MSc: Statistical Analysis (Spring Term)

Multivariate Analysis: Exercises 2 - Solutions

1.
```
> 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)
> S <- var(spend)
> S.inv <- solve(S)
> n <- dim(spend)[1]
> p <- dim(spend)[2]
> n
[1] 5
> p
[1] 3

> m.spend <- apply(spend, 2, mean)
> m.spend
 groceries leisure  income
    226.12 116.766 790.508
> mu0 <- c(180, 113, 750)
> mu0
[1] 180 113 750
> T2 <- n * t(m.spend - mu0) %*% S.inv %*% (m.spend - mu0)
> T2
         [,1]
[1,] 126.1583
> T2 <- drop(T2)
> T2
[1] 126.1583
> F.obs <- ((n - p) * T2)/(p * (n - 1))
> F.obs
[1] 21.02638
> p.value <- 1 - pf(F.obs, p, n - p)
> p.value
[1] 0.04574171
```

Conclusion: There is evidence to reject the null hypothesis $\boldsymbol{\mu} = (180, 113, 750)^T$ at the 5% level of significance.

2. In essence we ask whether the mean amount of activity, in the population under consideration, is symmetrical or not.

   Model the readings of the i-th member of a random sample of size n as $\mathbf{X}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i=1,\ldots,n$, where p is odd, and $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ are both unknown.

   Symmetry can be posed as

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

   i.e.

$$\begin{array}{ccccc} \mu_1 & - & \mu_p & = & 0 \\ \mu_2 & - & \mu_{p-1} & = & 0 \\ & \vdots & & & \vdots \\ \mu_{\frac{p-1}{2}} & - & \mu_{\frac{p+3}{2}} & = & 0 \end{array}$$

   or $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$,

   where

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & \ldots & \ldots & 0 & 0 & 0 & \ldots & \ldots & 0 & 0 & -1 \\ 0 & 1 & 0 & \ldots & \ldots & 0 & 0 & 0 & \ldots & \ldots & 0 & -1 & 0 \\ 0 & 0 & 1 & \ldots & \ldots & 0 & 0 & 0 & \ldots & \ldots & -1 & 0 & 0 \\ 0 & 0 & 0 & \ldots & \ldots & 0 & 0 & 0 & \ldots & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \ldots & 1 & 0 & -1 & \ldots & \ldots & 0 & 0 & 0 \end{bmatrix}$$

   which is a $\frac{p-1}{2} \times p$ matrix. Since

$$\mathbf{C}\overline{\mathbf{X}} \ \sim\ \mathrm{N}_{(p-1)/2}\left(\mathbf{C}\boldsymbol{\mu}, \frac{1}{n}\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T\right)$$

   and, independently

$$(n-1)\mathbf{C}\mathbf{S}_U\mathbf{C}^T \ \sim\ W_{(p-1)/2}(n-1, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T),$$

   then, under H_0,

$$T^2 \ =\ n(\mathbf{C}\overline{\mathbf{X}})^T(\mathbf{CSC}^T)^{-1}(\mathbf{C}\overline{\mathbf{X}}) \ =\ n\,\overline{\mathbf{X}}^T\mathbf{C}^T(\mathbf{CSC}^T)^{-1}\mathbf{C}\overline{\mathbf{X}} \ \sim\ T^2_{\frac{p-1}{2}}(n-1).$$

   Thus, we reject the hypothesis of symmetry at the $100\alpha\%$ level of significance if

$$\frac{n-(p-1)/2}{(p-1)(n-1)/2}\,T^2_{obs} \ >\ F_{(p-1)/2\,,\,n-(p-1)/2}\,(\alpha)$$

   and accept H_0 otherwise.

3. Testing the given structural relationship we have:

```
> b.height <- c(78, 76, 92, 81, 81, 84)
> b.chest <- c(60.6, 58.1, 63.2, 59, 60.8, 59.5)
> b.muac <- c(16.5, 12.5, 14.5, 14, 15.5, 14)
> boys <- data.frame(b.height, b.chest, b.muac)

> m.boys <- apply(boys, 2, mean)
> S <- var(boys)

> C <- matrix(c(2, 1, -3, 0, 0, -6), 2, 3)
> C
     [,1] [,2] [,3]
[1,]    2   -3    0
[2,]    1    0   -6

> Cm.boys <- C %*% m.boys
> Cm.boys
      [,1]
[1,] -16.6
[2,]  -5.0

> CSC <- C %*% S %*% t(C)
> CSC
      [,1]  [,2]
[1,] 58.47 56.66
[2,] 56.66 94.00

> T2 <- n * t(Cm.boys) %*% solve(CSC) %*% Cm.boys
> drop(T2)
[1] 47.14

> p <- 3
> n <- 6
> q <- p - 1
> df <- n - 1

> F.obs <- ((df - q + 1)/(df * q)) * T2
> drop(F.obs)
[1] 18.86

> prob <- 1 - pf(F.obs, q, df - q + 1)
> prob
[1] 0.009195
```

Clearly we reject the null hypothesis that the means of height, chest circumference, and MUAC in the population from which the sample of two-year old boys was drawn, occur in the ratio 6:4:1.

Note: Alternatively, the matrix **C** may be defined using `rbind` as in:

```
> c1 <- c(2, -3, 0)
> c2 <- c(1, 0, -6)
> C <- rbind(c1, c2)
> C
   [,1] [,2] [,3]
c1    2   -3    0
c2    1    0   -6
```

Try choosing a different matrix **C** that tests the same structural relationship. Check that the calculated T^2 statistic is unchanged.

4. (Assuming we still have the *objects* relating to the boys' sample from question 3.)

```
> n1 <- 6
> S1 <- S

> g.height <- c(80, 75, 78, 75, 79, 78, 75, 64, 80)
> g.chest <- c(58.4, 59.2, 60.3, 57.4, 59.5, 58.1, 58, 55.5, 59.2)
> g.muac <- c(14, 15, 15, 13, 14, 14.5, 12.5, 11, 12.5)
> girls <- data.frame(g.height, g.chest, g.muac)

> m.girls <- apply(girls, 2, mean)
> S2 <- var(girls)
> n2 <- 9

> m.boys
 b.height b.chest b.muac
       82    60.2   14.5
> m.girls
 g.height g.chest g.muac
       76    58.4   13.5

> S1
         b.height b.chest b.muac
b.height    31.60   8.040   0.50
 b.chest     8.04   3.172   1.31
  b.muac     0.50   1.310   1.90

> S2
         g.height g.chest g.muac
g.height   24.500   5.638  4.313
 g.chest    5.638   1.970  1.456
  g.muac    4.313   1.456  1.813

> m.boys - m.girls
 b.height b.chest b.muac
        6     1.8      1

> Sp <- ((n1 - 1) * S1 + (n2 - 1) * S2)/(n1 + n2 - 2)
> Sp
         b.height b.chest b.muac
b.height   27.231   6.562  2.846
 b.chest    6.562   2.432  1.400
  b.muac    2.846   1.400  1.846

> Sp.inv <- solve(Sp)

> T2 <- ((n1 * n2)/(n1 + n2)) * t(m.boys - m.girls) %*% Sp.inv %*% (m.boys - m.girls)
> drop(T2)
[1] 5.312

> df <- n1 + n2 - 2
> Fobs <- ((df - p + 1)/(df * p)) * T2
> drop(Fobs)
[1] 1.498

> prob <- 1 - pf(Fobs, p, df - p + 1)
> prob
[1] 0.2693
```

There is no evidence to reject the null hypothesis that there is no difference in the mean vectors for 2-year-old boys and girls.