MSc: Statistical Analysis (Spring Term)

Multivariate Analysis: Exercises 1

1. Let $Y_1 = \mathbf{a_1}'\mathbf{X}$ and $Y_2 = \mathbf{a_2}'\mathbf{X}$ be two different linear compounds of the $p \times 1$ random vector $\mathbf{X}$. Show, from the definition of covariance, that

$$\mathrm{cov}(Y_2, Y_1) = \mathrm{cov}(\mathbf{a}_2'\mathbf{X}, \mathbf{a}_1'\mathbf{X}) = \mathbf{a}_2'\boldsymbol{\Sigma}\mathbf{a}_1.$$

   More generally, if A and B are matrices of constants of dimensions $r \times p$ and $s \times p$ respectively, then show that the corresponding covariance matrices of

$$\mathbf{Y} = A\mathbf{X} \quad \text{and} \quad \mathbf{Z} = B\mathbf{X}$$

   are given by

$$Cov(\mathbf{Y}, \mathbf{Y}) = A\boldsymbol{\Sigma}\mathbf{A}^T$$
$$Cov(\mathbf{Z}, \mathbf{Z}) = B\boldsymbol{\Sigma}B^T$$

   and

$$Cov(\mathbf{Y}, \mathbf{Z}) = A\boldsymbol{\Sigma}B^T.$$

2. Suppose $\mathbf{X} = (X_1, X_2)'$ is a 2×1 random vector on which a sample of n independent observation vectors has been drawn. Write the sample covariance matrix $\mathbf{s}$ as:

$$\mathbf{S} = \begin{pmatrix} s_1^2 & rs_1s_2 \\ rs_1s_2 & s_2^2 \end{pmatrix}$$

   where r is the sample correlation coefficient between X_1 and X_2. Show that if X_1 and X_2 are standardized so that each has sample variance 1, then the principal components of the standardized observations do not depend on s_1^2, s_2^2 or r. What proportion of the variance is explained by the first component in this case? Suppose now that X_1 and X_2 are left unstandardized and that $s_1 = 2$, $s_2 = 3$ and $r = 1/\sqrt{6}$. Find the principal components for these data and find the percentage of total variance accounted for by each component. Show that the principal components for the unstandardized variables do not correspond to those for the standardized variables and use a geometric argument to explain why.
   (Remember to compare the two sets of components in the same units.)

3. The principal component analysis shown below was carried out using the correlation matrix of six bone measurements on 276 fowls. Briefly discuss these results and give interpretations for the components where you can.

	Principal Components					
Measurements	1	2	3	4	5	6
x_1, skull length	0.35	0.53	0.76	-0.05	0.04	0.00
x_2, skull breadth	0.33	0.70	-0.64	0.00	0.00	-0.04
x_3, wing humerus	0.44	-0.19	-0.05	0.53	0.19	0.59
x_4, wing ulna	0.44	-0.25	0.02	0.48	-0.15	-0.63
x_5, leg femur	0.43	-0.28	-0.06	-0.51	0.67	-0.48
x_6, leg tibia	0.44	-0.22	-0.05	-0.48	-0.70	0.15
Component variance	4.57	0.71	0.41	0.17	0.08	0.06

4. The table below shows the marks obtained (out of 20) for three assignments taken by a Chemistry class of 10 students.

Assignment 1	Assignment 2	Assignment 3
8	14	7
12	13	13
14	11	8
12	13	10
9	10	12
10	12	11
11	10	10
11	15	12
12	13	10
14	10	9

Each line represents the marks obtained by each student; the 10 sets of marks are assumed to constitute a random sample.

Let X_1, X_2 and X_3 be random variables denoting the marks for assignments 1,2, and 3 respectively for a student in the class.

(a) Based on the above data, find the sample covariance and sample correlation matrices for $\mathbf{X} = (X_1, X_2, X_3)'$.

(b) Find the principal components of $\mathbf{X}$ based on:

(i) the sample covariance matrix

(ii) the sample correlation matrix.