Introductory Multivariate Methods
Stat 3133

Lecture Notes

Awol S.
Department of Statistics
College of Computing & Informatics
Haramaya University
Dire Dawa, Ethiopia

Contents

Chapter 1

Review of Matrix Algebra

The study of multivariate methods is greatly facilitated by the use of matrix algebra. This chapter presents a review of basic concepts of matrix algebra which are essential to both geometrical interpretations and algebraic explanations of subsequent multivariate statistical techniques.

1.1 Definition of Matrix and Vector

A rectangular array of numbers with, for instance, n rows and p columns is called a matrix of dimension $n \times p$. It is written as:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

A vector is a matrix of $n \times 1$ real numbers $x_1, x_2, \dots, x_n$ and it is written as:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ or } \boldsymbol{x}' = (x_1, x_2, \cdots, x_n) \text{ or } \boldsymbol{x} = (x_1, x_2, \cdots, x_n)'.$$

A vector has both magnitude (length) and direction. The length of a vector, $\boldsymbol{x}' = (x_1, x_2, \cdots, x_n)$, is defined by

$$L_{\boldsymbol{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\boldsymbol{x}'\boldsymbol{x}}.$$

The length of a vector can be expanded and contracted by multiplying with a constant a. That is,

$$a\boldsymbol{x} = \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{bmatrix}.$$

Such multiplication of a vector $\boldsymbol{x}$ by a scalar a changes the length as

$$L_{a\boldsymbol{x}} = \sqrt{a^2x_1^2 + a^2x_2^2 + \cdots + a^2x_n^2} = |a|\sqrt{\boldsymbol{x}'\boldsymbol{x}}.$$

When $|a| > 1$, vector $\boldsymbol{x}$ is expanded. When $|a| < 1$, vector $\boldsymbol{x}$ is contracted. When $|a| = 1$, there is no change. If $a < 0$, the direction of vector $\boldsymbol{x}$ is changed.

Choosing $a = L_{\boldsymbol{x}}^{-1}$, we obtain the unit vector $L_{a\boldsymbol{x}}$, which has length 1 and lies in the direction of $\boldsymbol{x}$.

Example 1.1. If $n = 2$, consider the vector $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. The length of $\boldsymbol{x}$ is $L_{\boldsymbol{x}} = \sqrt{x_1^2 + x_2^2}$. Geometrically, the length of a vector in two dimensions can be viewed as the hypotenuse of a right triangle.

1.1.1 Matrix Characteristics

- **Rank**: The rank of a matrix $\boldsymbol{A}$ is the maximum number of linearly independent rows (columns).
  - A set of k vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k$ is said to be linearly independent if $a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_3 + \cdots + a_k\boldsymbol{x}_k = \sum_{i=1}^{k} a_i\boldsymbol{x}_i = 0$ only if $a_1 = a_2 = \cdots = a_k = 0$. That is, if every a_i is zero, the $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k$ (columns) are linearly independent. Linear independence implies every vector can not be written as a linear combination of the other vectors. Vectors of the same dimension that are not linearly independent are said to be linearly dependent which means at least one vector can be written as a linear combination of the other vectors.

    Example 1.2. $\boldsymbol{x}_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, $\boldsymbol{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

    $a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 = 0 \Rightarrow$

    $$3a_1 + 2a_2 = 0$$
    $$4a_1 + a_2 = 0$$

    holds only if $a_1 = a_2 = 0$. This confirms that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are linearly independent. In other words, the columns of matrix $\boldsymbol{A} = \begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix}$ are linearly independent.

    Example 1.3. $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$, $\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 5 \\ 1 \end{bmatrix}$, $\boldsymbol{x}_3 = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$

    $a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + a_3\boldsymbol{x}_3 = 0 \Rightarrow$

    $$a_1 + a_2 + a_3 = 0$$
    $$2a_1 + 5a_2 - a_3 = 0$$
    $$a_2 - a_3 = 0$$

    $\Rightarrow a_1 + 2a_2 = 0$. $\Rightarrow$ If $a_1 = a_2 = 0$, then $a_3 = 0$. If $a_1 = 1$, then $a_2 = a_3 = 0.5$. Therefore, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ are not linearly independent.

  - The row and column rank of a matrix are equal.
    * Rank $(\boldsymbol{A}) \geq 0$
    * Rank $(\boldsymbol{A}) \leq \min(n, p)$
    * Rank $(\boldsymbol{A})$ = Rank $(\boldsymbol{A}')$
    * Rank $(\boldsymbol{A})$ = Rank $(\boldsymbol{A}'\boldsymbol{A})$ = Rank $(\boldsymbol{A}\boldsymbol{A}')$

- **Trace**: The trace of a matrix is the sum of its diagonal elements: $\text{tr}(\boldsymbol{A}) = \sum_{i=1}^{k} a_{ii}$.
  - $\text{tr}(\boldsymbol{A} \pm \boldsymbol{B}) = \text{tr}(\boldsymbol{A}) \pm \text{tr}\boldsymbol{B})$
  - $\text{tr}(c\boldsymbol{A}) = c\ \text{tr}(\boldsymbol{A})$
  - $\text{tr}(\boldsymbol{A}_{n\times p}\boldsymbol{B}_{p\times n}) = \text{tr}(\boldsymbol{B}\boldsymbol{A})$
  - $\text{tr}(\boldsymbol{A}_{n\times p}\boldsymbol{B}_{p\times q}\boldsymbol{C}_{q\times n}) = \text{tr}(\boldsymbol{C}\boldsymbol{A}\boldsymbol{B}) = \text{tr}(\boldsymbol{B}\boldsymbol{C}\boldsymbol{A})$

- **Determinant**: Det $(\mathbf{A}) = |\boldsymbol{A}|$
  - $|a\boldsymbol{A}| = a^n|\boldsymbol{A}|$
  - $|\boldsymbol{A}\boldsymbol{B}| = |\boldsymbol{B}\boldsymbol{A}| = |\boldsymbol{A}||\boldsymbol{B}|$

- **Inverse**: A square matrix $\boldsymbol{A}$ is said to be non-singular if its rank is equal to the number of rows (columns).
  - If a $k \times k$ matrix $\boldsymbol{A}$ is non-singular, then there exist a unique $k \times k$ matrix $\boldsymbol{B}$ such that $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A} = \boldsymbol{I}_{k\times k}$.
    * The matrix $\boldsymbol{B}$ is called inverse of $\boldsymbol{A}$ denoted by $\boldsymbol{A}^{-1}$.
    * $\boldsymbol{A}^{-1}$ exists if and only if the determinant of $\boldsymbol{A}$ is non-zero. And hence, $|\boldsymbol{A}^{-1}| = |\boldsymbol{A}|^{-1}$.

- **Positive Definite Matrix**: A symmetric matrix $\boldsymbol{A}$ is said to be positive definite if the quadratic form $Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$ where $\boldsymbol{x}' = (x_1, x_2, \cdots, x_n)$.

  Example 1.4. $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

$$Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} 2, -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 0$$

  $\Rightarrow \boldsymbol{A}$ is not positive definite.

  - A symmetric matrix $\boldsymbol{A}$ is said to be positive semi-definite if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \neq \mathbf{0}$.

1.1.2 Eigenvalues and Eigenvectors

Let $\boldsymbol{A}$ be a $k \times k$ matrix and $\boldsymbol{I}$ be a $k \times k$ identity matrix. The scalars $\lambda_1, \lambda_2, \cdots, \lambda_k$ satisfying the polynomial equation:

$$|\boldsymbol{A} - \lambda\boldsymbol{I}| = 0$$

are called the eigenvalues (characteristics roots) of matrix $\boldsymbol{A}$. These eigenvalues are unique unless two or more eigenvalues are equal.

- The equation $|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0$ as a function of λ is called characteristic equation.
- The eigenvalues of a symmetric matrix with real elements are real. λ_i's can be complex numbers if the matrix is not symmetric.
- The eigenvalues of a positive definite matrix are all positive. If a $k \times k$ symmetric matrix is positive semi-definite of rank r ($r < k$), then it has r positive and $(k - r)$ zero eigenvalues.
- The eigenvalues of a diagonal matrix are the diagonal elements themselves.
- The eigenvalues of an idempotent matrix $\boldsymbol{A}$, that is, $\boldsymbol{A} = \boldsymbol{A}^2$ are 1 and 0.

Associated with every eigenvalue λ_i of a square matrix $\boldsymbol{A}$, there is an eigenvector $\boldsymbol{x}_i$ whose elements satisfy the homogenous system of equations:

$$(\boldsymbol{A} - \lambda_i \boldsymbol{I})\boldsymbol{x}_i = \boldsymbol{0} \Leftrightarrow \boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$$

- If $|\boldsymbol{A} - \lambda_i \boldsymbol{I}| = 0$, there exist at least one non-trivial solution ($\boldsymbol{x}_i \neq \boldsymbol{0}$).
- The elements of the vector $\boldsymbol{x}_i$ are determined only up to a scaled factor because the system is homogenous, we get only relationship like $x_{1i} = 5x_{2i}$ because the number of unknowns is greater than the number of equations.
  - Since the values of the eigenvectors are trivial, normalizing makes them unique, that is, the eigenvectors have a unit length.
  - The normalized eigenvector, $\boldsymbol{e}_i$, of $\boldsymbol{x}_i$ is:

$$\boldsymbol{e}_i = \frac{1}{L_{\boldsymbol{x}_i}} \boldsymbol{x}_i = \frac{\boldsymbol{x}_i}{\sqrt{\boldsymbol{x}_i' \boldsymbol{x}_i}}$$

    * $||\boldsymbol{e}_i|| = \boldsymbol{e}_i' \boldsymbol{e}_i = 1$, for all i.
    * $\boldsymbol{e}_i' \boldsymbol{e}_j = \dfrac{\boldsymbol{x}_i'}{\sqrt{\boldsymbol{x}_i' \boldsymbol{x}_i}} \dfrac{\boldsymbol{x}_j}{\sqrt{\boldsymbol{x}_j' \boldsymbol{x}_j}} = 0$ for all $i \neq j$.
  - The normalized eigenvectors are chosen to satisfy $\boldsymbol{e}_1' \boldsymbol{e}_1 = \boldsymbol{e}_2' \boldsymbol{e}_2 = \cdots = \boldsymbol{e}_k' \boldsymbol{e}_k = 1$ and be mutually perpendicular, $\boldsymbol{e}_i' \boldsymbol{e}_j = 0, i \neq j$.

Example 1.5. Find the eigenvalues and eigenvectors of $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$.

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0 \Rightarrow \left| \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0 \Rightarrow \begin{vmatrix} 1 - \lambda & 2 \\ 3 & 2 - \lambda \end{vmatrix} = 0$$

$$(1 - \lambda)(2 - \lambda) - 6 = 0 \Rightarrow \lambda^2 - 3\lambda - 4 = 0$$

Thus, the eigenvalues of $\boldsymbol{A}$ are $\lambda_1 = 4$ and $\lambda_2 = -1$.

To find the corresponding eigenvectors:

$$\boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i, \; i = 1, 2$$

- For $\lambda_1 = 4$,

$$\boldsymbol{A}\boldsymbol{x}_1 = \lambda_1\boldsymbol{x}_1 \Rightarrow \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = 4\begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix}$$

$$\Rightarrow x_{11} + 2x_{21} = 4x_{11} \Rightarrow x_{21} = \frac{3}{2}x_{11}$$

Let $x_{11} = 2 \Rightarrow x_{21} = 3$. Thus, $\boldsymbol{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ - - not unique.

The normalized eigenvector of $\boldsymbol{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ is

$$\boldsymbol{e}_1 = \frac{1}{\sqrt{\boldsymbol{x}_1'\boldsymbol{x}_1}}\boldsymbol{x}_1 = \frac{1}{\sqrt{4+9}}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{13}} \\ \frac{3}{\sqrt{13}} \end{bmatrix}.$$

Note that $\boldsymbol{e}_1'\boldsymbol{e}_1 = 1$.

- For $\lambda_2 = -1$,

$$\boldsymbol{A}\boldsymbol{x}_2 = \lambda_2\boldsymbol{x}_2 \Rightarrow \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} = -1\begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix}$$

$$\Rightarrow x_{12} + 2x_{22} = -x_{12} \Rightarrow x_{22} = -x_{12}$$

Let $x_{12} = 1 \Rightarrow x_{22} = -1$. Thus, $\boldsymbol{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ - - not unique.

The normalized eigenvector of $\boldsymbol{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is

$$\boldsymbol{e}_2 = \frac{1}{\sqrt{\boldsymbol{x}_2'\boldsymbol{x}_2}}\boldsymbol{x}_2 = \frac{1}{\sqrt{1+1}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Note that $\boldsymbol{e}_2'\boldsymbol{e}_2 = 1$. Also, $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are orthogonal (perpendicular), that is, $\boldsymbol{e}_1'\boldsymbol{e}_2 = 0$.

Example 1.6. Find the eigenvalues and corresponding eigenvectors of the following two matrices:

$\boldsymbol{A} = \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix} \Rightarrow \lambda_1 = 6,\ \lambda_2 = -4$ and $\boldsymbol{B} = \begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix} \Rightarrow \lambda_1 = 18,\ \lambda_2 = 9,\ \lambda_3 = 9$.

1.2 Spectral Decomposition

Any symmetric square matrix can be can be constructed from its eigenvalues and eigenvectors.

Let $\boldsymbol{A}$ be a $k \times k$ symmetric matrix having k non-zero eigenvalues $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_k$ with normalized eigenvectors $\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \cdots,\ \boldsymbol{e}_k$. Then, the spectral decomposition of $\boldsymbol{A}$ is given by:

$$\boldsymbol{A} = \lambda_1\boldsymbol{e}_1\boldsymbol{e}_1' + \lambda_2\boldsymbol{e}_2\boldsymbol{e}_2' + \cdots + \lambda_k\boldsymbol{e}_k\boldsymbol{e}_k' = \sum_{j=1}^{k}\lambda_j\boldsymbol{e}_j\boldsymbol{e}_j'.$$

Example 1.7. $\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix}$

- The eigenvalues: $|\boldsymbol{A} - \lambda \boldsymbol{I}| = \begin{vmatrix} 1-\lambda & 2 \\ 2 & -2-\lambda \end{vmatrix} = 0$

  $\Rightarrow \lambda^2 + \lambda - 6 = 0$. Thus, the eigenvalues of $\boldsymbol{A}$ are $\lambda_1 = 2$ and $\lambda_2 = -3$.

- The eigenvectors are:
  - For $\lambda_1 = 2$,

$$\boldsymbol{A}\boldsymbol{x}_1 = \lambda_1 \boldsymbol{x}_1 \Leftrightarrow \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} = 2 \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix}$$

  $\Rightarrow x_{21} = \frac{1}{2} x_{11} \Rightarrow \boldsymbol{x}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$

  The normalized eigenvector corresponding to $\lambda_1 = 2$ is $\boldsymbol{e}_1 = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}$.

  - For $\lambda_2 = -3$,

$$\boldsymbol{A}\boldsymbol{x}_2 = \lambda_2 \boldsymbol{x}_2 \Rightarrow \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} x_{12} \\ x_{22} \end{pmatrix} = 2 \begin{pmatrix} x_{12} \\ x_{22} \end{pmatrix}$$

  $\Rightarrow x_{22} = -2x_{12} \Rightarrow \boldsymbol{x}_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$

  The normalized eigenvector corresponding to $\lambda_2 = -3$ is $\boldsymbol{e}_2 = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{pmatrix}$.

  Note that $\boldsymbol{e}_1'\boldsymbol{e}_1 = \boldsymbol{e}_2'\boldsymbol{e}_2 = 1$ and $\boldsymbol{e}_1'\boldsymbol{e}_2 = \boldsymbol{e}_2'\boldsymbol{e}_1 = 0$.

We need to show $\boldsymbol{A} = \lambda_1 \boldsymbol{e}_1\boldsymbol{e}_1' + \lambda_2 \boldsymbol{e}_2\boldsymbol{e}_2'$.

$$\begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix} = 2 \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} - 3 \begin{pmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{pmatrix}$$

The matrix is written as a function of eigenvalues and normalized eigenvectors.

In matrix form, the spectral decomposition of $\boldsymbol{A}$ is:

$$\boldsymbol{A} = \boldsymbol{O}\boldsymbol{\Lambda}\boldsymbol{O}'$$

where $\boldsymbol{O} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_k)$ and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_k) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}$.

Note here that $\boldsymbol{O}'\boldsymbol{O} = \boldsymbol{O}\boldsymbol{O}' = \boldsymbol{I}_{k\times k}$ ($\boldsymbol{O}$ is orthogonal, $\boldsymbol{O}^{-1} = \boldsymbol{O}'$).
In the above example,

$$\boldsymbol{O} = (\boldsymbol{e}_1, \boldsymbol{e}_2) = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix}, \ \Lambda = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}$$

$$\Rightarrow \boldsymbol{A} = \boldsymbol{O}\boldsymbol{\Lambda}\boldsymbol{O}'.$$

Again, using spectral decomposition, the inverse of $\boldsymbol{A}$ is given by

$$\boldsymbol{A}^{-1} = \boldsymbol{O}\boldsymbol{\Lambda}^{-1}\boldsymbol{O}'$$

where $\boldsymbol{O} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_k)$ and $\boldsymbol{\Lambda}^{-1} = \text{diag}\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \cdots, \frac{1}{\lambda_k}\right)$.

$$\Rightarrow \boldsymbol{A}^{-1} = \sum_{j=1}^{k} \frac{1}{\lambda_j} \boldsymbol{e}_j \boldsymbol{e}_j'.$$

Also, for a positive definite matrix $\boldsymbol{A}$,

$$\boldsymbol{A}^{\frac{1}{2}} = \boldsymbol{O}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{O}'$$

where $\boldsymbol{O} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_k)$ and $\boldsymbol{\Lambda}^{\frac{1}{2}} = \text{diag}\left(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_k}\right)$.

$$\Rightarrow \boldsymbol{A}^{\frac{1}{2}} = \sum_{j=1}^{k} \sqrt{\lambda_j} \boldsymbol{e}_j \boldsymbol{e}_j'.$$

Example 1.8. Find $\boldsymbol{A}^{-1}$ and $\boldsymbol{A}^{\frac{1}{2}}$. $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ and $\boldsymbol{A} = \begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix}$

1.3 Singular Value Decomposition

Let $\boldsymbol{A}$ be an $m \times k$ matrix. Then there exist an $m \times m$ orthogonal matrix $\boldsymbol{U}$ (i.e., $\boldsymbol{U}\boldsymbol{U}' = \boldsymbol{I}$) and a $k \times k$ orthogonal matrix $\boldsymbol{V}$ (i.e., $\boldsymbol{V}\boldsymbol{V}' = \boldsymbol{I}$) such that $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}'$ where $\boldsymbol{\Lambda}$ is an $m \times k$ matrix with (i, i) entry $\lambda_i \geq 0$ for $i = 1, 2, \cdots, min(m, k)$ and the other entries are zero.

