)$ and $Cov(Z_3)$. (Hint: For a $p\times1$ random vector X and $p\times1$ scalar vectors a and b, $Cov(a'X, b'X) = a'Cov(X)b$.)

(f) Let $Z' = (Z_1, Z_2, Z_3)$. Calculate $E(Z)$ and $Cov(Z)$.

Q2. Data (RADIOTHERAPY.DAT attached with this file) measures average ratings over the course of treatment for cancer patients undergoing radiotherapy. The variables in the data set indicate

X_1	Number of symptoms
X_2	Amount of activity (1–5 scale)
X_3	Amount of sleep (1–5 scale)
X_4	Amount of food consumed (1–3 scale)
X_5	Appetite (1–5 scale)
X_6	Skin reaction (0–3 scale)

(a) Is there any error in the data?

(b) Find the sample mean vector $\bar{x}$ and the sample covariance matrix **S**.

(c) Derive a sample correlation matrix **R**.

(d) Investigate whether each variable follows a univariate normal distribution using probability plotting techniques.

(e) For the first three patients, calculate generalized distances.

(f) Investigate whether this data set follows a multivariate normal distribution using probability plotting techniques.

(f) In (d), suggest an appropriate transformation for a variable that does not follow a univariate normal distribution. For the transformed variable, check whether that variable follows a univariate normal distribution. (Hint: Consider a Box-Cox transformation.)