MTL-766 : Multivariate Statistical Analysis- Practice Sheet 1

1. The variance covariance matrix of a 3-dimensional random vector $(X_1, X_2, , X_3)$ is

$$\begin{pmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{pmatrix}$$

(a) Find the correlation matrix. (b) Find the correlation between X_1 and $0.5(X_2 + X_3)$.

2. Suppose the random vector $\underset{\sim}{X}$ is such that $E(\underset{\sim}{X}) = \underset{\sim}{\mu}$ and $Cov(\underset{\sim}{X}) = \Sigma$. Find $E(\underset{\sim}{X}\underset{\sim}{X}^T)$. Let $\underset{\sim}{Y}$ be another random vector with $E(\underset{\sim}{Y}) = \underset{\sim}{\delta}$ and $Cov(\underset{\sim}{X}, \underset{\sim}{Y}) = \Sigma_{XY}$. Derive $E(\underset{\sim}{Y}\underset{\sim}{X}^T)$.

3. The Indian companies yield the following data

Company	Sales (x_1)	Profits (x_2)	Assets (x_3)
C1	7269	422	5733
C2	9693	383	6608
C3	8665	351	8322
C4	6343	375	7773

Compute $\bar{\underset{\sim}{x}}$ and S for (x_1, x_2, x_3). Use the distance measure $d(\underset{\sim}{x}, \underset{\sim}{y}) = \sqrt{(\underset{\sim}{x} - \underset{\sim}{y})^T S^{-1} (\underset{\sim}{x} - \underset{\sim}{y})}$ to compute the company that is nearest to mean vector $\bar{\underset{\sim}{x}}$.

4. Show that the sample covariance matrix S of data on p variables is a semi definite matrix. Prove that S is positive definite unless observations on one of the variables is a linear function of observations on the remaining $p-1$ variables.

5. Two different visual stimuli S_1 and S_2 produced responses in both the left eye (L) and right eye (R) of subjects having Multiple Sclerosis. The following is data on 3 variables viz
x_1 = age, x_2 = total response of both eyes to S_1, x_3 = total response of both eyes to S_2, for 8 subjects

Subject	x_1	x_2	x_3
1	23	148.0	205.4
2	25	195.2	262.8
3	25	158.0	209.8
4	38	190.2	243.8
5	57	165.6	229.2
6	58	238.4	304.4
7	58	164.0	216.8
8	59	199.8	250.2

(a) Suppose a distance measure of a standardized data point $P(x_1, x_2, x_3)$ from the center of standardized data in (a) is defined

$$d(O, P) = \sqrt{\underset{\sim}{x}^T A \underset{\sim}{x}} \quad , \text{ where } A = \begin{pmatrix} 4 & 2 & 0 \\ 2 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Using eigenvectors of A, find a transformation from $\underset{\sim}{x} \to \underset{\sim}{y}$ that makes transformed variables $\underset{\sim}{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$ uncorrelated.

(b) What is the transformed distance of a point $\underset{\sim}{y}$ from the center?

(c) Find the three principal axis of the largest hyper-ellipsoid that covers all standardized data points.

6. Given the data matrix

$$X = \begin{bmatrix} 3 & 4 \\ 6 & -1 \\ 3 & 4 \\ -4 & -3 \end{bmatrix}$$

Calculate the lengths and the angle between the deviation vectors and hence find the data covariance matrix S. Compute the generalized sample variance of the data.

7. Consider data of National League teams as below.

Teams	X_1 = Player Pay	X_2 = Won-Lost %
A	3497900	.623
B	2485475	.593
C	1782875	.512
D	1725450	.500
E	1645575	.463
F	1469800	.395

Use six observations each on X_1 and X_2 to find their projections on $\underset{\sim}{1}$. Calculate the angle between deviation vectors for data on X_1 and X_2. Use this to comment on the dependency between X_1 and X_2.