- $\boldsymbol{U} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_{min(m,k)})$ where $\boldsymbol{e}_i$ $(i = 1, 2, \cdots, min(m, k))$ is the normalized eigenvector corresponding to λ_i of the matrix $\boldsymbol{A}\boldsymbol{A}'$.
- $\boldsymbol{V} = (\boldsymbol{e}_1^*, \boldsymbol{e}_2^*, \cdots, \boldsymbol{e}_{min(m,k)}^*)$ where $\boldsymbol{e}_i^*$ $(i = 1, 2, \cdots, min(m, k))$ is the normalized eigenvector corresponding to λ_i^* of the matrix $\boldsymbol{A}'\boldsymbol{A}$.
- $\boldsymbol{\Lambda} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_{min(m,k)}} \end{bmatrix}$

  Note that $\sqrt{\lambda_i}$ is the eigenvalue of matrix $\boldsymbol{A}$ where λ_i is the eigenvalue of $\boldsymbol{A}'\boldsymbol{A}$ or $\boldsymbol{A}\boldsymbol{A}'$.

Example 1.9. $\boldsymbol{A} = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$

$\boldsymbol{A}\boldsymbol{A}' = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$ and

$\boldsymbol{A}'\boldsymbol{A} = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}$

- Eigenvalues and eigenvectors corresponding to $\boldsymbol{AA'}$.

  $|\boldsymbol{AA'} - \lambda\boldsymbol{I}| = 0 \Rightarrow \begin{vmatrix} 11-\lambda & 1 \\ 1 & 11-\lambda \end{vmatrix} = 0 \Rightarrow (11-\lambda)^2 - 1 = 0$

  $\lambda^2 - 22\lambda + 120 = 0 \Rightarrow \lambda = 12$ or $\lambda = 10$.

  The eigenvalues of $\boldsymbol{AA'}$ or $\boldsymbol{A'A}$ are $\lambda_1 = 12$ and $\lambda_2 = 10$ which implies the eigenvalues of $\boldsymbol{A}$ to be $\lambda_1 = \sqrt{12}$ and $\lambda_2 = \sqrt{10}$.

  - Eigenvector corresponding to $\lambda_1 = 12$,

    $$\boldsymbol{AA'}\boldsymbol{x}_1 = \lambda_1\boldsymbol{x}_1 \Rightarrow \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}\begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = 12\begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} \Rightarrow x_{21} = x_{11}$$

    Let $x_{11} = 1 \Rightarrow x_{21} = 1 \Rightarrow \boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \boldsymbol{e}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

  - Eigenvector corresponding to $\lambda_2 = 10$,

    $$\boldsymbol{AA'}\boldsymbol{x}_2 = \lambda_2\boldsymbol{x}_2 \Rightarrow \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}\begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} = 10\begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} \Rightarrow x_{22} = -x_{12}$$

    Let $x_{12} = 1 \Rightarrow x_{22} = -1 \Rightarrow \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Rightarrow \boldsymbol{e}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$

  $$\boldsymbol{U} = (\boldsymbol{e}_1, \boldsymbol{e}_2) = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \text{ and } \boldsymbol{\Lambda} = \text{diag}\left(\sqrt{\lambda_1}, \sqrt{\lambda_2}\right) = \begin{bmatrix} \sqrt{12} & 0 \\ 0 & \sqrt{10} \end{bmatrix}$$

- Eigenvalues and eigenvectors corresponding to $\boldsymbol{A'A}$.

  $$|\boldsymbol{A'A} - \lambda\boldsymbol{I}| = 0 \Rightarrow \begin{vmatrix} 10-\lambda & 0 & 2 \\ 0 & 10-\lambda & 4 \\ 2 & 4 & 2-\lambda \end{vmatrix} = 0$$

  $$(10-\lambda)\begin{vmatrix} 10-\lambda & 4 \\ 4 & 2-\lambda \end{vmatrix} - 0\begin{vmatrix} 0 & 4 \\ 2 & 2-\lambda \end{vmatrix} + 2\begin{vmatrix} 0 & 10-\lambda \\ 2 & 4 \end{vmatrix} = 0$$

  $\lambda^2 - 12\lambda = 0 \Rightarrow \lambda = 12$ or $\lambda = 10$ or $\lambda = 0$.

  - Eigenvector corresponding to $\lambda_1 = 12$,

    $$\boldsymbol{A'A}\boldsymbol{x}_1 = \lambda_1\boldsymbol{x}_1 \Rightarrow \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}\begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix} = 12\begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix}$$

    $$10x_{11} + 0x_{21} + 2x_{31} = 12x_{11}$$
    $$0x_{11} + 10x_{21} + 4x_{31} = 12x_{21}$$
    $$2x_{11} + 4x_{21} + 2x_{31} = 12x_{31}$$

    $\Rightarrow x_{21} = 2x_{11}$ and $x_{31} = x_{11}$

    Let $x_{11} = 1 \Rightarrow x_{21} = 2$ and $x_{31} = 1 \Rightarrow \boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \Rightarrow \boldsymbol{e}_1^* = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}$

– Eigenvector corresponding to $\lambda_2 = 10$,

$$\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x}_2 = \lambda_2\boldsymbol{x}_2 \Rightarrow \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \end{bmatrix} = 10 \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \end{bmatrix}$$

$$10x_{12} + 0x_{22} + 2x_{32} = 10x_{12}$$

$$0x_{12} + 10x_{22} + 4x_{32} = 10x_{22}$$

$$2x_{12} + 4x_{22} + 2x_{32} = 10x_{32}$$

$\Rightarrow x_{32} = 0$ and $x_{22} = -\frac{1}{2}x_{12}$.

Let $x_{12} = 2 \Rightarrow x_{22} = 1 \Rightarrow \boldsymbol{x}_1 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \Rightarrow \boldsymbol{e}_2^* = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \\ 0 \end{bmatrix}$

$$\boldsymbol{V} = (\boldsymbol{e}_1^*, \boldsymbol{e}_2^*) = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{6}} & 0 \end{bmatrix}$$

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}' = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 \\ 0 & \sqrt{10} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 0 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

Chapter 2

Aspects of Multivariate Analysis

2.1 Introduction

Multivariate statistical analysis is concerned with data collected with several dimensions of the same individual (subject or experimental unit). Using multivariate analysis, the variables can be examined simultaneously in order to access the key features of the process. It enables us to

- explore the joint performance of the variables and
- determine the effect of each variable in the presence of the others.

As in the univariate case, it is assumed that a random sample of the multi-component observations has been collected from different individuals. The data consists of simultaneous measurements on many response variables. The common source of each individual observation will generally lead to dependence or correlation among the dimension (components). And this is the feature that distinguishes multivariate data and techniques from their univariate counterparts.

2.1.1 Objectives of Multivariate Analysis

The objectives of scientific investigations to which multivariate methods most naturally lend themselves include the following:

1. **Data reduction or structural simplification**. The phenomenon being studied is represented as simply as possible without sacrificing valuable information. This will make interpretation easier. Example: principal component analysis.
2. **Sorting and grouping.** Groups of "similar" objects or variables are created, based upon measured characteristics. Example: discriminant analysis.
3. **Investigation of the dependence among variables**. Studying the covariance structure will help determine the nature of the relationships among variables. In multivariate study, the interest is on the off-diagonals (covariances). Are all the variables mutually independent or are one or more variables dependent on the others? If so, how? Example: canonical correlation analysis.

4. **Prediction**. The relationship between variables can be determined for the purpose of predicting the values of one or more variables on the basis of observations on the other variables. Example: multivariate linear regression, multivariate analysis of variance.

5. **Hypothesis testing**. Specific statistical hypotheses can be tested to validate assumptions or to reinforce prior convictions.

2.1.2 Organization of Multivariate Data

Most multivariate data sets can be represented in *a rectangular format*, in which the elements of each row correspond to the variables values of a particular unit and the elements of the columns correspond to the values taken by a particular variable.

Suppose there are $p \geq 2$ variables (characteristics) measured from n items. Let x_{ij} denote the value of the j^{th} variable on the i^{th} item ($i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, p$, $n >> p$). Consequently, the data can be displayed as follows:

	Variable 1	Variable 2	$\cdots$	Variable j	$\cdots$	Variable p
Item 1	x_{11}	x_{12}	$\cdots$	x_{lj}	$\cdots$	x_{lp}
Item 2	x_{21}	x_{22}	$\cdots$	x_{2j}	$\cdots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
Item i	x_{i1}	x_{i2}	$\cdots$	x_{ij}	$\ldots$	x_{ip}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
Item n	x_{n1}	x_{n2}	$\cdots$	x_{nj}	$\cdots$	x_{np}

This can be written as a rectangular array, matrix, $\boldsymbol{X}$ of n rows and p columns:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{lp} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \ldots & x_{ij} & \ldots & x_{ip} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{bmatrix}_{n \times p}$$

A single multivariate observation is the collection of measurements on p different variables on the same item. Each row of $\boldsymbol{X}$ represents a multivariate observation.

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \ldots & x_{ij} & \ldots & x_{ip} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_i' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix} \begin{matrix} \\ \leftarrow 1^{st} \text{ multivariate observation} \\ \leftarrow 2^{nd} \text{ multivariate observation} \\ \\ \leftarrow i^{th} \text{ multivariate observation} \\ \\ \leftarrow n^{th} \text{ multivariate observation} \end{matrix}$$

Descriptive Statistics

A large data set is bulky, and its very mass poses a serious obstacle to any attempt to visually extract pertinent information. Much of the information contained in the data can be assessed by calculating certain summary numbers, known as descriptive statistics. For example, the arithmetic average, or sample mean, is a descriptive statistic that provides a measure of location-that is, a "central value" for a set of numbers. And the average of the squares of the distances of all of the numbers from the mean provides a measure of the spread, or variation, in the numbers.

- Sample mean: $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij};\ j = 1, 2, \cdots, p$
- Sample variance: $s_j^2 = s_{jj} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2;\ j = 1, 2, \cdots, p$
- Sample covariance between X_j and X_k: $s_{jk} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k);\ j, k = 1, 2, \cdots, p;\ j \neq k$. Note $s_{jk} = s_{kj}$ for all j and k.
- Sample correlation coefficient between variable j and k: $r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}}\sqrt{s_{kk}}};\ j, k = 1, 2, \cdots, p$. Note $r_{jk} = r_{kj}$ and $r_{jk} = 1$ if $j = k$.

  Although, the sign of the sample correlation and sample covariance are the same, the correlation is ordinarily easier to interpret as:

  - its magnitude is bounded, that is, $-1 \leq r_{jk} \leq 1$ for all j and k.
  - it is unitless.
  - it takes the variability into account.

  But the major disadvantages of correlation are it does not measure non-linear associations and it does not indicate any cause and effect.

The descriptive statistics for all the p variables in terms of vector and matrix operations are:

- Sample mean vector: $\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix}_{p\times 1}$.

  Also,

$$\bar{\boldsymbol{x}} = \frac{1}{n}\boldsymbol{X}'\boldsymbol{1} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{i1} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{i2} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{1j} & x_{2j} & \ldots & x_{ij} & \ldots & x_{nj} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{ip} & \cdots & x_{np} \end{bmatrix}_{p\times n} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n\times 1}$$

- Sample variance-covariance matrix: $\boldsymbol{S}_n = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$.

$$\boldsymbol{S}_n = \frac{1}{n}\left[\boldsymbol{X}'\boldsymbol{X} - \frac{1}{n}\boldsymbol{X}'\boldsymbol{1}(\boldsymbol{X}'\boldsymbol{1})'\right] = \frac{1}{n}\left(\boldsymbol{X}'\boldsymbol{X} - \frac{1}{n}\boldsymbol{X}'\boldsymbol{1}\boldsymbol{1}'\boldsymbol{X}\right)$$

$$\Rightarrow \boldsymbol{S}_n = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}_{p\times p}$$

Consequently, the sample standard deviation matrix is written as:

$$\boldsymbol{V}_n^{\frac{1}{2}} = \begin{bmatrix} \sqrt{s_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{s_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{s_{pp}} \end{bmatrix}_{p\times p}$$

- Sample correlation matrix: $\boldsymbol{R} = (\boldsymbol{V}_n^{\frac{1}{2}})^{-1}\boldsymbol{S}_n(\boldsymbol{V}_n^{\frac{1}{2}})^{-1}$

$$\Rightarrow \boldsymbol{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}_{p\times p}$$

Note $\boldsymbol{S}_n = \boldsymbol{V}_n^{\frac{1}{2}}\boldsymbol{R}\boldsymbol{V}_n^{\frac{1}{2}}$. Note also that $\boldsymbol{S}_n$ and $\boldsymbol{R}$ are symmetric and positive definite.

Example 2.1. Find the sample mean vector, covariance and correlation matrices for the following data matrix.

$$\begin{bmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 5 \end{bmatrix}$$

We find three observations, and here is what we observe (with notation: $x_{11} = 4$, $x_{21} = -1$, $x_{31} = 3$ and $x_{12} = 1$, $x_{22} = 3$, $x_{32} = 5$). The data array would the look like:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 5 \end{bmatrix}$$

2.2 Random Vectors and Matrices

A random vector (matrix) is a vector (matrix) whose elements are random variables. Let X_j be the j^{th} variable, then

- Mean$(X_j) = \mu_j = E(X_j)$; $j = 1, 2, \cdots, p$

- $\text{Var}(X_j) = \sigma_j^2 = E(X_j - \mu_j)^2;\ j = 1, 2, \cdots, p$
- $\text{Cov}(X_j, X_k) = \sigma_{jk} = E(X_j - \mu_j)(X_k - \mu_k);\ j, k = 1, 2, \cdots, p$
- $\text{Cor}(X_j, X_k) = \rho_{jk} = \dfrac{\sigma_{jk}}{\sqrt{\sigma_{jj}}\sqrt{\sigma_{kk}}};\ j, k = 1, 2, \cdots, p$

Let $\boldsymbol{X}$ be an $n \times p$ random vector, i.e, $\boldsymbol{X} = (X_1, X_2, \cdots, X_p)'$. Then the mean vector is:

$$E(\boldsymbol{X}) = E\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu}$$

The population variance-covariance matrix is:

$$\boldsymbol{\Sigma} = E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})' = E\left(\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_{p-\mu_p} \end{bmatrix} \begin{bmatrix} X_1 - \mu_1, X_2 - \mu_2, \cdots, X_p - \mu_2 \end{bmatrix} \right)$$

$$\Rightarrow \boldsymbol{\Sigma} = E\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \ldots & (X_p - \mu_p)^2 \end{bmatrix}$$

$$= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) & \ldots & E(X_1 - \mu_1)(X_p - \mu_p) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 & \ldots & E(X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_p - \mu_p)(X_1 - \mu_1) & E(X_p - \mu_p)(X_2 - \mu_2) & \ldots & E(X_p - \mu_p)^2 \end{bmatrix}$$

Thus, $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}_{p \times p}$.

If the p components are independently distributed (which rarely happens), then all the $\frac{1}{2}p(p-1)$ covariances of $\boldsymbol{\Sigma}$ will be zero. This cannot dealt with multivariate analysis rather univariate analysis.

The population standard deviation matrix is written as:

$$\boldsymbol{\sigma}^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix}_{p \times p}$$

Also, the population correlation matrix is $\boldsymbol{\rho} = (\boldsymbol{\sigma}^{\frac{1}{2}})^{-1}\boldsymbol{\Sigma}(\boldsymbol{\sigma}^{\frac{1}{2}})^{-1}$, that is,

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}_{p\times p}$$

Note $\boldsymbol{\Sigma} = \boldsymbol{\sigma}^{\frac{1}{2}}\boldsymbol{\rho}\boldsymbol{\sigma}^{\frac{1}{2}}$. Also here, $\boldsymbol{\Sigma}$ and $\boldsymbol{\rho}$ are symmetric and positive definite.

2.3 Distance of Vectors

Most multivariate techniques are based upon the simple concept of distance. If the point $P = (x_1, x_2)$ is the point on the XY plane, then the Euclidean (straight line) distance from P to the origin $O = (0, 0)$ is given by the Pythagorean theorem. That is,

$$d_E(O, P) = \sqrt{(x_1 - 0)^2 + (x_2 - 0)^2} = \sqrt{x_1^2 + x_2^2}.$$

All the points (x_1, x_2) that lie a constant distance, say c, from the origin satisfying the equation $c = \sqrt{x_1^2 + x_2^2} \Rightarrow c^2 = x_1^2 + x_2^2$ is called equation of a circle with radius c.

The Euclidean distance between two points $P = (x_1, x_2)$ and $Q = (y_1, y_2)$ in the two dimensional space is

$$d_E(P, Q) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$$

Similarly, the Euclidean distance between $P = (x_1, x_2, \cdots, x_p)$ and $Q = (y_1, y_2, \cdots, y_p)$ in the p dimensional space is

$$d_E(P, Q) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_p - y_p)^2} = \sqrt{(\boldsymbol{x} - \boldsymbol{y})'(\boldsymbol{x} - \boldsymbol{y})}.$$

Suppose $\boldsymbol{X}' = (X_1, X_2, \cdots, X_p)$ follows a p dimensional distribution with mean $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and covariance matrix $\text{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}$. And suppose again $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_p)'$ is a vector of means based on an $n \times p$ observed data matrix.

The Euclidean distance between the sample mean $\bar{\boldsymbol{x}}$ and the theoretical mean $\boldsymbol{\mu}$ is given by

$$d_S(\bar{\boldsymbol{x}}, \boldsymbol{\mu}) = \sqrt{(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'(\bar{\boldsymbol{x}} - \boldsymbol{\mu})}.$$

Straight line or Euclidean distance is unsatisfactory for most statistical purposes. This is because each co-ordinate contributes equally to the calculation of Euclidean distance. This suggests a statistical measure of distance.

