Revision Problems

1.

X1	3	4	2	6	8	2	5
X2	5	5.5	4	7	10	5	7.5

Calculate the sample means, sample variances and sample covariances

2.

Verify the following properties of the transpose when

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 4 & 2 \\ 5 & 0 & 3 \end{bmatrix}, \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix}$$

(a) $(\mathbf{A}')' = \mathbf{A}$

(b) $(\mathbf{C}')^{-1} = (\mathbf{C}^{-1})'$

(c) $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$

3.

Calculate the generalised sample variance for

$$\mathbf{X} = \begin{bmatrix} 9 & 1 \\ 5 & 3 \\ 1 & 2 \end{bmatrix}$$

4

Let $\mathbf{X}$ be distributed as $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}' = [1, -1, 2]$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

Which of the following random variables are independent? Explain.

(a) X_1 and X_2

(b) X_1 and X_3

(c) X_2 and X_3

(d) (X_1, X_3) and X_2

(e) X_1 and $X_1 + 3X_2 - 2X_3$

5

(a) Evaluate T^2, for testing H_0: $\boldsymbol{\mu}' = [7, \quad 11]$, using the data

$$\mathbf{X} = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix}$$

(b) Specify the distribution of T^2 for the situation in (a).

(c) Using (a) and (b), test H_0 at the $\alpha = .05$ level. What conclusion do you reach?

6

Determine the population principal components Y_1 and Y_2 for the covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$$

Also, calculate the proportion of the total population variance explained by the first principal component.

7

$$\boldsymbol{\rho} = \begin{bmatrix} 1.0 & .63 & .45 \\ .63 & 1.0 & .35 \\ .45 & .35 & 1.0 \end{bmatrix}$$

for the $p = 3$ standardized random variables Z_1, Z_2, and Z_3 can be generated by the $m = 1$ factor model

$$Z_1 = .9F_1 + \varepsilon_1$$
$$Z_2 = .7F_1 + \varepsilon_2$$
$$Z_3 = .5F_1 + \varepsilon_3$$

where $\mathrm{Var}(F_1) = 1$, $\mathrm{Cov}(\boldsymbol{\varepsilon}, F_1) = \mathbf{0}$, and

$$\boldsymbol{\Psi} = \mathrm{Cov}(\boldsymbol{\varepsilon}) = \begin{bmatrix} .19 & 0 & 0 \\ 0 & .51 & 0 \\ 0 & 0 & .75 \end{bmatrix}$$

(a) Calculate communalities h_i^2, $i = 1, 2, 3$, and interpret these quantities.

(b) Calculate $\mathrm{Corr}(Z_i, F_1)$ for $i = 1, 2, 3$. Which variable might carry the greatest weight in "naming" the common factor? Why?

8

Consider the matrix of distances

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ 11 & 2 & 0 & \\ 5 & 3 & 4 & 0 \end{bmatrix} \end{array}$$

Cluster the four items using each of the following procedures.

(a) Single linkage hierarchical procedure.

(b) Complete linkage hierarchical procedure.

(c) Average linkage hierarchical procedure.

Draw the dendrograms and compare the results in (a), (b), and (c).