Multivariate Analysis
Chapter 1 - Introduction and Review

Jingyuan Liu
Department of Statistics, School of Economics
Wang Yanan Institute for Studies in Economics
Xiamen University

Outline

Outline

About the Syllabus

Instructor:	**Jingyuan Liu**
Office:	D307 Economics Building
Office Hours:	By appointment
Email:	jingyuan@xmu.edu.cn
TA:	**Ao Sun**
Email:	654225305@qq.com
QQ Group:	1046221272

Textbook:
Methods of Multivariate Analysis, by Alvin C.Rencher and William F. Christensen.

Reference Book:
Applied Multivariate Statistical Analysis, by Richard A. Johnson and Dean W. Wichern.
An Introduction to Applied Multivariate Analysis with R, by Brian Everitt and Torsten Hothorn.

Programing:
Any statistical software is acceptable, and some guide and support in R will be provided.

Grade Policy:

- Homework – 20%
- Quiz – 5%
- Midterm Exam – 30%
- Final Project – 10%
- Final Exam – 35%

Outline

Topics to Be Covered in This Course

- Characterizing and displaying multivariate data
- Multivariate normal distributions
- Test on one or two mean vectors
- Discrimination and classification
- Multivariate multiple regression
- Principle component analysis
- Canonical correlation analysis
- Factor analysis
- Clustering methods

Outline

2 Introduction to Multivariate Analysis
- Multivariate Data
- Multivariate Data Analysis

A multivarite data example

Fisher's iris flower data set:
This famous dataset was collected to quantify the morphologic variation of iris flowers of three species - setosa, versicolor, virginica.

Part of the data is tabulated as follows:

Sepal length ⬍	Sepal width ▲	Petal length ⬍	Petal width ⬍	Species ⬍
5.0	2.0	3.5	1.0	*I. versicolor*
6.2	2.2	4.5	1.5	*I. versicolor*
6.0	2.2	5.0	1.5	*I. virginica*
6.0	2.2	4.0	1.0	*I. versicolor*
6.3	2.3	4.4	1.3	*I. versicolor*
5.5	2.3	4.0	1.3	*I. versicolor*
5.0	2.3	3.3	1.0	*I. versicolor*
4.5	2.3	1.3	0.3	*I. setosa*
5.5	2.4	3.8	1.1	*I. versicolor*
5.5	2.4	3.7	1.0	*I. versicolor*
4.9	2.4	3.3	1.0	*I. versicolor*
6.7	2.5	5.8	1.8	*I. virginica*
6.3	2.5	5.0	1.9	*I. virginica*

Outline

2 Introduction to Multivariate Analysis
- Multivariate Data
- Multivariate Data Analysis

What is Multivariate Analysis?

Multivariate analysis deals with statistical methods designed to elicit information from multivariate data sets which include *simultaneous* measurements on *many* variables.

What does Multivariate Analysis do?

- **Data reduction or simplification:** To represent the phenomenon (data) as simply as possible without sacrificing any valuable information.
- **Sorting and grouping:** To group "similar" objects or variables based on measured characteristics; or to set up rules for classifying objects into well-defined groups.
- **Investigation of the dependence among variables:** To examine the nature of relationship between variables.
- **Prediction:** To predict the values of one or more variables based on observations of the other variables utilizing the relationship among these variables.
- **Hypothesis testing:** To test hypotheses formulated by the parameters of multivariate populations.

Where did Multivariate Analyses Come from?

John Tukey
Bell Labs & Princeton Univ.

"The best thing about being a statistician is that you get to play in everyone else's backyard."

Much of the early developmental work in multivariate analysis was motivated by problems from **social and behavioral sciences**, especially education and psychology.

- Factor analysis - explaining psychological theories of human ability and behavior
- Principal component analysis - analyzing student scores on a battery of different tests
- Canonical correlation - exploring relationship between student scores on two separate batteries of tests

Some multivariate methods were motivated by problems in other scientific areas.