The statistical distance between $\bar{\boldsymbol{x}}$ and $\boldsymbol{\mu}$ is given by

$$d_S(\bar{\boldsymbol{x}}, \boldsymbol{\mu}) = \sqrt{(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})}.$$

$$\Rightarrow d_S(\bar{\boldsymbol{x}}, \boldsymbol{\mu}) = \sqrt{\begin{bmatrix}\bar{x}_1 - \mu_1, \bar{x}_2 - \mu_2, \cdots, \bar{x}_p - \mu_p\end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}^{-1} \begin{bmatrix}\bar{x}_1 - \mu_1 \\ \bar{x}_2 - \mu_2 \\ \vdots \\ \bar{x}_p - \mu_p\end{bmatrix}}$$

Hence, a statistical distance takes into account the variability as well as the correlation unlike the Euclidean distance.

- If $\boldsymbol{\Sigma} = \boldsymbol{I}$, the Euclidean and statistical distances are equal.
- If $\sigma_{ij} = 0$ for $i \neq j$, the statistical distance is given by:

$$d_S(\bar{\boldsymbol{x}}, \boldsymbol{\mu}) = \sqrt{\frac{(x_1 - \mu_1)^2}{\sigma_{11}} + \frac{(x_2 - \mu_2)^2}{\sigma_{22}} + \cdots + \frac{(x_p - \mu_p)^2}{\sigma_{pp}}}.$$

If one component has much larger variance than another, it will contribute less to the squared distance. Two highly correlated variables will contribute less than two variables that are nearly uncorrelated. Essentially, the use of the inverse of the covariance matrix eliminates the effect of correlation and standardizes all of the variables.

Example 2.2. Let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$. The variability in the x_1 direction is greater than that in the x_2 direction as $\sigma_{11} = 4 > \sigma_{22} = 1$.

Euclidean distance: $d_E = \sqrt{(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2}$.

Statistical distance: $d_S = \sqrt{(\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})}$.

$$\Rightarrow d_S = \sqrt{(x_1 - \mu_1, x_2 - \mu_2) \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}}$$

$$\Rightarrow d_S = \sqrt{\frac{(x_1 - \mu_1)^2}{4} + \frac{(x_2 - \mu_2)^2}{1}}$$

This is simply an equation of ellipse which is centered at $\boldsymbol{\mu} = (\mu_1, \mu_2)'$. All points that lie a constant distance, say $c = 2$ (the boundary of the ellipse), from the theoretical mean $\boldsymbol{\mu}$ satisfy the equation

$$\frac{(x_1 - \mu_1)^2}{4} + \frac{(x_2 - \mu_2)^2}{1} = c^2 = 4.$$

At $x_1 = \mu_1$, $(x_2 - \mu_2)^2 = 4 \Rightarrow x_2 - \mu_2 = \pm 2 \Rightarrow x_2 = \mu_2 \pm 2$.
At $x_2 = \mu_2$, $(x_1 - \mu_1)^2 = 16 \Rightarrow x_1 - \mu_1 = \pm 4 \Rightarrow x_1 = \mu_1 \pm 4$.

Plotoftheellipse

The ellipse stretches in the x_1 direction as compared to that in the x_2 direction because of the larger variance in x_1 (the ellipse is parallel to the x_1). Having the same variance in both axes, the equation will be simply a circle.

2.4 Linear Combinations of Random Vectors

1. Univariate case: For a single random variable X, $E(X) = \mu$ and $\text{Var}(X) = E(X - \mu)^2 = \sigma^2$. Then, for a linear combination $Z = aX$, $E(aX) = aE(X) = a\mu$, and $\text{Var}(aX) = a^2\text{Var}(X) = a^2\sigma^2$ for $a \in \mathcal{R}$.

2. Bivariate case: For two random variables, X_1 and X_2, $E(X_1) = \mu_1$ and $E(X_2) = \mu_2$; $\text{Var}(X_1) = E(X_1 - \mu_1)^2 = \sigma_{11}$ and $\text{Var}(X_2) = E(X_2 - \mu_2)^2 = \sigma_{22}$; $\text{Cov}(X_1, X_2) = E(X_1 - \mu_1)(X_2 - \mu_2) = \sigma_{12}$.

   If $\boldsymbol{X}' = (X_1, X_2)$ and $\boldsymbol{a}' = (a_1, a_2)$, then the linear combination $Z = \boldsymbol{a}'\boldsymbol{X} = a_1X_2 + a_2X_2$ is a one-dimensional random variable.

   - $E(\boldsymbol{a}'\boldsymbol{X}) = E(a_1X_1 + a_2X_2) = a_1E(X_1) + a_2E(X_2) = a_1\mu_1 + a_2\mu_2$

     $$\Rightarrow E(\boldsymbol{a}'\boldsymbol{X}) = (a_1, a_2)\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} = \boldsymbol{a}'\boldsymbol{\mu}$$

   - $\text{Var}(\boldsymbol{a}'\boldsymbol{X}) = \text{Var}(a_1X_1 + a_2X_2) = E[a_1X_1 + a_2X_2 - (a_1\mu_1 + a_2\mu_2)]^2$

     $$\begin{aligned}\Rightarrow \text{Var}(\boldsymbol{a}'\boldsymbol{X}) &= a_1^2\sigma_{11} + a_2^2\sigma_{22} + 2a_1a_2\sigma_{12}\\ &= (a_1, a_2)\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}\begin{bmatrix}a_1\\ a_2\end{bmatrix}\\ &= \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}\end{aligned}$$

3. Multivariate case: If $\boldsymbol{X}$ a p-dimensional random vector, $\boldsymbol{X} = (X_1, X_2, \cdots, X_p)$, and $\boldsymbol{a} \in \mathcal{R}^p$, then the linear combination $Z = \boldsymbol{a}'\boldsymbol{X} = a_1X_1 + a_2X_2 + \cdots + a_pX_p$ is univariate. That is,

   - $E(\boldsymbol{a}'\boldsymbol{X}) = \boldsymbol{a}'E(\boldsymbol{X}) = \boldsymbol{a}'\boldsymbol{\mu}$
   - $\text{Var}(\boldsymbol{a}'\boldsymbol{X}) = E(\boldsymbol{a}'\boldsymbol{X} - \boldsymbol{a}'\boldsymbol{\mu})^2 = \boldsymbol{a}'E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{a} = \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}$

4. Consider q linear combinations of p random variables.

   $$\begin{aligned}Z_1 &= \boldsymbol{a}_1'\boldsymbol{X} = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p = \sum_{j=1}^{p} a_{1j}X_j\\ Z_2 &= \boldsymbol{a}_2'\boldsymbol{X} = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p = \sum_{j=1}^{p} a_{2j}X_j\\ &\vdots\\ Z_q &= \boldsymbol{a}_q'\boldsymbol{X} = a_{q1}X_1 + a_{q2}X_2 + \cdots + a_{qp}X_p = \sum_{j=1}^{p} a_{qj}X_j\end{aligned}$$

   In matrix form:

   $$\begin{bmatrix}Z_1\\ Z_2\\ \vdots\\ Z_q\end{bmatrix} = \begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1p}\\ a_{21} & a_{22} & \cdots & a_{2p}\\ \vdots & \vdots & \ddots & \vdots\\ a_{q1} & a_{q2} & \cdots & a_{qp}\end{bmatrix}\begin{bmatrix}X_1\\ X_2\\ \vdots\\ X_p\end{bmatrix} \Leftrightarrow \boldsymbol{Z} = \boldsymbol{A}\boldsymbol{X}$$

   - $E(\boldsymbol{Z}) = E(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}E(\boldsymbol{X}) = \boldsymbol{A}\boldsymbol{\mu}$
   - $\text{Cov}(\boldsymbol{Z}) = \text{Cov}(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{A}' = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$

Example 2.3. Find the mean vector and covariance matrix for the linear combinations $Z_1 = X_1 - X_2$ and $Z_2 = X_1 + X_2$.

$$\boldsymbol{Z} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \boldsymbol{AX}$$

- $E(\boldsymbol{Z}) = \boldsymbol{A}E(\boldsymbol{X}) = \boldsymbol{A\mu} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 + \mu_2 \end{bmatrix}$
- $\text{Cov}(\boldsymbol{Z}) = \boldsymbol{A}\text{Cov}(\boldsymbol{Z})\boldsymbol{A}'$

  $\Rightarrow \text{Cov}(\boldsymbol{Z}) = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{11} - \sigma_{22} \\ \sigma_{11} - \sigma_{22} & \sigma_{11} + 2\sigma_{12} + \sigma_{22} \end{bmatrix}$

2.5 Expected Value of the Sample Mean Vector and Covariance Matrix

Let $\boldsymbol{X}$ be a random matrix given by:

$$\boldsymbol{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{lp} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1' \\ \boldsymbol{X}_2' \\ \vdots \\ \boldsymbol{X}_n' \end{bmatrix}$$

If $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ is a random sample from some joint distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, then

- $E(\bar{\boldsymbol{X}}) = \frac{1}{n}\sum_{i=1}^{n} E(\boldsymbol{X}_i) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\mu} = \boldsymbol{\mu}$. Thus, $\bar{\boldsymbol{X}}$ is an unbiased estimator of the mean vector $\boldsymbol{\mu}$.
- $\text{Cov}(\bar{\boldsymbol{X}}) = E(\bar{\boldsymbol{X}} - \boldsymbol{\mu})(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'$.

$$\begin{aligned} \text{Cov}(\bar{\boldsymbol{X}}) &= \begin{bmatrix} E(\bar{X}_1 - \mu_1)^2 & E(\bar{X}_1 - \mu_1)(\bar{X}_2 - \mu_2) & \cdots & E(\bar{X}_1 - \mu_1)(\bar{X}_p - \mu_p) \\ E(\bar{X}_2 - \mu_2)(\bar{X}_1 - \mu_1) & E(\bar{X}_2 - \mu_2)^2 & \cdots & E(\bar{X}_2 - \mu_2)(\bar{X}_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(\bar{X}_p - \mu_p)(\bar{X}_1 - \mu_1) & E(\bar{X}_p - \mu_p)(\bar{X}_2 - \mu_2) & \cdots & E(\bar{X}_p - \mu_p)^2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{n}\sigma_{11} & \frac{1}{n}\sigma_{12} & \cdots & \frac{1}{n}\sigma_{1p} \\ \frac{1}{n}\sigma_{21} & \frac{1}{n}\sigma_{22} & \cdots & \frac{1}{n}\sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n}\sigma_{p1} & \frac{1}{n}\sigma_{p2} & \cdots & \frac{1}{n}\sigma_{pp} \end{bmatrix} \\ &= \frac{1}{n}\boldsymbol{\Sigma} \end{aligned}$$

Recall the sample variance-covariance matrix $\boldsymbol{S}_n = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'$. It can be shown that $E(\boldsymbol{S}_n) = \frac{n-1}{n}\boldsymbol{\Sigma}$. Thus, $\boldsymbol{S}_n$ is a biased estimator of $\boldsymbol{\Sigma}$. This implies

$\boldsymbol{S} = \frac{n}{n-1}\boldsymbol{S}_n = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'$ is an unbiased estimator of $\boldsymbol{\Sigma}$, i.e., $E(\boldsymbol{S}) = \frac{n}{n-1}E(\boldsymbol{S}_n) = \frac{n}{n-1}\frac{n-1}{n}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}$.

Chapter 3

The Multivariate Normal Distribution

Just as the normal distribution dominates univariate techniques, the multivariate normal distribution plays an important role in most multivariate procedures, because

- the multivariate normal distribution is mathematically tractable and "nice" results can be obtained. Mathematical complexity of other data generating distributions may prevent the development of sampling distribution of the usual test statistics and estimators.
- the sampling distribution of many multivariate statistics are approximately normal, regardless of the the form of the parent population, because of the central limit effect; i.e., as the number of source random vectors increases without bound.

3.1 The Multivariate Normal Density and Its Properties

Univariate case:

Let X be a random variable with $E(X) = \mu$ and $var(X) = \sigma^2$. Then if $X \sim \mathcal{N}(\mu, \sigma^2)$, its pdf is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right].$$

Note the term $\dfrac{(x-\mu)^2}{\sigma^2} = (x-\mu)(\sigma^2)^{-1}(x-\mu)$ measures the squared statistical distance form x to μ in standard deviation units.

Multivariate case:

Suppose $\boldsymbol{X}' = (X_1, X_2, \cdots, X_p)$ is a $p \times 1$ vector with $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $\text{Cov}(\boldsymbol{X}) = E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})' = \boldsymbol{\Sigma}$. The joint density of p independent normal variates, $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, is:
$f(\boldsymbol{x}) = f(x_1, x_2, \cdots, x_p) = f(x_1)f(x_2)\cdots f(x_p)$

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2}\frac{(x_1-\mu_1)^2}{\sigma_1^2}\right] \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{1}{2}\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right] \cdots \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left[-\frac{1}{2}\frac{(x_p-\mu_p)^2}{\sigma_p^2}\right]$$

$$\Rightarrow f(\boldsymbol{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^p} \frac{1}{\sigma_1\sigma_2\cdots\sigma_p} \exp\left[-\frac{1}{2}\sum_{j=1}^{p}\frac{(x_j-\mu_j)^2}{\sigma_j^2}\right]$$

Since $\boldsymbol{\Sigma} = \text{diag}(\sigma_{11}, \sigma_{22}, \cdots, \sigma_{pp})$, $|\boldsymbol{\Sigma}| = \sigma_{11}\sigma_{22}\cdots\sigma_{pp}$ and $|\boldsymbol{\Sigma}|^{\frac{1}{2}} = \sigma_1\sigma_2\cdots\sigma_p$. Also,

$$\sum_{j=1}^{p}\frac{(x_j-\mu_j)^2}{\sigma_j^2} = \sum_{j=1}^{p}(x_j-\mu_j)(\sigma_j^2)^{-1}(x_j-\mu_j) = (\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

Therefore, the joint density is

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right].$$

The general p dimensional normal density function is obtained by letting $\boldsymbol{\Sigma}$ to be any $p \times p$ symmetric matrix,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}.$$

Here, the j^{th} element of $\boldsymbol{\mu}$ is still $E(X_j) = \mu_j$. And the j^{th} element of $\boldsymbol{\Sigma}$ is still $\sigma_{jj} = E(X_j-\mu_j)^2$ but the $(j,k)^{th}$ element of $\boldsymbol{\Sigma}$ is now $\sigma_{jk} = E(X_j-\mu_j)(X_k-\mu_k)$, $i \neq k$.

For a general p dimensional normal density function, $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = c^2$ is the squared statistical distance from $\boldsymbol{x}$ to $\boldsymbol{\mu}$.

Note the symmetric matrix $\boldsymbol{\Sigma}$ is positive definite (all eigenvalues are positive). Let $\boldsymbol{a}' = (a_1, a_2, \cdots, a_p)$. We need to show $\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} > 0$. Hence, $\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} = \boldsymbol{a}'E(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{a}$ since $\boldsymbol{\Sigma} = E(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'$. Since $(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{a}$ is a scalar, its transpose makes no change, i.e., $\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}$ can be written as $E[\boldsymbol{a}'(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{a}] = E[\boldsymbol{a}'(\boldsymbol{X}-\boldsymbol{\mu})\boldsymbol{a}'(\boldsymbol{X}-\boldsymbol{\mu})] > 0$. Therefore, $\boldsymbol{\Sigma}$ is positive definite.

Example 3.1. Bivariate normal distribution ($p = 2$).

$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$

$\rho_{12} = \dfrac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} \Rightarrow \sigma_{12} = \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$

$\Rightarrow \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{22} \end{bmatrix} \Rightarrow |\boldsymbol{\Sigma}| = \sigma_{11}\sigma_{22} - \rho_{12}^2\sigma_{11}\sigma_{22} = (1-\rho_{12}^2)\sigma_{11}\sigma_{22}$

$\Rightarrow \boldsymbol{\Sigma}^{-1} = \dfrac{1}{(1-\rho_{12}^2)\sigma_{11}\sigma_{22}} \begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix}$

The squared statistical distance is $(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})=c$

$$
\begin{aligned}
c &= [X_1-\mu_1, X_2-\mu_2]\frac{1}{(1-\rho^2)\sigma_{11}\sigma_{22}}\begin{bmatrix}\sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}\\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11}\end{bmatrix}\begin{bmatrix}X_1-\mu_1\\ X_2-\mu_2\end{bmatrix}\\
&= \frac{1}{(1-\rho_{12}^2)\sigma_{11}\sigma_{22}}\begin{matrix}[\sigma_{22}(x_1-\mu_1)-\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(x_2-\mu_2),\\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(x_1-\mu_1)+\sigma_{11}(x_2-\mu_2)]\end{matrix}\begin{bmatrix}x_1-\mu_1\\ x_2-\mu_2\end{bmatrix}\\
&= \frac{1}{(1-\rho_{12}^2)\sigma_{11}\sigma_{22}}[\sigma_{22}(x_1-\mu_1)^2-2\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(x_1-\mu_1)(x_2-\mu_2)+\sigma_{11}(x_2-\mu_2)^2]\\
&= \frac{1}{(1-\rho_{12}^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_{11}}-2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)+\frac{(x_2-\mu_2)^2}{\sigma_{22}}\right]
\end{aligned}
$$

Therefore, the bivariate normal density is given by:

$$f(x_1,x_2)=\frac{1}{2\pi\sqrt{(1-\rho_{12}^2)\sigma_{11}\sigma_{22}}}\exp(-\frac{1}{2}c).$$

3.1.1 Principal Axis of the Multivariate Normal Density

The component $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ specifies the equation of an ellipsoid in the p dimensional space when it is set equal to some positive constant c. The family of ellipsoids generated by varying c have a common point $\boldsymbol{\mu}$, that is, each ellipsoid is centered at $\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_p)'$. For example, figure 3.1 represents an ellipse for $p=2$ obtained by varying the boundary of the ellipse c.

