Multivariate Analysis - Homework 7

1. We present the results of several factor analyses on bone and skull measurements of white leghorn fowl. The full data set consists of $n = 276$ measurements on bone dimensions:

$$
\text{Head:} \quad \begin{cases} X_1 = \text{skull length} \\ X_2 = \text{skull breadth} \end{cases}
$$

$$
\text{Leg:} \quad \begin{cases} X_3 = \text{femur length} \\ X_4 = \text{tibia length} \end{cases}
$$

$$
\text{Wing:} \quad \begin{cases} X_5 = \text{humerus length} \\ X_6 = \text{ulna length} \end{cases}
$$

The sample correlation matrix is

$$
\mathbf{R} = \begin{bmatrix}
1.000 & .505 & .569 & .602 & .621 & .603 \\
.505 & 1.000 & .422 & .467 & .482 & .450 \\
.569 & .422 & 1.000 & .926 & .877 & .878 \\
.602 & .467 & .926 & 1.000 & .874 & .894 \\
.621 & .482 & .877 & .874 & 1.000 & .937 \\
.603 & .450 & .878 & .894 & .937 & 1.000
\end{bmatrix}
$$

(a) Obtain the estimated factor loadings by the principal component method, suppose 3 factors are used.

(b) Estimate the specific variances.

(c) Estimate the communalities.

(d) Estimate the proportion of variance explained by each factor.

(e) Compute the residual matrix and comment the result.

(f) Suppose the rotated factor loadings are as follows, give the interpretation of the factors.

Rotated estimated loadings

F_1^*	F_2^*	F_3^*
.355	.244	.902
.235	.949	.211
.921	.164	.218
.904	.212	.252
.888	.228	.283
.908	.192	.264

2. Prove that the Ward's method for hierarchical clustering is proportional to the centroid method.

3. The distances between pairs of five items are as follows:

$$
\begin{array}{c}
 & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} &
\left[\begin{array}{ccccc}
0 & & & & \\
4 & 0 & & & \\
6 & 9 & 0 & & \\
1 & 7 & 10 & 0 & \\
6 & 3 & 5 & 8 & 0
\end{array}\right]
\end{array}
$$

Cluster the five items using the single linkage, complete linkage, and average linkage. Sketch the dendrograms by hand (the relative distances in the dendrogram do not need to be exact). Compare the results.

4. Suppose we measure two variables for four items A, B, C and D. The data are as follows:

	Observations	
Item	x_1	x_2
A	5	4
B	1	-2
C	-1	1
D	3	1

Use the k-means clustering technique to divide the items into 2 clusters.

(a) Start with the initial groups (AB) and (CD).

(b) Start with the initial groups (AC) and (BD).

(c) Compare the results of (a) and (b). Are they the same? Graph the items in terms of their (x_1, x_2) coordinates, and comment on the solutions.