Problems

1. (i) Three variables were measured at 6 different locations from a region. The data are given by the next table:

No.	y_1	y_2	y_3
1	33	4.5	2.7
2	20	2.9	2.8
3	35	10.1	3.3
4	10	2.3	3.2
5	30	1.5	2.9
6	35	8.0	4.4

   Calculate the sample covariance matrix and the sample correlation matrix corresponding to these data.

   (ii) Which is, in the multivariate setting, the statistical distance of the p–dimensional point $\mathbf{y}$ to $\boldsymbol{\mu}$? What means that the statistical distance of $\mathbf{y}$ to $\boldsymbol{\mu}$ is constant if the sample covariance matrix has the particular form $\mathbf{S} = \begin{pmatrix} 2 & 2 \\ 2 & 3 \end{pmatrix}$?

2. Three psychological tests were given to 30 men and 30 women. Using the date obtained, the mean vectors and covariances matrices are

$$\bar{\mathbf{y}}_1 = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}, \quad \bar{\mathbf{y}}_2 = \begin{pmatrix} 2 \\ 4 \\ 3 \end{pmatrix}$$

   and

$$\mathbf{S}_1 = \begin{pmatrix} 5 & -1 & -3 \\ -1 & 2 & 2 \\ -3 & 2 & 1 \end{pmatrix}, \quad \mathbf{S}_2 = \begin{pmatrix} 2 & 3 & 2 \\ 3 & 1 & 5 \\ 2 & 5 & 1 \end{pmatrix}.$$

   Test the hypothesis that the two mean vectors are equal with a significance level of 0.01.

   Also test if the first variable is redundant with respect to the last two variables.

3. We assume that we have the following independent samples:

Population	y_1	y_2
1	6	8
1	4	6
1	2	6
2	3	8
2	-4	3
3	-3	2
3	-4	-5
3	3	-3

   Use the MANOVA unbalanced model in order to test if the population means differ with a significance level of 0.05.

4. Determine the least square estimates of the parameters in the bivariate straight line regression model

$$\begin{aligned} y_{i1} &= \beta_{01} + \beta_{11} x_{i1} + \epsilon_{i1} \\ y_{i2} &= \beta_{02} + \beta_{12} x_{i1} + \epsilon_{i2}\,, \quad i = \overline{1,5}. \end{aligned}$$

corresponding to the following data

x_1	-2	-1	0	1	2
y_1	5	3	4	2	1
y_2	-3	-1	-1	2	3

Write the regression lines. Compute the matrices of fitted values $\hat{\mathbf{Y}}$ and residuals $\hat{\mathbf{\Xi}}$.

Provide the meaning of $\hat{\mathbf{B}}$ and the geometrical interpretation of the least squares techniques. Make the connection with the geometrical interpretation of the first principal component. Also compute $\mathbf{X}'\hat{\mathbf{\Xi}}$ and $\hat{\mathbf{Y}}'\hat{\mathbf{\Xi}}$.

5. (i) We assume that we have n measurements of 2–dimensional vectors $\mathbf{y}$. We obtain the vector mean $\bar{\mathbf{y}}' = \begin{pmatrix} 1 & 2 \end{pmatrix}$ and the sample covariance matrix $\mathbf{S} = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$. Find the principal components and provide the geometric and the algebraic interpretation of the first and the second principal components. Also find the covariance matrix S_z.

   Comment the results obtained.

   (ii) If the covariance matrix has the particular form $\mathbf{S} = \begin{pmatrix} 2 & 2r & 2r \\ 2r & 2 & 2r \\ 2r & 2r & 2 \end{pmatrix}$, then determine the correlation matrix $\mathbf{R}$. Find the principal components associated to $\mathbf{R}$.

   Comment the results obtained.