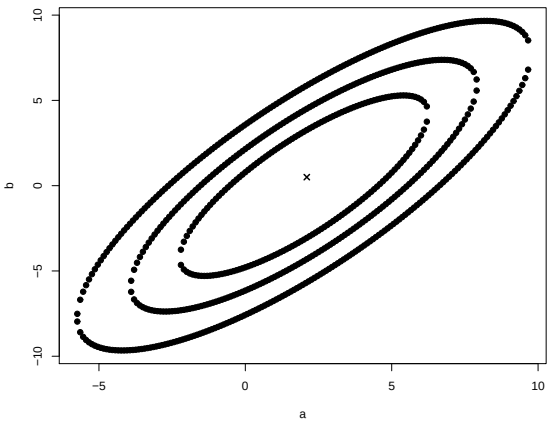

Figure 3.1: Plot of an ellipse for a bivariate normal distribution ($p=2$)

The first principal axis of each ellipsoid is the line passing through its largest dimension and the second which is perpendicular. If any line through $\boldsymbol{\mu}$ of an ellipsoid is represented by its coordinates $\boldsymbol{x}$ on the surface, then the first principal axis will have coordinates $\boldsymbol{x}$ that maximized its squared half length, $d^2=(\boldsymbol{x}-\boldsymbol{\mu})'(\boldsymbol{x}-\boldsymbol{\mu})$.

So, to maximize $d^2 = (\boldsymbol{x}-\boldsymbol{\mu})'(\boldsymbol{x}-\boldsymbol{\mu})$ subject to the constraint $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = c$ (c is fixed), the lagrange function is used as

$$f(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{\mu})'(\boldsymbol{x}-\boldsymbol{\mu}) - \lambda[(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - c]$$

where λ is the lagrange multiplier. Thus, the coordinates of the longest axis must satisfy the equation

$$\frac{\partial}{\partial \boldsymbol{x}} f(\boldsymbol{x}) = \boldsymbol{0} \Rightarrow 2(\boldsymbol{x}-\boldsymbol{\mu}) - 2\lambda\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \boldsymbol{0} \Rightarrow (\boldsymbol{I} - \lambda\boldsymbol{\Sigma}^{-1})(\boldsymbol{x}-\boldsymbol{\mu}) = \boldsymbol{0}.$$

Then, pre-multiplying by $\boldsymbol{\Sigma}$ gives

$$(\boldsymbol{\Sigma} - \lambda\boldsymbol{I})(\boldsymbol{x}-\boldsymbol{\mu}) = \boldsymbol{0}.$$

The trivial solution is $(\boldsymbol{x}-\boldsymbol{\mu}) = \boldsymbol{0}$. In order to have a non-trivial solution, $|\boldsymbol{\Sigma} - \lambda\boldsymbol{I}| = 0$. Hence, λ is the eigenvalues of $\boldsymbol{\Sigma}$. But, to which of the p eigenvalues of $\boldsymbol{\Sigma}$, does the vector $\boldsymbol{x}$ correspond? From above, we have $(\boldsymbol{I} - \lambda\boldsymbol{\Sigma}^{-1})(\boldsymbol{x}-\boldsymbol{\mu}) = \boldsymbol{0}$. This implies, $(\boldsymbol{x}-\boldsymbol{\mu}) = \lambda\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$. Pre-multiplying this equation by $(\boldsymbol{x}-\boldsymbol{\mu})'$, yields

$$\begin{aligned}(\boldsymbol{x}-\boldsymbol{\mu})'(\boldsymbol{x}-\boldsymbol{\mu}) &= \lambda(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \\ \Rightarrow d^2 &= \lambda c.\end{aligned}$$

For a fixed c, the length of the principal axis is maximized by taking λ as the largest eigenvalue λ_1 of $\boldsymbol{\Sigma}$. Thus, the half length of the major (principal) axis is equal to $d_1 = \sqrt{\lambda_1 c}$ in the direction of $\boldsymbol{e}_1$ (where $\boldsymbol{e}_1$ is the normalized eigenvector corresponding to the eigenvalue λ_1 of $\boldsymbol{\Sigma}$). Consequently, the full length of the principal axis is equal to $2d_1 = 2\sqrt{\lambda_1 c}$.

Example 3.2. Consider the bivariate case, $p = 2$. That is, $\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ with $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$. Assume $\mu_1 > \mu_2$, $\sigma_{11} = \sigma_{22}$ and $\sigma_{12} > 0$.

Let us plot the ellipse. Note that the ellipse could not be parallel to the X or Y axis as the off-diagonal of $\boldsymbol{\Sigma}$ is not zero.

To find the eigenvalues of $\boldsymbol{\Sigma}$: $|\boldsymbol{\Sigma} - \lambda\boldsymbol{I}| = 0 \Rightarrow \begin{vmatrix} \sigma_{11}-\lambda & \sigma_{12} \\ \sigma_{12} & \sigma_{11}-\lambda \end{vmatrix} = 0 \Rightarrow \lambda^2 + 2\sigma_{11}\lambda + \sigma_{11}^2 - \sigma_{12}^2 = 0$. This equation is quadratic in λ. Therefore, $\lambda_1 = \sigma_{11} + \sigma_{12}$ and $\lambda_2 = \sigma_{11} - \sigma_{12}$.

The eigenvectors (orientations) of the major and minor axes:

- For $\lambda_1 = \sigma_{11} + \sigma_{12} \Rightarrow \boldsymbol{\Sigma}\boldsymbol{x}_1 = \lambda_1\boldsymbol{x}_1$.

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = (\sigma_{11} + \sigma_{12}) \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix}$$

$$\begin{aligned} \Rightarrow \quad & \sigma_{11}x_{11} + \sigma_{12}x_{21} = \sigma_{11}x_{11} + \sigma_{12}x_{11} \\ & \sigma_{12}x_{11} + \sigma_{11}x_{21} = \sigma_{11}x_{21} + \sigma_{12}x_{21} \\ & \Rightarrow x_{21} = x_{11}. \end{aligned}$$

The major axis will be parallel to the line $x_{21} = x_{11}$. Let $x_{11} = 1 \Rightarrow x_{21} = 1$. Thus, the eigenvector corresponding to $\lambda_1 = \sigma_{11} + \sigma_{12}$ is $\boldsymbol{x}_1 = (1, 1)'$. The normalized eigenvector corresponding to $\lambda_1 = \sigma_{11} + \sigma_{12}$ is $\boldsymbol{e}_1 = (1/\sqrt{2}, 1/\sqrt{2})'$ which is the coordinates of the major axis. Hence, the first principal axis lies along the 45° through the center point $\boldsymbol{\mu} = (\mu_1, \mu_2)'$.

- For $\lambda_2 = \sigma_{11} - \sigma_{12} \Rightarrow \boldsymbol{\Sigma x}_2 = \lambda_2 \boldsymbol{x}_2$.

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} = (\sigma_{11} - \sigma_{12}) \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix}$$

$$\Rightarrow \begin{array}{c} \sigma_{11}x_{12} + \sigma_{12}x_{22} = \sigma_{11}x_{12} - \sigma_{12}x_{12} \\ \sigma_{12}x_{12} + \sigma_{11}x_{22} = \sigma_{11}x_{22} - \sigma_{12}x_{22} \end{array}$$

$$\Rightarrow x_{22} = -x_{12}.$$

The minor axis will be parallel to the line $x_{22} = -x_{12}$. Let $x_{12} = 1 \Rightarrow x_{22} = -1$. Thus, the eigenvector corresponding to $\lambda_2 = \sigma_{11} - \sigma_{12}$ is $\boldsymbol{x}_2 = (1, -1)'$. The normalized eigenvector corresponding to $\lambda_2 = \sigma_{11} - \sigma_{12}$ is $\boldsymbol{e}_2 = (1/\sqrt{2}, -1/\sqrt{2})'$ which is the coordinates of the minor axis.

Note the major and minor axes are perpendicular, that is, $\boldsymbol{e}_1'\boldsymbol{e}_2 = \boldsymbol{e}_2'\boldsymbol{e}_1 = 0$.

Beginning at the center $\boldsymbol{\mu}$, the half length of the major axis is $d_1 = \sqrt{\lambda_1 c} = \sqrt{(\sigma_{11} + \sigma_{12})c}$ in the direction of $\boldsymbol{e}_1$ and the half length of the minor axis is $d_2 = \sqrt{\lambda_2 c} = \sqrt{(\sigma_{11} - \sigma_{12})c}$ in the direction of $\boldsymbol{e}_2$.

Now, the plot of the ellipse is as shown below.

InsertthePlothere

Along the ellipse shown above (on the boundary of the ellipse) the bivariate normal density is constant. This path along the surface is called a contour.

Note:

- If $\sigma_{12} = 0$, then the equations of the major and minor axes will be reversed. That is, the major axis will be parallel to the line $x_{11} = -x_{21}$ while the minor axis will be parallel to the line $x_{12} = x_{22}$.
- If $\sigma_{12} = 0$ ($\rho_{12} = 0$), then the concentration ellipse would simply be a circle and an infinity of perpendicular axes can be given as "principal", each with half length $\sqrt{\sigma_{11}c}$.

Remarks: Recall spectral decomposition. Let $\boldsymbol{O}$ be a matrix whose columns are the normalized eigenvectors of $\boldsymbol{\Sigma}$ and let Λ be a diagonal matrix whose diagonals are the eigenvalues of $\boldsymbol{\Sigma}$. Then, using spectral decomposition

$$\boldsymbol{\Sigma} = \sum_{j=1}^{p} \lambda_j \boldsymbol{e}_j \boldsymbol{e}_j' = \boldsymbol{O\Lambda O}', \boldsymbol{\Sigma}^{-1} = \sum_{j=1}^{p} \frac{1}{\lambda_j} \boldsymbol{e}_j \boldsymbol{e}_j' = \boldsymbol{O\Lambda}^{-1}\boldsymbol{O}' \text{ and } \boldsymbol{\Sigma}^{\frac{1}{2}} = \sum_{j=1}^{p} \sqrt{\lambda_j} \boldsymbol{e}_j \boldsymbol{e}_j' = \boldsymbol{O\Lambda}^{\frac{1}{2}}\boldsymbol{O}'$$

Generally,

- $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = c$ defines ellipsoids of different sizes depending on c.
- Each ellipsoid is centered at $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_p)'$.
- The half lengths of the axes are $d_j = \sqrt{\lambda_j c}$ in the direction of $\boldsymbol{e}_j$; $j = 1, 2, \cdots, p$.

3.1.2 Further Properties of the Multivariate Normal Density

Let $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

1. Linear combinations of the components of $\boldsymbol{X}$ are also normally distributed. That is, if $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{a}'\boldsymbol{X} = a_1X_1 + a_2X_2 + \cdots + a_pX_p$ will have a univariate normal distribution. That is, $\boldsymbol{a}'\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a})$.

   More specifically, the marginal distribution of any component X_j of $\boldsymbol{X}$ is $\mathcal{N}(\mu_j, \sigma_{jj})$. Let $\boldsymbol{a}' = (0, 0, \cdots, \underbrace{1}_{j^{th}\text{position}}, \cdots, 0)$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_j, \cdots, \mu_p)'$. Then $\boldsymbol{a}'\boldsymbol{X} = X_j \sim \mathcal{N}(\mu_j, \sigma_{jj})$.

   Similarly, if $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the q linear combinations

$$\boldsymbol{Z} = \boldsymbol{A}\boldsymbol{X} = \begin{bmatrix} a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\ a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + a_{q2}X_2 + \cdots + a_{qp}X_p \end{bmatrix} \sim \mathcal{N}_q(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}').$$

2. All subsets of the components of $\boldsymbol{X}$ have a (multivariate) normal distribution. That is, if $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\begin{bmatrix} \boldsymbol{X}_{q\times 1} \\ ---- \\ \boldsymbol{X}_{(p-q)\times 1} \end{bmatrix}$ with have a multivariate normal distribution with $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_{q\times 1} \\ ---- \\ \boldsymbol{\mu}_{(p-q)\times 1} \end{bmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}^{11}_{q\times q} & | & \boldsymbol{\Sigma}^{12}_{q\times(p-q)} \\ ---- & | & ---- \\ \boldsymbol{\Sigma}^{21}_{(p-q)\times q} & | & \boldsymbol{\Sigma}^{22}_{(p-q)\times(p-q)} \end{bmatrix}$.
3. Zero covariance implies that the corresponding components are independently distributed (for normal distribution only). X_1 and X_2 are independent if and only if $\text{cov}(X_1, X_2) = 0$. That is, if $\text{cov}(X_1, X_2) = 0$, $f(x_1, x_2) = f(x_1)f(x_2)$.
4. The conditional distributions of the components are (multivariate) normal. $f(x_1|x_2) = \dfrac{f(x_1, x_2)}{f(x_2)} \sim \mathcal{N}\left[\mu_1 + \dfrac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \dfrac{\sigma_{12}^2}{\sigma_{22}}\right]$.

3.2 Sampling from the Multivariate Normal Distribution

3.2.1 The Multivariate Normal Likelihood

Recall if $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the multivariate normal density is given by

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right].$$

Let $\boldsymbol{X}_i$; $i = 1, 2, \cdots, n$ represent a (vector) random sample from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Since $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ are mutually independent and each is distributed as $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the joint distribution is

$$f(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{n} f(\boldsymbol{x}_i) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right]$$
$$= \frac{1}{(2\pi)^{\frac{np}{2}}|\boldsymbol{\Sigma}|^{\frac{n}{2}}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right].$$

This expression as a function of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for a fixed set of observations $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ is called likelihood function denoted by $\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$. That is,

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) = \frac{1}{(2\pi)^{\frac{np}{2}}|\boldsymbol{\Sigma}|^{\frac{n}{2}}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right].$$

To get the ML estimate of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, $\frac{\partial}{\partial \boldsymbol{\mu}} \log \ell = 0$ and $\frac{\partial}{\partial \boldsymbol{\Sigma}} \log \ell = 0$. Thus, it results $\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i = \bar{\boldsymbol{x}}$ and $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})' = \frac{n-1}{n}\boldsymbol{S}$.

3.2.2 The Sampling Distribution of $\bar{\boldsymbol{X}}$ and $\boldsymbol{S}$

- Univariate case: X_i; $i = 1, 2, \cdots, n$ be a random sample from $N(\mu, \sigma^2)$. Then,
  a. $\bar{X} \sim N(\mu, \frac{1}{n}\sigma^2)$.
  b. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ where $n > 1$ and $\sigma^2 > 0$.
  c. If $n > 1$, then $\bar{X}$ and S are independent where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
- Multivariate case: Let $\boldsymbol{X}_i$; $i = 1, 2, \cdots, n$ be a random sample of (vectors) from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then,
  a. $\bar{\boldsymbol{X}} \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$.
  b. $(n-1)\boldsymbol{S}$ is distributed as Wishart distribution (matrix) with $n-1$ degrees of freedom.
  c. $\bar{\boldsymbol{X}}$ and $\boldsymbol{S}$ are independent.

The sampling distribution of the sample covariance matrix is called the Wishart distribution. It is defined as the sum of independent products of multivariate normal random vectors, $\boldsymbol{Z}_i$. $\rightarrow W_n(\cdot|\boldsymbol{\Sigma})$ Wishart distribution with n degrees of freedom = distribution of $\sum_{i=1}^{n} \boldsymbol{Z}_i\boldsymbol{Z}_i'$. (Note for univariate distribution, $\sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$.

3.2.3 Large Sample Behaviour of $\bar{\boldsymbol{X}}$ and $\boldsymbol{S}$

- Recall if $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $|\boldsymbol{\Sigma}| > 0$, then
  - $(\boldsymbol{X} - \boldsymbol{\mu}) \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$
  - $\boldsymbol{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{I}_p)$
  - $\boldsymbol{Z}'\boldsymbol{Z} = (\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) = Z_1^2 + Z_2^2 + \cdots + Z_p^2 \sim \chi^2(p)$.
- Let $\boldsymbol{X}_i$; $i = 1, 2, \cdots, n$ be a random sample from any distribution with mean $\boldsymbol{\mu}$ and finite covariance $\boldsymbol{\Sigma}$. Then,
  - $(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim \mathcal{N}_p(\mathbf{0}, \frac{1}{n}\boldsymbol{\Sigma})$ for large n.
  - $\sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ for large n.
  - Since for large n, $\boldsymbol{S}$ is close to $\boldsymbol{\Sigma}$ with high probability, $\sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{S})$.
  - $\boldsymbol{Z} = \sqrt{n}\boldsymbol{\Sigma}^{-\frac{1}{2}}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{I}_p)$
  - $\boldsymbol{Z}'\boldsymbol{Z} = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim \chi^2(p)$ for large $n - p$.

Chapter 4

Inference about a Mean Vector

One of the central messages of multivariate analysis is that the p correlated variables must be analysed jointly.

4.1 The Plausibility of $\boldsymbol{\mu}_0$ as a Value for a Normal Population Mean $\boldsymbol{\mu}$

Univariate case:

Suppose a random sample of $X_1, X_2, \cdots, X_n$ is drawn from a normal population with mean μ and variance σ^2 (in practice σ^2 is unknown, s is used instead). Given $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. The test statistic is

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t(n-1).$$

The null hypothesis is rejected if $|t|$ is large. Rejecting H_0 when $|t|$ is large is equivalent to rejecting H_0 if t^2 is large. Hence, the test statistic becomes

$$\begin{aligned} t^2 = \left(\frac{\bar{X} - \mu_0}{s/\sqrt{n}}\right)^2 &= (\bar{X} - \mu_0)\left[\frac{s^2}{n}\right]^{-1}(\bar{X} - \mu_0) \\ &= n(\bar{X} - \mu_0)(s^2)^{-1}(\bar{X} - \mu_0) \sim t^2(n-1) = F(1, n-1) \end{aligned}$$

Given a sample of n observations $x_1, x_2, \cdots, x_n$, H_0 should be rejected, that μ_0 is a plausible value for μ, if the observed

$$|t| = \left|\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right|$$

exceeds $t_{\alpha/2}(n-1)$ or if the observed $t^2 = n(\bar{x}-\mu_0)[s^2]^{-1}(\bar{x}-\mu_0) > t^2_{\alpha/2}(n-1) = F_\alpha(1, n-1)$.

Multivariate case:

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ be a random sample from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The hypothesis to be tested is $H_0\boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. To test it, the squared statistical distance from $\bar{\boldsymbol{X}}$ to $\boldsymbol{\mu}_0$ is considered. Thus, the test statistic which is analog of the univariate t^2 is

$$T^2 = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0) \sim \frac{(n-1)p}{n-p}F(p, n-p)$$

where

$$\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i \text{ and } \boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'.$$

This test statistic is called Hotelling's T^2 statistic. If T^2 is "too large", i.e., $\bar{\boldsymbol{X}}$ is "too far" from $\boldsymbol{\mu}_0$, then $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is rejected which means $\boldsymbol{\mu}_0$ is not a plausible value for $\boldsymbol{\mu}$.

