MSc: Statistical Analysis (Spring Term)

Multivariate Analysis: Exercises 2

1. Consider the Household Spending data of Chapter 8. Suppose we wish to determine whether there is any evidence to reject the hypothesis that the mean amounts spent on groceries, leisure, and income are £180, £113, and £750, respectively in households across the population.

   Data can be input as in Chapter 8.

   ```
   groceries <- c(227.01, 241.42, 188.08, 238.23, 235.86)
   leisure <- c(96.98, 140.44, 85.13, 158.22, 103.06)
   income <- c(741.29, 854.07, 812.07, 813.69, 731.42)
   spend <- data.frame(groceries, leisure, income)
   ```

   Now carry out an appropriate one-sample Hotelling T^2 test. What is your conclusion?

2. Sensors are placed in a symmetrical fashion across patients' heads in order to measure the amounts of electrical activity taking place. Suppose that the $p \times 1$ random vector $\mathbf{X}$, with p odd, represents the series of measurements taken, and suppose further that $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown.

   Assume that we have a random sample of size n taken from this distribution and construct a test to determine whether or not the amount of activity across the head is asymmetric or not. That is, consider the null hypothesis that

   $$\mu_1 = \mu_p,\ \mu_2 = \mu_{p-1},\ \ldots,\ \mu_{\frac{p-1}{2}} = \mu_{\frac{p+3}{2}}$$

   Give the form of the test statistic and the appropriate degrees of freedom, under the null hypothesis, in terms of p and the sample size n.

3. Consider again the artificial data given in Chapter 9 referring to a random sample of six approximately 2-year old boys from a high altitude region in Asia. The variables recorded on each member of the sample were Height (X_1), Chest Circumference (X_2), and Middle Upper Arm Circumference (X_3), measured in c.m.

Individual	Height (cm)	Chest circumference (cm)	MUAC (cm)
1	78	60.6	16.5
2	76	58.1	12.5
3	92	63.2	14.5
4	81	59.0	14.0
5	81	60.8	15.5
6	84	59.5	14.0

We wish to test whether the mean of the height, chest circumference, and MUAC of the population from which the sample was drawn, occur in the ratio 6:4:1.

Satisfy yourself that the hypothesis to be tested can be expressed as:

$$H_0 : \frac{1}{6}\mu_1 = \frac{1}{4}\mu_2 = \mu_3$$

or, alternatively, in the form $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ where

$$\mathbf{C} = \begin{bmatrix} 2 & -3 & 0 \\ 1 & 0 & -6 \end{bmatrix}$$

This fits the general framework for testing,

$$H_0 : \mathbf{C}\boldsymbol{\mu} = \boldsymbol{\phi} \text{ vs. } H_1 : \mathbf{C}\boldsymbol{\mu} \neq \boldsymbol{\phi}$$

where $\mathbf{C}$ is an $m \times p$ matrix of constants, $\text{rank}(\mathbf{C}) = m < n$, and $\boldsymbol{\phi}$ is a $m \times 1$ vector of constants.

Then the relevant test statistic is given by

$$T^2 = n(\mathbf{C}\overline{\mathbf{X}} - \boldsymbol{\phi})^T(\mathbf{C}\mathbf{S}_U\mathbf{C}^T)^{-1}(\mathbf{C}\overline{\mathbf{X}} - \boldsymbol{\phi}).$$

where, under H_0, $T^2 \sim T^2_m(n-1)$.

Carry out the above test for these data.

4. The following table is for 2-year-old girls and corresponds to the data in Exercise 3 for boys.

Individual	Height (cm)	Chest circumference (cm)	MUAC (cm)
1	80	58.4	14.0
2	75	59.2	15.0
3	78	60.3	15.0
4	75	57.4	13.0
5	79	59.5	14.0
6	78	58.1	14.5
7	75	58.0	12.5
8	64	55.5	11.0
9	80	59.2	12.5

We wish to test the hypothesis that there is no difference in the mean vectors for 2-year-old boys and girls. That is,

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \text{ vs. } \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2.$$

You will need to use a two-sample Hotelling T^2-test. In general this takes the form,

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2)^T \mathbf{S}_U^{-1} (\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2).$$

Under H_0, $T^2 \sim T_p^2(n_1 + n_2 - 2)$.

Note: In this case $\mathbf{S}_U$ is the pooled *within-groups* maximum likelihood sample covariance matrix given by

$$\mathbf{S}_U = \frac{(n_1 - 1)\mathbf{S}_{U1} + (n_2 - 1)\mathbf{S}_{U2}}{n_1 + n_2 - 2}.$$