Use the data in Table 1 below for Questions 1, 2 and 3.

	Features	
Point	X_1	X_2
p_1	2.70	2.50
p_2	0.75	1.00
p_3	2.25	4.00
p_4	1.90	2.25
p_5	4.00	2.50
p_6	2.30	2.70

Table 1: Two features with six points for question

1. **PCA**. Use the data in Table 1 manually perform calculation (with the help of a calculator or spreadsheet) for (a) - (d).
   (a) Please construct a zero-mean 2×2 variance-covariance matrix, Σ_x.
   (b) Find the eigenvalues and unit eigenvectors of the matrix Σ_x. Also demonstrate that the eigenvectors are orthogonal.
   (c) What are the two principle components. And demonstrate that the variance of the first principle component and the second is the 1st and the 2nd eigenvalue respectively.
   (d) Please calculate PVE (percent of variation explained) by each principle component.
   (e) Please accomplish (b) - (d) in R, using R functions, eigen(), prcomp(), and svd().
   (f) Please compare results of eigenvalues, eigenvectors, PCs, and PVE from (b) - (d) with the corresponding results from (e).
   (g) Please biplots the two PCs in R. And comments on the plots.
2. Please use the data Table 1 for Question 2 and Question 3 below.

   Hierarchical Clustering. Please manually calculate applying Hierarchical Clustering with complete linkage to produce TWO clusters for (a) - (b). (You can use calculator or spreadsheet.)
   (a) Please build the dendrogram and identify the two clusters.
   (b) Please estimate the proportion of total variation accounted for by the two groups.
   (c) Please accomplish (a) - (b) using R.
3. **K-means Clustering** You are interested to find TWO clusters using K-means clustering method.
   (a) Use K-means clustering algorithm to find the two clusters manually. Please start with initial assignment $C_1 = \{P_1, P_4, P_5\}$ and $C_2 = \{P_2, P_3, P_6\}$

(b) Using R to randomly perform K-means clustering for 2 clusters. Plot the clusters (with different color for each cluster).

4. The loading vectors of the first two principal components for a particular data set is given in the following table.

Variable j	u_{j1}	u_{j2}
1	0.5359	-0.4182
2	0.5832	-0.1880
3	0.2782	0.8728
4	0.5434	0.1673

Here are two observations of the original variables.

Variable i	x_{i1}	x_{i2}	x_{i3}	x_{i4}
1	1.2426	0.7828	-0.5209	-0.0034
2	0.5079	1.1068	-1.2118	2.4842

calculate the following:

(a) calculate the first principal component score for the first observation.

(b) calculate the second principal component score for the second observation.

(c) Approximation of x_{14} by the first two principal components.

(d) Approximation error of x_{13} by the first two principal components.

(e) The distance between the first observation and its approximation by the first two principal components.