If n independent observation vectors $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\cdots$, $\boldsymbol{x}_n$ are collected, then $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is rejected if $T^2 = n(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0) > c^*$ where $c^* = \dfrac{(n-1)p}{n-p}F_\alpha(p, n-p)$.

Example 4.1. Laboratory analysis of two different nutrients (A and B) for each of a sample of size $n = 10$ of the same food (in mg per 100 gram portion) revealed the following.

A	3.17	3.45	3.73	1.82	4.39	2.91	3.54	4.09	2.85	2.05
B	3.45	2.35	5.09	3.88	3.64	4.63	2.88	3.98	3.74	4.36

Does it appear that the sample come from a food with mean nutrient amount vector $\boldsymbol{\mu}_0 = (3, 5)'$?

Summary statistics: $p = 2$, $n = 10$

$\bar{x}_j = \dfrac{1}{n}\sum_{i=1}^{n} x_{ij};\ j = 1, 2$

$$\Rightarrow \bar{x}_1 = \frac{1}{10}\sum_{i=1}^{10} x_{i1} = \frac{1}{10}(3.17 + 3.45 + \cdots + 2.05) = 3.20$$

$$\Rightarrow \bar{x}_2 = \frac{1}{10}\sum_{i=1}^{10} x_{i2} = \frac{1}{10}(3.45 + 2.35 + \cdots + 4.36) = 3.80$$

Thus, the sample mean vector is: $\bar{\boldsymbol{x}} = \begin{bmatrix} 3.20 \\ 3.80 \end{bmatrix}$.

$$s_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k);\ j, k = 1, 2$$

$$\Rightarrow s_{11} = \frac{1}{10-1}\sum_{i=1}^{10}(x_{i1} - \bar{x}_1)^2 = 0.678 \text{ and } s_{22} = \frac{1}{10-1}\sum_{i=1}^{10}(x_{i2} - \bar{x}_2)^2 = 0.645$$

$$\Rightarrow s_{12} = \frac{1}{10-1}\sum_{i=1}^{10}(x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) = -0.109$$

Thus, the sample covariance matrix is:

$$\boldsymbol{S} = \begin{bmatrix} 0.678 & -0.109 \\ -0.109 & 0.645 \end{bmatrix} \Rightarrow \boldsymbol{S}^{-1} = \begin{bmatrix} 1.517 & 0.257 \\ 0.257 & 1.594 \end{bmatrix}$$

1. Hypothesis: $H_0 : \boldsymbol{\mu} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$ vs $H_1 : \boldsymbol{\mu} \neq \begin{bmatrix} 3 \\ 5 \end{bmatrix}$.

2. Critical value: $c^* = \frac{(n-1)p}{n-p}F_\alpha(p, n-p) = \frac{(10-1)2}{10-2}F_{0.05}(2, 10-2) = 2.25(4.459) = 10.033$.

3. Test statistic: $T^2 = n(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)$

$$\Rightarrow T^2 = 10\begin{bmatrix} 3.20-3 \\ 3.80-5 \end{bmatrix}' \begin{bmatrix} 1.517 & 0.257 \\ 0.257 & 1.594 \end{bmatrix} \begin{bmatrix} 3.20-3 \\ 3.80-5 \end{bmatrix}$$
$$= 10\begin{bmatrix} 0.2 & -1.2 \end{bmatrix} \begin{bmatrix} 1.517 & 0.257 \\ 0.257 & 1.594 \end{bmatrix} \begin{bmatrix} 0.2 \\ -1.2 \end{bmatrix}$$
$$= 22.322$$

4. Since $T^2 = 22.322 > c^* = 10.033$, H_0 is rejected. Thus, the sample does not appear to come from a food with mean nutrient $[3, 5]'$ at 5% level of significance.

4.2 Confidence Region for the Mean Vector $\boldsymbol{\mu}$

Ordinarily, instead of testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, it is preferable to find regions of $\boldsymbol{\mu}$ values that are plausible in the light of the observed data.

Univariate case:

For a random sample of n observations x_1, x_2, $\cdots$, x_n is drawn from a normal population with mean μ and variance σ^2, the $(1-\alpha)100\%$ confidence interval for μ is given by

$$\left|\frac{\bar{x} - \mu}{s/\sqrt{n}}\right| \leq t_{\alpha/2}(n-1)$$

which is equivalent to

$$t^2 = \frac{(\bar{x} - \mu)^2}{s^2/n} = n(\bar{x} - \mu)(s^2)^{-1}(\bar{x} - \mu) \leq F_\alpha(1, n-1).$$

Multivariate case:

A $(1-\alpha)100\%$ confidence region for the p dimensional multivariate normal population with mean $\boldsymbol{\mu}$ is given by

$$n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) \leq \frac{(n-1)p}{n-p}F_\alpha(p, n-p).$$

The confidence region is an ellipsoid centered at the sample mean vector $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_p)'$. This implies, the boundary of the ellipsoid is

$$(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) = \frac{c^*}{n} \text{ where } c^* = \frac{(n-1)p}{(n-p)}F_\alpha(p, n-p).$$

Beginning at the center $\bar{\boldsymbol{x}}$, the half lengths of the axes are given by

$$\sqrt{\lambda_j c} = \sqrt{\lambda_j \frac{c^*}{n}} = \sqrt{\lambda_j \frac{(n-1)p}{(n-p)n}F_\alpha(p, n-p)}$$

in the direction of $\boldsymbol{e}_j$ which is the normalized eigenvector corresponding to the eigenvalue λ_j; $j = 1, 2, \cdots, p$ of $\boldsymbol{S}$.

Example 4.2. Recall example 4.1. The 95% confidence region for $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ is given by

$$(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) \leq \frac{c^*}{n}$$

$$\Rightarrow \begin{bmatrix} 3.20 - \mu_1 \\ 3.80 - \mu_2 \end{bmatrix}' \begin{bmatrix} 1.517 & 0.257 \\ 0.257 & 1.594 \end{bmatrix} \begin{bmatrix} 3.20 - \mu_1 \\ 3.80 - \mu_2 \end{bmatrix} \leq \frac{10.033}{10}$$

This confidence region for $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ will be an equation of ellipse like the form

$$\ell_1(x_1 - \mu_1)^2 + \ell_2(x_2 - \mu_2)^2 + \ell_3(x_1 - \mu_1)(x_2 - \mu_2) \leq \ell_4.$$

For all points inside the ellipse (satisfying the equation), H_0 will not be rejected. For example, we can easily check that $\boldsymbol{\mu} = (3, 5)'$ does not lie in the region.

Let us plot of the confidence region. First, let us find the eigenvalues of $\boldsymbol{S}$. That is,

$$|\boldsymbol{S} - \lambda\boldsymbol{I}| = 0 \Rightarrow \begin{vmatrix} 0.678 - \lambda & -0.109 \\ -0.109 & 0.645 - \lambda \end{vmatrix} = 0$$

This gives the quadratic equation $\lambda^2 - 1.323\lambda + 0.425 = 0$. Hence, the eigenvalues are $\lambda_1 = 0.774$ and $\lambda_2 = 0.550$.

Second, let us obtain the orientations (eigenvectors) associated with both eigenvalues. That is, $\boldsymbol{S}\boldsymbol{x}_j = \lambda_j\boldsymbol{x}_j;\ j = 1, 2$

For $\lambda_1 = 0.774$: $\boldsymbol{S}\boldsymbol{x}_1 = \lambda_1\boldsymbol{x}_1$

$$\Rightarrow \begin{bmatrix} 0.678 & -0.109 \\ -0.109 & 0.645 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = 0.774 \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} \Rightarrow \boldsymbol{x}_1 = \begin{bmatrix} 1.00 \\ -0.88 \end{bmatrix}$$

Thus, the orientation of the major axis is: $\boldsymbol{e}_1 = \begin{bmatrix} 0.751 \\ -0.661 \end{bmatrix}$.

For $\lambda_2 = 0.550$: $\boldsymbol{S}\boldsymbol{x}_2 = \lambda_1\boldsymbol{x}_2$

$$\Rightarrow \begin{bmatrix} 0.678 & -0.109 \\ -0.109 & 0.645 \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} = 0.550 \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} \Rightarrow \boldsymbol{x}_2 = \begin{bmatrix} 1.000 \\ 1.174 \end{bmatrix}$$

Thus, the orientation of the minor axis is: $\boldsymbol{e}_1 = \begin{bmatrix} 0.648 \\ 0.761 \end{bmatrix}$.

The half lengths of the major and minor axes are $\sqrt{\lambda_1 c} = \sqrt{0.774(1.0033)} = 0.881$ and $\sqrt{\lambda_2 c} = \sqrt{0.550(1.0033)} = 0.743$, respectively.

Beginning at $\bar{\boldsymbol{x}} = [3.20, 3.80]'$, the plot is as follows.

Insertttheplothere

4.3 Simultaneous Confidence Statements

Once the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is rejected, then the component which is responsible for rejection has to be determined.

It would be erroneous to carry out univariate t tests for this purpose because the number of tests and the correlation among the responses would lead to a greatly different values of significance level (α) than the one chosen for the critical value of the t distribution. For example, let $\boldsymbol{X} \sim \mathcal{N}_6(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and assume each component mean equals to a specified value. There would be $p = 6$ univariate t-tests. Let $\alpha = 0.05$. Then, the probability of *not rejecting* the hypothesis of no difference from the specified value in each case would be $1 - 0.05 = 0.95$. If the tests are independent of each other, the probability of *not rejecting* H_0 in all of the 6 cases is $(0.95)(0.95)\cdots(0.95) = (0.95)^6 = 0.7351$. The probability of rejecting at least one hypothesis of no difference from the specified value is $1 - 0.7351 = 0.2649 = \alpha$ for a univariate t-test. This means that type I error is committed 26% of the time in testing all the 6 univariate tests. In general, the probability of committing type I error increases as the number of components is larger.

Simultaneous confidence statements are proposed to avoid such drawback of univariate confidence intervals by using linear combinations of the components. Recall, if $\boldsymbol{X}$ follows an $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the linear combination of the components of $\boldsymbol{X}$, $\boldsymbol{a'X} = a_1X_1 + a_2X_2 + \cdots + a_pX_p$, has also a normal distribution with mean $\boldsymbol{a'\mu}$ and variance $\boldsymbol{a'\Sigma a}$, that is, $\boldsymbol{a'X} \sim \mathcal{N}(\boldsymbol{a'\mu}, \boldsymbol{a'\Sigma a})$. Consequently, $\boldsymbol{a'\bar{X}} \sim \mathcal{N}(\boldsymbol{a'\mu}, \boldsymbol{a'\Sigma a}/n)$.

In constructing the simultaneous confidence statements, all the separate confidence intervals hold simultaneously a specified high confidence level (low significance level). That is, a simultaneous confidence interval uses linear combination of the components of $\boldsymbol{\mu}$ which is given by a set of $\boldsymbol{a'\mu}$ values such that the observed t^2 is relatively small for *all* choices of $\boldsymbol{a}$.

Then, a $(1-\alpha)100\%$ simultaneous confidence interval for $\boldsymbol{a'\mu}$ is

$$\frac{(\boldsymbol{a'\bar{x}} - \boldsymbol{a'\mu})^2}{\boldsymbol{a'Sa}/n} \leq c^* \Rightarrow \left|\frac{\boldsymbol{a'\bar{x}} - \boldsymbol{a'\mu}}{\sqrt{\boldsymbol{a'Sa}/n}}\right| \leq \sqrt{c^*} \Rightarrow (\boldsymbol{a'\bar{x}} \pm \sqrt{c^*}\sqrt{\boldsymbol{a'Sa}/n})$$

where $\boldsymbol{a'\bar{x}}$ is an estimate of $\boldsymbol{a'\mu}$, and $\boldsymbol{a'Sa}/n$ is an estimate of $\text{cov}(\boldsymbol{a'\bar{x}})$.

In particular, if $\boldsymbol{a'} = (0, 0, \cdots, \underbrace{1}_{j^{th} \text{ position}}, \cdots, 0)$, then the confidence interval for $\boldsymbol{a'\mu} = \mu_j$ is

$\left(\bar{x}_j \pm \sqrt{c^*}\sqrt{\frac{s_{jj}}{n}}\right)$ where $c^* = \frac{(n-1)p}{n-p}F_\alpha(p, n-p)$.

Example 4.3. Consider again example 4.1. Find the 95% confidence interval for mean nutrient A and B. The sample mean vector and sample variance-covariance matrix, respectively, were

$$\bar{\boldsymbol{x}} = \begin{bmatrix} 3.20 \\ 3.80 \end{bmatrix} \text{ and } \boldsymbol{S} = \begin{bmatrix} 0.678 & -0.109 \\ -0.109 & 0.645 \end{bmatrix}.$$

Also, the critical value for the Hotelling's T^2 was $c^* = \frac{(10-1)2}{10-2}F_{0.05}(2, 10-2) = 10.033$.

A 95% simultaneous confidence interval interval for μ_1 is

$$\left(\bar{x}_1 \pm \sqrt{c^*}\sqrt{\frac{s_{11}}{n}}\right) = \left(3.20 \pm \sqrt{10.033}\sqrt{\frac{0.678}{10}}\right) = (2.375, 4.025).$$

Similarly, a 95% simultaneous confidence interval interval for μ_2 is

$$\left(\bar{x}_2 \pm \sqrt{c^*}\sqrt{\frac{s_{22}}{n}}\right) = \left(3.80 \pm \sqrt{10.033}\sqrt{\frac{0.645}{10}}\right) = (2.995, 4.650).$$

Note that $\mu_{01} = 3$ is found inside in the confidence interval for μ_1 while $\mu_{02} = 5$ is found outside the confidence interval for μ_2. Hence, the second component (nutrient B) is responsible for the rejection of $H_0 : \boldsymbol{\mu} = (3, 5)'$.

4.4 The Bonferroni Method of Multiple Comparisons

The Bonferroni confidence interval makes an adjustment on the univariate t-test critical value, not to increase type I error, by considering the total number of confidence intervals required. The $(1-\alpha)100\%$ Bonferroni confidence interval for $\boldsymbol{\mu}_j$ is

$$\left(\bar{x}_j \pm t_{\alpha/2p}(n-1)\sqrt{\frac{s_{jj}}{n}}\right)$$

where p is the number of confidence intervals required.

Example 4.4. Find the Bonferroni confidence interval based on the data given in example 4.1. $t_{\frac{0.05}{2(2)}}(10-1) = t_{0.0125}(9) = 3.111$.

$$\mu_1 : \left(3.2 \pm 3.111\sqrt{\frac{0.678}{10}}\right) = (2.39, 4.01) \text{ and } \mu_2 : \left(3.8 \pm 3.111\sqrt{\frac{0.645}{10}}\right) = (3.01, 4.59)$$

Again using the Bonferroni confidence interval, the second component (nutrient B) is responsible for the rejection of $H_0 : \boldsymbol{\mu} = (3, 5)'$.

4.5 Likelihood-Ratio Test

Likelihood-ratio test is a general principle for constructing test procedures. It is the ratio of the restricted likelihood function to the unrestricted likelihood function.

Recall for a random sample $\boldsymbol{X}_i$; $i = 1, 2, \cdots, n$ from an $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the likelihood function is:

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{np}{2}}|\boldsymbol{\Sigma}|^{\frac{n}{2}}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right].$$

Also recall the ML estimate of $\boldsymbol{\mu}$ is $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}$ and that of $\boldsymbol{\Sigma}$ is $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$.

The exponent of the likelihood function can be

$$\begin{aligned}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}) &= \operatorname{tr}\left\{\sum_{i=1}^{n}\underbrace{(\boldsymbol{x}_i-\boldsymbol{\mu})'}_{\boldsymbol{A}_{1\times p}}\underbrace{\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})}_{\boldsymbol{B}_{p\times 1}}\right\}\\ &= \operatorname{tr}\left\{\sum_{i=1}^{n}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})'\right\} \text{ as } \operatorname{tr}(AB)=\operatorname{tr}(BA)\\ &= \operatorname{tr}\left\{\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})'\right\}.\end{aligned}$$

Thus,

$$\ell(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{np}{2}}|\boldsymbol{\Sigma}|^{\frac{n}{2}}}\exp\left[-\frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})'\right\}\right].$$

The (unrestricted) maximum of the likelihood function is

$$\begin{aligned}\ell(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}}) &= \frac{1}{(2\pi)^{\frac{np}{2}}|\hat{\boldsymbol{\Sigma}}|^{\frac{n}{2}}}\exp\left[-\frac{1}{2}\operatorname{tr}\left\{\hat{\boldsymbol{\Sigma}}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'\right\}\right]\\ &= \frac{1}{(2\pi)^{\frac{np}{2}}|\hat{\boldsymbol{\Sigma}}|^{\frac{n}{2}}}\exp\left[-\frac{1}{2}\operatorname{tr}\left(n\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}\right)\right]\\ &= \frac{1}{(2\pi)^{\frac{np}{2}}|\hat{\boldsymbol{\Sigma}}|^{\frac{n}{2}}}\exp\left[-\frac{1}{2}n\operatorname{tr}\left(\boldsymbol{I_p}\right)\right]\\ &= \frac{1}{(2\pi)^{\frac{np}{2}}|\hat{\boldsymbol{\Sigma}}|^{\frac{n}{2}}}\exp\left[-\frac{1}{2}np\right].\end{aligned}$$

When the null hypothesis holds, there is no need of searching for $\boldsymbol{\mu}$ because it is given as fixed. Hence, under $H_0:\boldsymbol{\mu}=\boldsymbol{\mu}_0$, the restricted likelihood function is

$$\begin{aligned}\ell(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0) &= \frac{1}{(2\pi)^{\frac{np}{2}}|\boldsymbol{\Sigma}_0|^{\frac{n}{2}}}\exp\left[-\frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Sigma}_0^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu}_0)(\boldsymbol{x}_i-\boldsymbol{\mu}_0)'\right\}\right]\\ &= \frac{1}{(2\pi)^{\frac{np}{2}}|\boldsymbol{\Sigma}_0|^{\frac{n}{2}}}\exp\left[-\frac{1}{2}\operatorname{tr}\left(n\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_0\right)\right]\\ &= \frac{1}{(2\pi)^{\frac{np}{2}}|\boldsymbol{\Sigma}_0|^{\frac{n}{2}}}\exp\left[-\frac{1}{2}n\operatorname{tr}\left(\boldsymbol{I_p}\right)\right]\\ &= \frac{1}{(2\pi)^{\frac{np}{2}}|\boldsymbol{\Sigma}_0|^{\frac{n}{2}}}\exp\left[-\frac{1}{2}np\right].\end{aligned}$$

Therefore, the likelihood-ratio is

$$\Lambda = \frac{\ell(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)}{\ell(\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}})} = \left[\frac{|\hat{\boldsymbol{\Sigma}}|}{|\boldsymbol{\Sigma}_0|}\right]^{\frac{n}{2}}$$

$$\Rightarrow \Lambda^{\frac{2}{n}} = \frac{|\hat{\boldsymbol{\Sigma}}|}{|\boldsymbol{\Sigma}_0|}.$$

This likelihood-ratio test statistic $\Lambda^{\frac{2}{n}}$ is called Wilks' Lambda. The null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ should be rejected if the value of Λ is too small, that is, if

$$\Lambda = \left[\frac{|\hat{\boldsymbol{\Sigma}}|}{|\boldsymbol{\Sigma}_0|}\right]^{\frac{n}{2}} < c_\alpha$$

where c_α is the lower $(100\alpha)^{th}$ percentile of the distribution of Λ. But,

$$\Lambda^{\frac{2}{n}} = \left[1 + \frac{T^2}{n-1}\right]^{-1}$$

where $T^2 \sim \dfrac{(n-1)p}{n-p}F(p, n-p)$. Rejecting H_0 for small values of $\Lambda^{\frac{2}{n}}$ is equivalent to rejecting H_0 for large values of T^2.