8. The following data is on test scores, x_1 = score on first test, x_2 = score on second test and x_3 = total score on the two tests.

$$X = \begin{bmatrix} 12 & 17 & 29 \\ 18 & 20 & 38 \\ 14 & 16 & 30 \\ 20 & 18 & 35 \\ 16 & 19 & 35 \end{bmatrix}$$

   (a) Compute S and verify that the generalized sample variance zero. Find normalized eigen vector $\underset{\sim}{e}$ corresponding to the zero eigenvalue of S.

   (b) Use eigenvalue $\underset{\sim}{e}$ to demonstrate the linear dependence of columns of mean corrected data matrix.

9. Prove the following properties for the square root $A^{1/2}$ of a symmetric positive definite matrix A of order k.

   (a) $A^{1/2}$ is a symmetric matrix.

   (b) $A^{1/2}A^{1/2} = A$

   (c) $(A^{1/2})^{-1} = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}} \underset{\sim}{e}_i \underset{\sim}{e}_i^T = P\Lambda^{-1/2}P'$ (denoted by$A^{-1/2}$) where $P = [\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots, \underset{\sim}{e}_k]$, $\Lambda^{-1/2} = diag(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \ldots, \frac{1}{\sqrt{\lambda_k}})$ and $(\lambda_i, \underset{\sim}{e}_i)$, $i = 1, 2, \ldots, k$ are eigenvalues, normalized eigenvector pairs of A.

10. Let $X^T = (X_1, X_2, X_3)$ be a random vector with covariance matrix Σ. If X_1 and X_2 are independent, find covariance matrix for $Z^T = (Z_1, Z_2, Z_3, Z_4)$ where $Z_1 = X_1 - 2X_2$, $Z_2 = X_1 + X_2 + X_3$, $Z_3 = X_1 + 2X_2 - X_3$ and $Z_4 = 3X_1 - 4X_2$.

11. Show that $cov(a_1X_1 + \ldots + a_pX_p, b_1X_1 + \ldots + b_pX_p) = \underset{\sim}{a}^T\Sigma\underset{\sim}{b}$, where $\underset{\sim}{a}^T = (a_1, \ldots, a_p)$, $\underset{\sim}{b}^T = (b_1, \ldots, b_p)$ and Σ is the covariance matrix of $\underset{\sim}{X}^T = (X_1, \ldots, X_p)$.

12. Let $\underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{X}_1 \\ ------ \\ \underset{\sim}{X}_2 \end{pmatrix}$, where $\underset{\sim}{X}^{(1)} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $\underset{\sim}{X}^{(2)} = \begin{pmatrix} X_3 \\ X_4 \\ X_5 \end{pmatrix}$. Let $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \end{pmatrix}$. If $\underset{\sim}{X}$ has mean $\mu_X^T = [2, 4, -1, 3, 0]$, and $cov(\underset{\sim}{X}) = \begin{pmatrix} 4 & -1 & 1/2 & -1/2 & 0 \\ -1 & 3 & 1 & -1 & 0 \\ 1/2 & 1 & 6 & 1 & 1 \\ -1/2 & -1 & 1 & 4 & 0 \\ 0 & 0 & -1 & 0 & 2 \end{pmatrix}$, then find

   (a) $E(\underset{\sim}{X}^{(1)})$, $E(A\underset{\sim}{X}^{(1)})$, $Cov(\underset{\sim}{X}^{(1)})$, $Cov(A\underset{\sim}{X}^{(1)})$

   (b) $Cov(\underset{\sim}{X}^{(1)}, \underset{\sim}{X}^{(2)})$

   (c) $Cov(A\underset{\sim}{X}^{(1)}, B\underset{\sim}{X}^{(2)})$, and the covariance matrix of $\begin{pmatrix} A\underset{\sim}{X}^{(1)} \\ ------ \\ B\underset{\sim}{X}^{(1)} \end{pmatrix}$