4.6 Large Sample Inference about $\boldsymbol{\mu}$

When the sample size is large, tests of hypothesis and confidence intervals can be constructed without the assumption of a normal population.

Univariate case:

Suppose a random sample of large size n is drawn from any population with mean μ and variance σ^2. The test statistic for testing $H_0 : \mu = \mu_0$ is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Rejecting H_0 when $|Z|$ is large is equivalent to rejecting H_0 if Z^2 is large. That is,

$$\begin{aligned} Z^2 = \left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right)^2 &= (\bar{X} - \mu_0)\left[\frac{\sigma^2}{n}\right]^{-1}(\bar{X} - \mu_0) \\ &= n(\bar{X} - \mu_0)(\sigma^2)^{-1}(\bar{X} - \mu_0) \sim \chi^2(1) \end{aligned}$$

If σ is unknown, s is used instead.

Multivariate case:

All large sample multivariate inferences are based on the χ^2 distribution. When $n-p$ is large, $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ will be rejected if $T^2 = n(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0) > \chi^2_\alpha(p)$ since $\dfrac{(n-1)p}{n-p}F_\alpha(p, n-p)$ and $\chi^2_\alpha(p)$ are approximately equal for large sample size.

The $(1-\alpha)100\%$ simultaneous and Bonferroni confidence intervals for $\boldsymbol{a}'\boldsymbol{\mu}$ are given by $(\boldsymbol{a}'\bar{\boldsymbol{x}} \pm \sqrt{\chi^2_\alpha(p)}\sqrt{\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}/n})$ and $(\boldsymbol{a}'\bar{\boldsymbol{x}} \pm z_{\alpha/2p}\sqrt{\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}/n})$, respectively.

Chapter 5

Comparison of Several Multivariate Means

5.1 Dependent Samples

5.1.1 Paired Comparison

In paired comparison, the presence and absence of a single treatment or two treatments are compared by assigning both treatments to the *same* (e.g., persons) or *identical* (e.g., plots) experimental units. The paired responses are then analysed by computing their differences.

Univariate case:

Let X_{i1} and X_{i2} denote the responses to treatment I (response before treatment) and to treatment II (after treatment) for the i^{th}; $i = 1, 2, \cdots, n$ trial (experimental unit). That is, (X_{i1}, X_{i2}) are responses recorded on the i^{th} pair of like units. The differential effects of the treatments is $D_i = X_{i1} - X_{i2}$; $i = 1, 2, \cdots, n$.

Let the differences D_i; $i = 1, 2, \cdots, n$ represent independent observations from $\mathcal{N}(\mu_d, \sigma_d^2)$. Thus, $\bar{D} \sim \mathcal{N}(\mu_d, \sigma_d^2/n)$. The hypothesis to be tested is $H_0 : \mu_d = 0$ versus $H_1 : \mu_d \neq 0$. Then, the test statistic is

$$t = \frac{\bar{D} - \mu_d}{s_d/\sqrt{n}} \sim t(n-1)$$

where $\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$ and $s_d = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2$.

Consequently, H_0 should be rejected if the observed $|t| > t_{\alpha/2}(n-1)$.

Multivariate case:

Given p responses, 2 treatments and n experimental units. Let X_{1ij} denote the j^{th} response of the i^{th} unit to treatment I (response before treatment) and let X_{2ij} denote the j^{th} response of the i^{th} unit to treatment II (response after treatment).

Pre-treatment matrix			
Var 1	Var 2	$\cdots$	Var p
X_{111}	X_{112}	$\cdots$	X_{11p}
X_{121}	X_{122}	$\cdots$	X_{12p}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
X_{1n1}	X_{1n2}	$\cdots$	X_{1np}

Post-treatment matrix			
Var 1	Var 2	$\cdots$	Var p
X_{211}	X_{212}	$\cdots$	X_{21p}
X_{221}	X_{222}	$\cdots$	X_{22p}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
X_{2n1}	X_{2n2}	$\cdots$	X_{2np}

The analysis is performed on the differences (before treatment - after treatment) of the type $D_{ij} = D_{1ij} - D_{2ij}$, $j = 1, 2, \cdots, p$; $i = 1, 2, \cdots, n$.

Let the differences $\boldsymbol{D}_1, \boldsymbol{D}_2, \cdots, \boldsymbol{D}_n$ represent independent observation vectors from $\mathcal{N}_p(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d)$. Thus, $\bar{\boldsymbol{D}} \sim \mathcal{N}_p(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d/n)$. The hypothesis of interest is $H_0 : \boldsymbol{\mu}_d = \boldsymbol{0}$ (no treatment effect for all p components) versus $H_1 : \boldsymbol{\mu}_d \neq \boldsymbol{0}$. Then the test statistic is

$$T^2 = n(\bar{\boldsymbol{D}} - \boldsymbol{\mu}_d)'(\boldsymbol{S}_d)^{-1}(\bar{\boldsymbol{D}} - \boldsymbol{\mu}_d) \sim \frac{(n-1)p}{n-p} F(p, n-p)$$

where $\bar{\boldsymbol{D}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{D}_i$ and $\boldsymbol{S}_d = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{D}_i - \bar{\boldsymbol{D}})(\boldsymbol{D}_i - \bar{\boldsymbol{D}})'$.

Given the observed differences $\boldsymbol{d}_i' = (d_{i1}, d_{i2}, \cdots, d_{ip})$; $i = 1, 2, \cdots, n$, $H_0 : \boldsymbol{\mu}_d = \boldsymbol{0}$ is rejected if the observed

$$T^2 = n\bar{\boldsymbol{d}}'(\boldsymbol{S}_d)^{-1}\bar{\boldsymbol{d}} > \frac{(n-1)p}{n-p} F_\alpha(p, n-p)$$

where $\bar{\boldsymbol{d}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{d}_i$ and $\boldsymbol{S}_d = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{d}_i - \bar{\boldsymbol{d}})(\boldsymbol{d}_i - \bar{\boldsymbol{d}})'$.

Note that $\bar{\boldsymbol{d}} = \begin{bmatrix} \bar{d}_1 \\ \bar{d}_2 \\ \vdots \\ \bar{d}_p \end{bmatrix}$ and $\boldsymbol{S}_d = \begin{bmatrix} s_{d_1d_1} & s_{d_1d_2} & \cdots & s_{d_1d_p} \\ s_{d_2d_1} & s_{d_2d_2} & \cdots & s_{d_2d_p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{d_pd_1} & s_{d_pd_2} & \cdots & s_{d_pd_p} \end{bmatrix}$

A $(1-\alpha)100\%$ confidence region for $\boldsymbol{\mu}_d$ is $n(\bar{\boldsymbol{d}} - \boldsymbol{\mu}_d)'(\boldsymbol{S}_d)^{-1}(\bar{\boldsymbol{d}} - \boldsymbol{\mu}_d) \leq c^*$ which is an ellipsoid passing through $\bar{\boldsymbol{d}}$. To plot the confidence ellipsoid, the sample covariance matrix of the sample differences, $\boldsymbol{S}_d$, is used.

A $(1-\alpha)100\%$ simultaneous confidence interval for a linear combination $\boldsymbol{a}'\boldsymbol{\mu}_d$ is given by

$$\left(\boldsymbol{a}'\bar{\boldsymbol{d}} \pm \sqrt{c^*}\sqrt{\boldsymbol{a}'\boldsymbol{S}_d\boldsymbol{a}/n}\right)$$

where $\boldsymbol{a}'\bar{\boldsymbol{d}}$ is an estimate of $\boldsymbol{a}'\boldsymbol{\mu}_d$, and $\boldsymbol{a}'\boldsymbol{S}_d\boldsymbol{a}/n$ is an estimate of $\text{cov}(\boldsymbol{a}'\bar{\boldsymbol{d}})$.

Particularly, a $(1-\alpha)100\%$ simultaneous confidence interval for the individual mean differences μ_{d_j}'s are given by

$$\left(\bar{d}_j \pm \sqrt{c^*}\sqrt{\frac{s_{d_jd_j}}{n}}\right);\ j = 1, 2, \cdots, p.$$

Also, a $(1-\alpha)100\%$ Bonferroni confidence interval for the individual mean differences μ_{d_j}'s are given by

$$\left(\bar{d}_j \pm t_{\alpha/2p}(n-1)\sqrt{\frac{s_{d_j d_j}}{n}}\right);\ j = 1, 2, \cdots, p.$$

Example 5.1. It is felt that three drugs (X_1, X_2 and X_3) may lead to changes in the level of a certain biochemical compound found in the brain. Thirty mice of the same stain were randomly divided into three groups and received the drugs. The amount of the compound (in micrograms per gram of brain tissue) is recorded before and after the treatments. The responses are in given in the following table. Test the hypothesis of no treatment effect at 5% level of significance.

Before treatment			After treatment		
x_{1i1}	x_{1i2}	x_{1i3}	x_{2i1}	x_{2i2}	x_{2i3}
1.21	0.61	0.70	1.26	0.50	0.81
0.92	0.43	0.71	1.07	0.39	0.69
0.80	0.35	0.71	1.33	0.24	0.70
0.85	0.48	0.68	1.39	0.37	0.72
0.98	0.42	0.71	1.38	0.42	0.71
1.15	0.52	0.72	0.98	0.49	0.70
1.10	0.50	0.75	1.41	0.41	0.70
1.02	0.53	0.70	1.30	0.47	0.67
1.18	0.45	0.70	1.22	0.29	0.68
1.09	0.40	0.69	1.00	0.30	0.70

The necessary calculations are obtained as follows. Here $d^*_{ij} = d_{ij} - \bar{d}_j$. Also the last row is the sum.

d_{i1}	d_{i2}	d_{i3}	d^{*2}_{i1}	d^{*2}_{i2}	d^{*2}_{i3}	$d^*_{i1}d^*_{i2}$	$d^*_{i1}d^*_{i3}$	$d^*_{i2}d^*_{i3}$
-0.050	0.110	-0.110	0.023716	0.000841	0.011881	0.004466	-0.016786	-0.003161
-0.150	0.040	0.020	0.002916	0.001681	0.000441	-0.002214	0.001134	-0.000861
-0.530	0.110	0.010	0.106276	0.000841	0.000121	-0.009454	-0.003586	0.000319
-0.540	0.110	-0.040	0.112896	0.000841	0.001521	-0.009744	0.013104	-0.001131
-0.400	0.000	0.000	0.038416	0.006561	0.000001	0.015876	-0.000196	-0.000081
0.170	0.030	0.020	0.139876	0.002601	0.000441	-0.019074	0.007854	-0.001071
-0.310	0.090	0.050	0.011236	0.000081	0.002601	-0.000954	-0.005406	0.000459
-0.280	0.060	0.030	0.005776	0.000441	0.000961	0.001596	-0.002356	-0.000651
-0.040	0.160	0.020	0.026896	0.006241	0.000441	0.012956	0.003444	0.001659
0.090	0.100	-0.010	0.086436	0.000361	0.000081	0.005586	-0.002646	-0.000171
-2.040	0.810	-0.010	0.554440	0.020490	0.018490	-0.000960	-0.005440	-0.004690

$$\bar{d}_j = \frac{1}{n}\sum_{i=1}^{n} d_{ij} = \frac{1}{10}\sum_{i=1}^{10} d_{ij};\ j = 1, 2, 3$$

$$\Rightarrow \bar{d}_1 = \frac{1}{10}\sum_{i=1}^{10} d_{i1} = \frac{1}{10}(-0.204) = -0.204$$

$$\Rightarrow \bar{d}_2 = \frac{1}{10}\sum_{i=1}^{10} d_{i2} = \frac{1}{10}(0.810) = 0.081$$

$$\Rightarrow \bar{d}_3 = \frac{1}{10}\sum_{i=1}^{10} d_{i3} = \frac{1}{10}(-0.01) = -0.001$$

$$\Rightarrow \bar{\boldsymbol{d}} = \begin{bmatrix} -0.204 \\ 0.081 \\ -0.001 \end{bmatrix}$$

$$s_{d_j d_k} = \frac{1}{n-1}\sum_{i=1}^{n}(d_{ij} - \bar{d}_j)(d_{ik} - \bar{d}_k) = \frac{1}{10-1}\sum_{i=1}^{10}(d_{ij} - \bar{d}_j)(d_{ik} - \bar{d}_k);\ j,k = 1,2,3$$

$$\Rightarrow s_{d_1 d_1} = \frac{1}{9}\sum_{i=1}^{10}(d_{i1} - \bar{d}_1)^2 = \frac{1}{9}(0.55444) = 0.06160$$

$$\Rightarrow s_{d_2 d_2} = \frac{1}{9}\sum_{i=1}^{10}(d_{i2} - \bar{d}_2)^2 = \frac{1}{9}(0.02049) = 0.00228$$

$$\Rightarrow s_{d_3 d_3} = \frac{1}{9}\sum_{i=1}^{10}(d_{i3} - \bar{d}_3)^2 = \frac{1}{9}(0.01849) = 0.00205$$

$$\Rightarrow s_{d_1 d_2} = \frac{1}{9}\sum_{i=1}^{10}(d_{i1} - \bar{d}_1)(d_{i2} - \bar{d}_2) = \frac{1}{9}(-0.00096) = -0.00011$$

$$\Rightarrow s_{d_1 d_3} = \frac{1}{9}\sum_{i=1}^{10}(d_{i1} - \bar{d}_1)(d_{i3} - \bar{d}_3) = \frac{1}{9}(-0.00544) = -0.00060$$

$$\Rightarrow s_{d_2 d_3} = \frac{1}{9}\sum_{i=1}^{10}(d_{i2} - \bar{d}_2)(d_{i3} - \bar{d}_3) = \frac{1}{9}(-0.00469) = -0.00052$$

$$\boldsymbol{S}_d = \begin{bmatrix} 0.06160 & -0.00011 & -0.00060 \\ -0.00011 & 0.00228 & -0.00052 \\ -0.00060 & -0.00052 & 0.00205 \end{bmatrix} \Rightarrow \boldsymbol{S}_d^{-1} = \begin{bmatrix} 16.28866 & 1.98818 & 5.27173 \\ 1.98818 & 465.77088 & 118.72867 \\ 5.27173 & 118.72867 & 519.46436 \end{bmatrix}$$

The hypothesis to be tested is $H_0 : \boldsymbol{\mu}_d = \mathbf{0}$ vs $H_1 : \boldsymbol{\mu}_d \neq \mathbf{0}$.

$$T^2 = n\bar{\boldsymbol{d}}'(\boldsymbol{S}_d)^{-1}\bar{\boldsymbol{d}}$$

$$\Rightarrow T^2 = 10\begin{bmatrix} -0.204 \\ 0.081 \\ -0.001 \end{bmatrix}'\begin{bmatrix} 16.28866 & 1.98818 & 5.27173 \\ 1.98818 & 465.77088 & 118.72867 \\ 5.27173 & 118.72867 & 519.46436 \end{bmatrix}\begin{bmatrix} -0.204 \\ 0.081 \\ -0.001 \end{bmatrix} = 36.515$$

The critical value is $c^* = \frac{(n-1)p}{n-p}F_\alpha(p, n-p) = \frac{(10-1)3}{10-3}F_{0.05}(3, 10-3) = 16.779$. There is a significant treatment effect at 5% level of significance.

The next question is which of the three drugs (X_1, X_2 or X_3) leads to changes in the level of the biochemical compound found in the brain? To answer this question, the simultaneous confidence intervals for the individual mean differences μ_{d_j} need to be constructed, which is given by

$$\left(\bar{d}_j \pm \sqrt{c^*}\sqrt{\frac{s_{d_j d_j}}{n}}\right);\ j = 1, 2, 3$$

Hence, the 95% confidence intervals are:

$$\mu_{d_1}: \left(\bar{d}_1 \pm \sqrt{c^*}\sqrt{\frac{s_{d_1 d_1}}{n}}\right) = \left(-0.204 \pm \sqrt{16.779}\sqrt{\frac{0.06160}{10}}\right) = (-0.5255, 0.1175)$$

$$\mu_{d_2}: \left(\bar{d}_2 \pm \sqrt{c^*}\sqrt{\frac{s_{d_2 d_2}}{n}}\right) = \left(0.081 \pm \sqrt{16.779}\sqrt{\frac{0.00228}{10}}\right) = (0.0191, 0.1429)$$

$$\mu_{d_3}: \left(\bar{d}_3 \pm \sqrt{c^*}\sqrt{\frac{s_{d_3 d_3}}{n}}\right) = \left(-0.001 \pm \sqrt{16.779}\sqrt{\frac{0.00205}{10}}\right) = (-0.0596, 0.0576)$$

The confidence interval for μ_{d_2} does not include zero. Thus, $H_0 : \boldsymbol{\mu}_d = \mathbf{0}$ was rejected due to the second component (X_2). In other words, it is the second drug (X_2) that led to a significant change in the level of the biochemical compound found in the brain at 5% level of significance.

5.1.2 A Repeated Measures Design for Comparing Treatments

A repeated measures design is another generalization of the univariate t statistic in which q treatments are compared with respect to a *single* response measured from the same (identical) sampling units over time or space. Each experimental unit receives each treatment once over successive period of time. The name *repeated measures* stems from the fact that all treatments are administered to each unit.

Let X_{ik} be the response of the i^{th}; $i = 1, 2, \cdots, n$ unit to the k^{th}; $k = 1, 2, \cdots, q$ treatment.

Item	Treatment 1	Treatment 2	$\cdots$	Treatment q	
1	X_{11}	X_{12}	$\cdots$	X_{1q}	$\boldsymbol{X}_1'$
2	X_{21}	X_{22}	$\cdots$	X_{1q}	$\boldsymbol{X}_2'$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
n	X_{n1}	X_{n2}	$\cdots$	X_{nq}	$\boldsymbol{X}_q'$

The hypothesis of interest is whether $\mu_1 = \mu_2 = \cdots = \mu_q$ (no treatment effect). For comparative purposes, *contrasts* of the components of $\boldsymbol{\mu} = E(\boldsymbol{X}_i)$ are considered. These

could be

$$\underbrace{\begin{bmatrix} \mu_1-\mu_2 \\ \mu_1-\mu_3 \\ \mu_1-\mu_4 \\ \vdots \\ \mu_1-\mu_q \end{bmatrix}}_{(q-1)\times 1} = \underbrace{\begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & -1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 0 & -1 \end{bmatrix}}_{(q-1)\times q} \underbrace{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_q \end{bmatrix}}_{q\times 1} = \boldsymbol{A\mu}$$

or

$$\underbrace{\begin{bmatrix} \mu_1-\mu_2 \\ \mu_2-\mu_3 \\ \mu_3-\mu_4 \\ \vdots \\ \mu_{q-1}-\mu_q \end{bmatrix}}_{(q-1)\times 1} = \underbrace{\begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix}}_{(q-1)\times q} \underbrace{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_q \end{bmatrix}}_{q\times 1} = \boldsymbol{B\mu}$$

Since each row is a contrast and the $q-1$ rows are linearly independent, both $\boldsymbol{A}$ and $\boldsymbol{B}$ are *contrast matrices*. If $\boldsymbol{A\mu} = \boldsymbol{B\mu} = \boldsymbol{0}$, then $\mu_1 = \mu_2 = \cdots = \mu_q$. Hence, the hypothesis of no difference in treatments (equal treatment means) is $\boldsymbol{A\mu} = \boldsymbol{0}$ for any choice of contrast matrix $\boldsymbol{A}$.

Consider an $\mathcal{N}_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. If $\boldsymbol{A}$ is a contrast matrix, then $\boldsymbol{AX} \sim \mathcal{N}_{q-1}(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A'})$. Hence, $\boldsymbol{A\bar{X}} \sim \mathcal{N}_{q-1}(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A'}/n)$.

Therefore, for testing $H_0 : \boldsymbol{A\mu} = \boldsymbol{0}$ vs $H_1 : \boldsymbol{A\mu} \neq \boldsymbol{0}$, the T^2 test statistic, which does not depend on the particular choice of $\boldsymbol{A}$, is

$$T^2 = n(\boldsymbol{A\bar{X}} - \boldsymbol{A\mu})'(\boldsymbol{ASA'})^{-1}(\boldsymbol{A\bar{X}} - \boldsymbol{A\mu}) \sim \frac{(n-1)(q-1)}{n-(q-1)} F[q-1, n-(q-1)].$$

As usual, $H_0 : \boldsymbol{A\mu} = \boldsymbol{0}$ is rejected if the observed $T^2 = n(\boldsymbol{A\bar{x}})'(\boldsymbol{ASA'})^{-1}(\boldsymbol{A\bar{x}}) > c^*$ where $c^* = \dfrac{(n-1)(q-1)}{n-(q-1)} F_\alpha[q-1, n-(q-1)]$.

The $(1-\alpha)100\%$ simultaneous confidence interval for a single contrast $\boldsymbol{a'\mu}$ for any contrast vector $\boldsymbol{a}$ of interest are

$$(\boldsymbol{a'\bar{x}} \pm \sqrt{c^*}\sqrt{\boldsymbol{a'Sa}/n})$$

where $\boldsymbol{a'\bar{x}}$ is an estimate of $\boldsymbol{a'\mu}$, and $\boldsymbol{a'Sa}/n$ is an estimate of $\text{cov}(\boldsymbol{a'\bar{x}})$.

Particularly, the confidence interval for the difference of the j^{th} and k^{th} treatment means, $\mu_j - \mu_k$, is obtained by letting $\boldsymbol{a'} = (0, \cdots, 0, \underbrace{1}_{j^{th}\text{ position}}, 0, \cdots, 0, \underbrace{-1}_{k^{th}\text{ position}}, 0, \cdots, 0)$:

$$\left[(\bar{x}_j - \bar{x}_k) \pm \sqrt{c^*}\sqrt{\frac{s_{jj} - 2s_{jk} + s_{kk}}{n}}\right];\ j \neq k.$$

Also, a $(1-\alpha)100\%$ Bonferroni confidence interval for the difference in treatment means $\mu_j - \mu_k$ are given by

$$\left(\bar{d}_j \pm t_{\alpha/[q(q-1)]}(n-1)\sqrt{\frac{s_{d_j d_j}}{n}}\right);\ j = 1, 2, \cdots, p.$$

Example 5.2. A researcher conducted three indices measuring severity of heart attacks. The values of the indices for $n = 40$ heart-attack patients arriving at a hospital emergency room produced the following summary statistics.

$$\bar{\boldsymbol{x}} = \begin{bmatrix} 46.1 \\ 57.3 \\ 50.4 \end{bmatrix} \text{ and } \boldsymbol{S} = \begin{bmatrix} 101.3 & 63.0 & 71.0 \\ 63.0 & 80.2 & 55.6 \\ 71.0 & 55.6 & 97.4 \end{bmatrix}$$

Test the equality of the mean indices and judge the differences in pairs of mean indices.

Since there are $q = 3$ treatments, let $\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$. Then the hypothesis to be tested is

$$H_0 : \boldsymbol{A\mu} = \boldsymbol{0} \Rightarrow H_0 : \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 - \mu_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$H_1 : \boldsymbol{A\mu} \neq \boldsymbol{0} \Rightarrow H_1 : \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 - \mu_3 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The test statistic is $T^2 = n(\boldsymbol{A}\bar{\boldsymbol{x}})'(\boldsymbol{ASA}')^{-1}(\boldsymbol{A}\bar{\boldsymbol{x}})$.

$$\boldsymbol{A}\bar{\boldsymbol{x}} = \begin{bmatrix} -11.2 \\ -4.3 \end{bmatrix},\ \boldsymbol{ASA}' = \begin{bmatrix} 55.5 & 22.9 \\ 22.9 & 56.7 \end{bmatrix} \Rightarrow (\boldsymbol{ASA}')^{-1} = \begin{bmatrix} 0.02162 & -0.00873 \\ -0.00873 & 0.02116 \end{bmatrix}$$

$$T^2 = \begin{bmatrix} -11.2 & -4.3 \end{bmatrix} \begin{bmatrix} 0.02162 & -0.00873 \\ -0.00873 & 0.02116 \end{bmatrix} \begin{bmatrix} -11.2 \\ -4.3 \end{bmatrix} = 90.49$$

$$c^* = \frac{(n-1)(q-1)}{n-(q-1)} F_\alpha[q-1, n-(q-1)] = \frac{(40-1)(3-1)}{40-(3-1)} F_{0.05}[3-1, 40-(3-1)] = 6.66$$

Hence, $H_0 : \boldsymbol{A\mu} = \boldsymbol{0}$ is rejected. The mean indices are not all equal.

The 95% simultaneous confidence interval for $\mu_j - \mu_k$ is

$$\left[(\bar{x}_j - \bar{x}_k) \pm \sqrt{c^*}\sqrt{\frac{s_{jj} - 2s_{jk} + s_{kk}}{n}}\right];\ j \neq k$$

$$\mu_1 - \mu_2: \quad \left[(\bar{x}_1 - \bar{x}_2) \pm \sqrt{c^*}\sqrt{\frac{s_{11} - 2s_{12} + s_{22}}{n}}\right]$$

$$\left[-11.2 \pm \sqrt{6.66}\sqrt{\frac{101.3 - 2(63.0) + 80.2}{40}}\right] = (-14.23986, -8.16014)$$

$$\mu_1 - \mu_3: \quad \left[(\bar{x}_1 - \bar{x}_3) \pm \sqrt{c^*}\sqrt{\frac{s_{11} - 2s_{13} + s_{33}}{n}}\right]$$

$$\left[-4.3 \pm \sqrt{6.66}\sqrt{\frac{101.3 - 2(71.0) + 97.4}{40}}\right] = (-7.37255, -1.22745)$$

$$\mu_2 - \mu_3: \quad \left[(\bar{x}_2 - \bar{x}_3) \pm \sqrt{c^*}\sqrt{\frac{s_{22} - 2s_{23} + s_{33}}{n}}\right]$$

$$\left[6.9 \pm \sqrt{6.66}\sqrt{\frac{80.2 - 2(55.6) + 97.4}{40}}\right] = (3.57500, 10.22500)$$

All the intervals do not contain zero. Thus, all mean indices are significantly different from each other ($\mu_2 > \mu_3 > \mu_1$).

5.2 Independent Samples

5.2.1 Comparing Mean Vectors from Two Populations

Univariate case:

- $X_{11}, X_{12}, \cdots, X_{1n_1} \sim \mathcal{N}(\mu_1, \sigma_1^2)$
- $X_{21}, X_{22}, \cdots, X_{2n_2} \sim \mathcal{N}(\mu_2, \sigma_2^2)$
- The two samples are independent.

The hypothesis to be tested is $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_1$. Assuming $\sigma_1^2 = \sigma_2^2$, the test statistic is

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) s^2_{pooled}}} \sim t(n_1 + n_2 - 2)$$

where $s^2_{pooled} = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$. Reject H_0 if the observed $|t| > t_{\alpha/2}(n_1 + n_2 - 2)$.

Multivariate case:

- $\boldsymbol{X}_{11}, \boldsymbol{X}_{12}, \cdots, \boldsymbol{X}_{1n_1} \sim \mathcal{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$
- $\boldsymbol{X}_{21}, \boldsymbol{X}_{22}, \cdots, \boldsymbol{X}_{2n_2} \sim \mathcal{N}_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$.
- $\boldsymbol{X}_{11}, \boldsymbol{X}_{12}, \cdots, \boldsymbol{X}_{1n_1}$ are independent of $\boldsymbol{X}_{21}, \boldsymbol{X}_{22}, \cdots, \boldsymbol{X}_{2n_1}$.

The data layout,

Population 1				
X_{111}	X_{112}	$\cdots$	X_{11p}	$\rightarrow \boldsymbol{X}'_{11}$
X_{121}	X_{122}	$\cdots$	X_{12p}	$\rightarrow \boldsymbol{X}'_{12}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
X_{1n_11}	X_{1n_12}	$\cdots$	X_{1n_1p}	$\rightarrow \boldsymbol{X}'_{1n_1}$

Population 2				
X_{211}	X_{212}	$\cdots$	X_{21p}	$\rightarrow \boldsymbol{X}'_{21}$
X_{221}	X_{222}	$\cdots$	X_{22p}	$\rightarrow \boldsymbol{X}'_{22}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
X_{2n_21}	X_{2n_22}	$\cdots$	X_{2n_2p}	$\rightarrow \boldsymbol{X}'_{2n_2}$

If n_1 and n_2 are small, then both populations should be multivariate normal and they should have the same covariance matrix (i.e., $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$). The second assumption is much stronger than its univariate counterpart because the several pairs of variances and covariances must be nearly equal.

If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, then both $\boldsymbol{S}_1 = \dfrac{1}{n_1 - 1}\sum_{i=1}^{n_1}(\boldsymbol{X}_{1i} - \bar{\boldsymbol{X}}_1)(\boldsymbol{X}_{1i} - \bar{\boldsymbol{X}}_1)'$ and $\boldsymbol{S}_2 = \dfrac{1}{n_2 - 1}\sum_{i=1}^{n_2}(\boldsymbol{X}_{2i} - \bar{\boldsymbol{X}}_2)(\boldsymbol{X}_{2i} - \bar{\boldsymbol{X}}_2)'$ estimate $\boldsymbol{\Sigma}$. Consequently, both samples can be pooled to estimate the common covariance $\boldsymbol{\Sigma}$. That is, $\boldsymbol{S}_{pooled} = \dfrac{(n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2}{n_1 + n_2 - 2}$ estimates $\boldsymbol{\Sigma}$.

To test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ vs $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$, the squared statistical distance from $\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2$ to $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ is considered. As $E(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, $(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)$ estimates $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. Since the two samples are independent, $\text{Cov}(\bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2) = 0$. This implies $\text{Cov}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) = \text{Cov}(\bar{\boldsymbol{X}}_1) + \text{Cov}(\bar{\boldsymbol{X}}_2) = \dfrac{1}{n_1}\boldsymbol{\Sigma} + \dfrac{1}{n_2}\boldsymbol{\Sigma} = \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)\boldsymbol{\Sigma}$. Thus, $\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)\boldsymbol{S}_{pooled}$ estimates $\text{Cov}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) = \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)\boldsymbol{\Sigma}$.

The test statistic is, therefore,

$$T^2 = \left[(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right]' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{S}_{pooled}\right]^{-1} \left[(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right].$$

Since $T^2 \sim \dfrac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1}F(p, n_1 + n_2 - p - 1)$, H_0 will be rejected if the observed $T^2 > c^*$ where $c^* = \dfrac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1}F_\alpha(p, n_1 + n_2 - p - 1)$.

A $(1 - \alpha)100\%$ confidence region for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ is given by

$$[(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{S}_{pooled}\right]^{-1} [(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] \leq c^*.$$

which is an ellipsoid centered at $(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$. The boundary of the ellipsoid is $c = \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)c^*$. The half lengths of the axes are $\sqrt{\lambda_j}\sqrt{\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)c^*}$; $j = 1, 2, \cdots, p$ in the direction of $\boldsymbol{e}_j$ which is the normalized eigenvector associated with the eigenvalue λ_j of $\boldsymbol{S}_{pooled}$.

A $(1 - \alpha)100\%$ simultaneous confidence interval for $\boldsymbol{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ is

$$\left[\boldsymbol{a}'(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) \pm \sqrt{c^*}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\boldsymbol{a}'\boldsymbol{S}_{pooled}\boldsymbol{a}}\right].$$

If $\boldsymbol{a}' = (0, 0, \cdots, \underbrace{1}_{j^{th} \text{ position}}, \cdots, 0)$, $\boldsymbol{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \mu_{1j} - \mu_{2j}$, $\boldsymbol{a}'(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) = \bar{x}_{1j} - \bar{x}_{2j}$ and $\boldsymbol{a}'\boldsymbol{S}_{pooled}\boldsymbol{a} = s_{jj}$. Thus, a $(1-\alpha)100\%$ simultaneous confidence interval for $\mu_{1j} - \mu_{2j}$ is

$$\left[(\bar{x}_{1j} - \bar{x}_{2j}) \pm \sqrt{c^*}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) s_{jj}}\right]$$

where s_{jj} is the j^{th} diagonal entry of the pooled covariance matrix, $\boldsymbol{S}_{pooled}$.

Example 5.3. Given the following data on academic performance of students (in preparatory school out of 100 and in university out of 4.00). Test the equality of the population mean vectors between the two groups.

Female	
Preparatory	University
97	3.40
95	3.45
85	3.50

Male	
Preparatory	University
86	3.90
84	3.75
70	2.25
80	3.05
75	2.80

Summary statistics

$$\text{Female: } \bar{\boldsymbol{x}}_1 = \begin{bmatrix} 92.3333 \\ 3.4500 \end{bmatrix} \text{ and } \boldsymbol{S}_1 = \begin{bmatrix} 41.3333 & 0.2500 \\ 0.2500 & 0.0025 \end{bmatrix}$$

$$\text{Male: } \bar{\boldsymbol{x}}_2 = \begin{bmatrix} 79.0000 \\ 3.1500 \end{bmatrix} \text{ and } \boldsymbol{S}_2 = \begin{bmatrix} 43.0000 & 4.4125 \\ 4.4125 & 0.4663 \end{bmatrix}$$

$$\boldsymbol{S}_{pooled} = \begin{bmatrix} 42.4444 & 3.0250 \\ 3.0250 & 0.3117 \end{bmatrix} \Rightarrow \boldsymbol{S}_{pooled}^{-1} = \begin{bmatrix} 0.0764 & -0.7415 \\ -0.7415 & 10.4048 \end{bmatrix}$$

The observed test statistic is:

$$\begin{aligned} T^2 &= \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{S}_{pooled}^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) \\ &= \left(\frac{1}{3} + \frac{1}{5}\right)^{-1} \begin{bmatrix} 13.3333 & 0.3000 \end{bmatrix} \begin{bmatrix} 0.0764 & -0.7415 \\ -0.7415 & 10.4048 \end{bmatrix} \begin{bmatrix} 13.3333 \\ 0.3000 \end{bmatrix} \\ &= 16.0999 \end{aligned}$$

Critical value: $c^* = \frac{(6)2}{5} F_{0.05}(2, 5) = 13.8866$. Reject $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$.

A $(1-\alpha)100\%$ simultaneous confidence interval for $\mu_{1j} - \mu_{2j}$ is

$$\left[(\bar{x}_{1j} - \bar{x}_{2j}) \pm \sqrt{c^*}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) s_{jj}}\right]$$

where s_{jj} is the j^{th} diagonal entry of the pooled covariance matrix, $\boldsymbol{S}_{pooled}$.

For component 1 (preparatory score):

$$\mu_{11}-\mu_{21}:\left[(\bar{x}_{11}-\bar{x}_{21})\pm\sqrt{c^*}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)s_{11}}\right]$$
$$\left(13.3333\pm\sqrt{13.8866}\sqrt{\left(\frac{1}{3}+\frac{1}{5}\right)42.4444}\right)=(-4.3967,31.0633)$$

For component 2 (university score):

$$\mu_{12}-\mu_{22}:\left[(\bar{x}_{12}-\bar{x}_{22})\pm\sqrt{c^*}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)s_{22}}\right]$$
$$\left(0.3000\pm\sqrt{13.8866}\sqrt{\left(\frac{1}{3}+\frac{1}{5}\right)0.3117}\right)=(-1.2194,1.8194)$$

Both simultaneous confidence intervals contain the value zero indicating that there is no significant difference in the mean vectors between females and males. But, this is in contradiction with the test of $H_0:\boldsymbol{\mu}_1=\boldsymbol{\mu}_2$. The possible reasons may be:

- The multivariate normality of the observation vectors might be violated because of the small sample sizes.
- The assumption of equality of the covariance matrices ($\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2$) may not hold.

5.2.2 Comparison of Several Multivariate Population Means

Often, more than two populations need to be compared. Random samples are collected from each of g populations.

Univariate ANOVA:

- Let $X_{\ell 1}$, $X_{\ell 2}$, $\cdots$, $X_{\ell n_\ell}$ be a random sample from an $\mathcal{N}(\mu_\ell,\sigma^2)$; $\ell=1,2,\cdots,g$.
- The samples from different populations are independent.
- All populations have a common variance, σ^2.

The null hypothesis of equality of means $H_0:\mu_1=\mu_2=\cdots=\mu_g$. Each population mean $\mu_\ell;\ \ell=1,2,\cdots,g$ can be considered as a sum of an overall mean (μ) and a component specific to each population (τ_ℓ), that is, $\mu_\ell=\mu+\tau_\ell$ where $\tau_\ell=\mu_\ell-\mu$ is the ℓ^{th} population (treatment) effect. The null hypothesis now becomes $H_0:\tau_1=\tau_2=\cdots=\tau_g=0$.

Since the response $X_{\ell i}\sim\mathcal{N}(\mu_\ell,\sigma^2)$, it can be expressed as $X_{\ell i}=\mu_\ell+e_{\ell i}=\mu+\tau_\ell+e_{\ell i}$ where the random error $e_{\ell i}$ are independent $\mathcal{N}(0,\sigma^2)$. A constraint $\sum_{\ell=1}^{g}n_\ell\tau_\ell=0$ is imposed to define the model parameters uniquely.

Hence, the analysis of variance is based on the decomposition of each observed value $x_{\ell i}$,

$$\begin{aligned} x_{\ell i} &= \bar{x} + (\bar{x}_\ell - \bar{x}) + (x_{\ell i} - \bar{x}_\ell) \\ x_{\ell i} - \bar{x} &= (\bar{x}_\ell - \bar{x}) + (x_{\ell i} - \bar{x}_\ell) \\ (x_{\ell i} - \bar{x})^2 &= (\bar{x}_\ell - \bar{x})^2 + 2(\bar{x}_\ell - \bar{x})(x_{\ell i} - \bar{x}_\ell) + (x_{\ell i} - \bar{x}_\ell)^2 \end{aligned}$$

Taking the summation over i,

$$\sum_{i=1}^{n_\ell}(x_{\ell i} - \bar{x})^2 = n_\ell(\bar{x}_\ell - \bar{x})^2 + 2(\bar{x}_\ell - \bar{x})\sum_{i=1}^{n_\ell}(x_{\ell i} - \bar{x}_\ell) + \sum_{i=1}^{n_\ell}(x_{\ell i} - \bar{x}_\ell)^2$$

Since $\sum_{i=1}^{n_\ell}(x_{\ell i} - \bar{x}_\ell) = \sum_{i=1}^{n_\ell} x_{\ell i} - n_\ell \bar{x}_\ell = 0$,

$$\sum_{i=1}^{n_\ell}(x_{\ell i} - \bar{x})^2 = n_\ell(\bar{x}_\ell - \bar{x})^2 + \sum_{i=1}^{n_\ell}(x_{\ell i} - \bar{x}_\ell)^2.$$

Now taking the summation over ℓ,

$$\underbrace{\sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}(x_{\ell i} - \bar{x})^2}_{SS_{\text{corrected}}} = \underbrace{\sum_{\ell=1}^{g} n_\ell(\bar{x}_\ell - \bar{x})^2}_{BSS\ (SS_{\text{treatment}})} + \underbrace{\sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}(x_{\ell i} - \bar{x}_\ell)^2}_{WSS\ (SS_{\text{residuals}})}.$$

The ANOVA table is

Sources of variation	Sum of squares (SS)	Degrees of freedom (df)
Between Group (Treatment)	$BSS = \sum_{\ell=1}^{g} n_\ell(\bar{x}_\ell - \bar{x})^2$	$g-1$
Within Group (Residual)	$WSS = \sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}(x_{\ell i} - \bar{x}_\ell)^2$	$n-g$
Total	$TSS_{\text{cor}} = \sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}(x_{\ell i} - \bar{x})^2$	$n-1$

The null hypothesis H_0 is rejected if $F = \dfrac{BSS/(g-1)}{WSS/(n-g)} > F_\alpha(g-1, n-g)$. Rejecting H_0 when F is too large is equivalent with rejecting H_0 if $\dfrac{BSS}{WSS}$ is too large or $\dfrac{BSS}{WSS} + 1$ is too large or $\dfrac{1}{\frac{BSS}{WSS}+1}$ too small or $\dfrac{WSS}{BSS+WSS}$ is too small. This is used for a multivariate generalization.

Multivariate ANOVA - MANOVA:

Population 1: $\boldsymbol{X}_{11}, \boldsymbol{X}_{12}, \cdots, \boldsymbol{X}_{1n_1}$

Population 2: $\boldsymbol{X}_{21}, \boldsymbol{X}_{22}, \cdots, \boldsymbol{X}_{2n_2}$

$\vdots$

Population g: $\boldsymbol{X}_{g1}, \boldsymbol{X}_{g2}, \cdots, \boldsymbol{X}_{gn_g}$

Let $\boldsymbol{X}_{\ell 1}$, $\boldsymbol{X}_{\ell 2}$, $\cdots$, $\boldsymbol{X}_{\ell n_\ell}$; $\ell = 1, 2, \cdots, g$ is a random sample of size n_ℓ from an $\mathcal{N}_p(\boldsymbol{\mu}_\ell, \boldsymbol{\Sigma})$. The random sample from the different populations are independent.

The sample mean of the ℓ^{th} group is $\bar{\boldsymbol{x}}_\ell = \dfrac{1}{n_\ell}\sum_{i=1}^{n_\ell} \boldsymbol{x}_{\ell i}$; $\ell = 1, 2, \cdots, g$ and the overall sample mean is $\bar{\boldsymbol{x}} = \dfrac{1}{n}\sum_{\ell=1}^{g} n_\ell \bar{\boldsymbol{x}}_\ell$. Also, the sample covariance matrix of the ℓ^{th} group is $\boldsymbol{S}_\ell = \dfrac{1}{n_\ell - 1}\sum_{i=1}^{n_\ell}(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)'$; $\ell = 1, 2, \cdots, g$. This implies the pooled covariance matrix is

$$\boldsymbol{S}_{pooled} = \frac{(n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2 + \cdots + (n_g - 1)\boldsymbol{S}_g}{n - g}.$$

The null hypothesis of equality of means $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g$. The model is $\boldsymbol{X}_{\ell i} = \boldsymbol{\mu} + \boldsymbol{\tau}_\ell + \boldsymbol{e}_{\ell i}$ where $\boldsymbol{\tau}_\ell = \boldsymbol{\mu}_\ell - \boldsymbol{\mu} = (\tau_{\ell 1}, \tau_{\ell 2}, \cdots, \tau_{\ell p})'$ is the ℓ^{th} group (treatment) effect with $\sum_{\ell=1}^{g} n_\ell \boldsymbol{\tau}_\ell = \boldsymbol{0}$ and $\boldsymbol{e}_{\ell i} \sim \mathcal{N}_p(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$.

Now to decompose the sum squares, matrix manipulation is used as follows.

$$\begin{aligned}
(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}})(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}})' &= [(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}}) + (\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)][(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}}) + (\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)]' \\
&= [(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}}) + (\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)][(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})' + (\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)'] \\
&= (\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})' + (\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)' \\
&\quad + (\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})' + (\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)'
\end{aligned}$$

When taking the summation over i, the middle two cross-products become zero vectors. Then, taking the summation over ℓ gives

$$\sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}})(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}})' = \underbrace{\sum_{\ell=1}^{g} n_\ell(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})'}_{\boldsymbol{B}} + \underbrace{\sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)'}_{\boldsymbol{W}}$$

Therefore, the MANOVA table is

Sources of variation	Matrix of SS and cross-products (SSP)	*df*
Between Group (Treatment)	$\boldsymbol{B} = \sum_{\ell=1}^{g} n_\ell(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})'$	$g - 1$
Within Group (Residual)	$\boldsymbol{W} = \sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)'$	$n - g$
Total	$\boldsymbol{B} + \boldsymbol{W} = \sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}})(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}})'$	$n - 1$

Note that $\boldsymbol{W} = \sum_{\ell=1}^{g}(n_\ell - 1)\boldsymbol{S}_\ell = (n - g)\boldsymbol{S}_{pooled} \Rightarrow \boldsymbol{S}_{pooled} = \dfrac{1}{n - g}\boldsymbol{W}$.

The null hypothesis is $H_0 : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_g = \mathbf{0}$ is to be rejected if

$$\Lambda^* = \frac{|\boldsymbol{W}|}{|\boldsymbol{B} + \boldsymbol{W}|}$$

is too small. The statistic Λ^* is known as Wilks' Lambda. The exact distribution of Λ^* for special cases is given on the text book on page 303, Table 6.3.

If n is large, Bartlett has shown that

$$-\left(n - 1 - \frac{p+g}{2}\right)\log(\Lambda^*) \sim \chi^2[p(g-1)].$$

Simultaneous Confidence Intervals

If $H_0 : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_g = \mathbf{0}$ is rejected, the next task is to identify which groups for which variable(s) (component(s)) are responsible for rejection. Bonferroni's approach can be used to construct simultaneous intervals for the components of the differences $(\boldsymbol{\tau}_\ell - \boldsymbol{\tau}_k)$ or $(\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_k)$, $\ell \neq k$, which adjusts the significance level to the $p(gC_2)$ confidence intervals required.

The $(1-\alpha)100\%$ confidence interval for the linear combination $\boldsymbol{a}'(\boldsymbol{\tau}_\ell - \boldsymbol{\tau}_k) = \boldsymbol{a}'(\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_k)$ is

$$\left[\boldsymbol{a}'(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}}_k) \pm t_{\alpha/[pg(g-1)]}(n-g)\sqrt{\left(\frac{1}{n_\ell} + \frac{1}{n_k}\right)\boldsymbol{a}'\boldsymbol{S}_{pooled}\boldsymbol{a}}\right].$$

Here, $\boldsymbol{\tau}_\ell = \boldsymbol{\mu}_\ell - \boldsymbol{\mu} = (\tau_{\ell 1}, \tau_{\ell 2}, \cdots, \tau_{\ell p})'$; $\ell = 1, 2, \cdots, g$. Let $\boldsymbol{a} = (0, 0, \cdots, \underbrace{1}_{j^{th}\text{ position}}, \cdots, 0)$. Thus, the linear combination $\boldsymbol{a}'(\boldsymbol{\tau}_\ell - \boldsymbol{\tau}_k) = \boldsymbol{a}'(\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_k) = \mu_{\ell j} - \mu_{kj}$ corresponds to component j. That is, for component j, $(\bar{X}_{\ell j} - \bar{X}_{kj})$ estimates $(\mu_{\ell j} - \mu_{kj})$. Hence, $\text{cov}(\bar{X}_{\ell j} - \bar{X}_{kj})$ is estimated by $\left(\frac{1}{n_\ell} + \frac{1}{n_k}\right)s_{jj}$ where s_{jj} is the j^{th} diagonal element of $\boldsymbol{S}_{pooled}$. Since $\boldsymbol{S}_{pooled} = \frac{1}{n-g}\boldsymbol{W}$, $s_{jj} = \frac{w_{jj}}{n-g}$ where w_{jj} is the j^{th} diagonal element of $\boldsymbol{W}$.

Therefore, a $(1-\alpha)100\%$ confidence interval for the difference $\tau_{\ell j} - \tau_{kj}$ is

$$\left[(\bar{x}_{\ell j} - \bar{x}_{kj}) \pm t_{\alpha/[pg(g-1)]}(n-g)\sqrt{\left(\frac{1}{n_\ell} + \frac{1}{n_k}\right)\frac{w_{jj}}{n-g}}\right]$$

where w_{jj} is the j^{th} diagonal element of $\boldsymbol{W}$.

Example 5.4. Given the following observation vectors on two responses collected for three treatments.

Treatment 1	6	5	8	4	7
	7	9	6	9	9
Treatment 2	3	1	2		
	3	6	3		
Treatment 3	2	5	3	2	
	3	1	1	3	

Construct oneway MANOVA and test for treatment effects at 5% significance level.

$$\bar{\boldsymbol{x}}_1 = \begin{bmatrix} 6 \\ 8 \end{bmatrix}, \bar{\boldsymbol{x}}_2 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \bar{\boldsymbol{x}}_3 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$\Rightarrow \bar{\boldsymbol{x}} = \frac{1}{n}\sum_{\ell=1}^{g} n_\ell \bar{\boldsymbol{x}}_\ell = \frac{1}{12}\left(5\begin{bmatrix} 6 \\ 8 \end{bmatrix} + 3\begin{bmatrix} 2 \\ 4 \end{bmatrix} + 4\begin{bmatrix} 3 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

$$\boldsymbol{B} = \sum_{\ell=1}^{g} n_\ell(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_\ell - \bar{\boldsymbol{x}})' = \begin{bmatrix} 36 & 48 \\ 48 & 84 \end{bmatrix}$$

$$\boldsymbol{W} = \sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)(\boldsymbol{x}_{\ell i} - \bar{\boldsymbol{x}}_\ell)' = \begin{bmatrix} 18 & -13 \\ -13 & 38 \end{bmatrix}$$

$$\boldsymbol{B} + \boldsymbol{W} = \begin{bmatrix} 54 & 35 \\ 35 & 122 \end{bmatrix}$$

The MANOVA table is

Sources of variation	Matrix of SS and cross-products (SSP)	*df*
Between Group (Treatment)	$\boldsymbol{B} = \begin{bmatrix} 36 & 48 \\ 48 & 84 \end{bmatrix}$	$3-1=2$
Within Group (Residual)	$\boldsymbol{W} = \begin{bmatrix} 18 & -13 \\ -13 & 38 \end{bmatrix}$	$12-3=9$
Total	$\boldsymbol{B}+\boldsymbol{W} = \begin{bmatrix} 54 & 35 \\ 35 & 122 \end{bmatrix}$	$12-1=11$

$$\Lambda^* = \frac{\boldsymbol{W}}{\boldsymbol{B}+\boldsymbol{W}} = \frac{515}{5363} = 0.096$$

For $p = 2$ and $g = 3$, the exact distribution of Λ^*

$$\frac{n-g-1}{g-1}\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F[2(g-1), 2(n-g-1)]$$

$$\Rightarrow \frac{7}{2}\left(\frac{1-\sqrt{0.096}}{\sqrt{0.096}}\right) = 8.908 \text{ and } F_{0.05}[4,16] = 3.01$$

Therefore, $H_0 : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \boldsymbol{\tau}_3 = \boldsymbol{0}$ should be rejected.

Next for the simultaneous confidence interval $\alpha/[pg(g-1)] = 0.004167$, $t_{0.004167}(n-g) = 3.808$. A $(1-\alpha)100\%$ confidence interval for the difference $\tau_{\ell j} - \tau_{kj}$ is

$$\left[(\bar{x}_{\ell j} - \bar{x}_{kj}) \pm t_{[\alpha/pg(g-1)]}(n-g)\sqrt{\left(\frac{1}{n_\ell} + \frac{1}{n_k}\right)\frac{w_{jj}}{n-g}}\right]$$

where w_{jj} is the j^{th} diagonal element of $\boldsymbol{W}$.

For component 1:

$$\tau_{11}-\tau_{21}: \left[(\bar{x}_{11}-\bar{x}_{21}) \pm 3.808\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\frac{w_{11}}{12-3}}\right]$$
$$\left[4 \pm 3.808\sqrt{\left(\frac{1}{5}+\frac{1}{3}\right)\frac{18}{12-3}}\right] = (0.067, 7.933)$$

$$\tau_{11}-\tau_{31}: \left[(\bar{x}_{11}-\bar{x}_{31}) \pm 3.808\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_3}\right)\frac{w_{11}}{12-3}}\right]$$
$$\left[3 \pm 3.808\sqrt{\left(\frac{1}{5}+\frac{1}{4}\right)\frac{18}{12-3}}\right] = (-0.613, 6.613)$$

$$\tau_{21}-\tau_{31}: \left[(\bar{x}_{21}-\bar{x}_{31}) \pm 3.808\sqrt{\left(\frac{1}{n_2}+\frac{1}{n_3}\right)\frac{w_{11}}{12-3}}\right]$$
$$\left[-1 \pm 3.808\sqrt{\left(\frac{1}{3}+\frac{1}{4}\right)\frac{18}{12-3}}\right] = (-5.113, 3.113)$$

For component 2:

$$\tau_{12}-\tau_{22}: \left[(\bar{x}_{12}-\bar{x}_{22}) \pm 3.808\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\frac{w_{22}}{12-3}}\right]$$
$$\left[4 \pm 3.808\sqrt{\left(\frac{1}{5}+\frac{1}{3}\right)\frac{38}{12-3}}\right] = (-1.714, 9.714)$$

$$\tau_{12}-\tau_{32}: \left[(\bar{x}_{12}-\bar{x}_{32}) \pm 3.808\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_3}\right)\frac{w_{22}}{12-3}}\right]$$
$$\left[6 \pm 3.808\sqrt{\left(\frac{1}{5}+\frac{1}{4}\right)\frac{38}{12-3}}\right] = (0.751, 11.249)$$

$$\tau_{22}-\tau_{32}: \left[(\bar{x}_{22}-\bar{x}_{32}) \pm 3.808\sqrt{\left(\frac{1}{n_2}+\frac{1}{n_3}\right)\frac{w_{22}}{12-3}}\right]$$
$$\left[2 \pm 3.808\sqrt{\left(\frac{1}{3}+\frac{1}{4}\right)\frac{38}{12-3}}\right] = (-3.976, 7.976)$$

Treatment 1 has significantly larger mean than treatment 2 for component 1. It has also significantly larger mean than treatment 3 for component 2.