- Linear discriminant analysis - taxonomic problem using multiple botanical measurements
- Multivariate analysis of variance - agricultural experiments
- Regression - heredity and the orbits of planets

Application Areas of Multivariate Analysis

- **Marketing:** Predict new purchasing trends. Identify "loyal" customers. Detect potential customers. Segment markets. Precise marketing.
- **Banking:** Evaluate loan policies using customer characteristics. Predict credit card switch.
- **Finance:** Identify relationships between financial indicators. Track changes in an investment portfolio and predict price turning points. Analyze volatility patterns in high-frequency stock transactions.
- **Insurance:** Identify characteristics of buyers of new policies. Find unusual claim patterns. Identify "risky" customers.

- **Healthcare:** Early warning of diseases. Predict doctor visits from patient characteristics. Precise medical care.
- **Molecular Biology:** Gene detection. Analyze DNA microarrays. Characterize biological function. Predict protein structure.
- **Astronomy:** Catalogue (as stars, galaxies, etc.) objects in the sky. Identify patterns and relationships of objects.
- **Forensic Accounting:** Detect fraud in insurance, credit card and medical claims. Identify instances of tax evasion. Identify insider-trading behaviors in stock market.
- **Sports:** Identify most effective stretagies. Discover hidden game patterns.

A Typical Knowledge Discovery Process

1. Crystalize scientific/industrial objective and extract the corresponding statistical problem.
2. Select the target/historical dataset; fix output and input variables.
3. Data cleaning (removal of noise, identification of potential outliers, imputing missing data)
4. Preprocess the data (data transformations, tracking time-dependent information)
5. Decide analysis methods (regression, classification, etc.)
6. Analyze the cleaned data (algorithms for data reduction, fitting models, prediction, extracting patterns)
7. Interprete and assess the knowledge from analysis results.

Outline

3 Review of Matrix Algebra
- Basic Definitions and Operations
- Characteristics of Matrices
- R Implementation

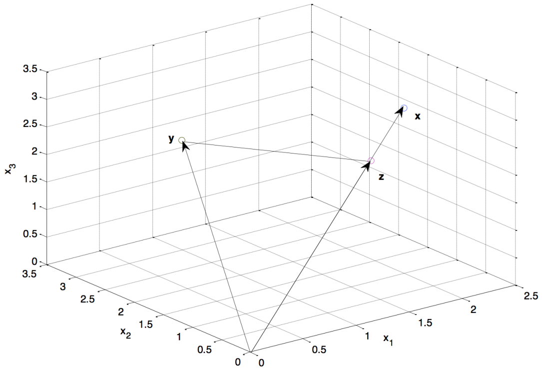
Review of Basic Definitions

- Matrix ($\mathbf{A}$), vector ($\mathbf{a}$), scalar (a)
- Equality of matrices and vectors
- Transpose ($\mathbf{A}'$) and symmetric matrix ($\mathbf{A}' = \mathbf{A}$)
- Special matrix: diagonal matrix (diag($\mathbf{A}$)), identity matrix ($\mathbf{I}$), square matrix, upper/lower triangular matrix
- Positive definite matrix ($\mathbf{A} > 0$) and positive semidefinite matrix ($\mathbf{A} \geq 0$)

Review of Operations

- **Addition**: $\mathbf{A} \pm \mathbf{B}$ or $\mathbf{a} \pm \mathbf{b}$
- **Multiplication**: $\mathbf{AB}$, $c\mathbf{A}$, $\mathbf{Aa}$, $\mathbf{a}'\mathbf{b}$, $\mathbf{ab}'$
- **Length** of vector $\mathbf{a}$: $L_{\mathbf{a}} = \sqrt{\mathbf{a}'\mathbf{a}} = \sqrt{\sum_{i=1}^{n} a_i^2}$;
  vector $\mathbf{a}$ is said to be **normalized** if $L_{\mathbf{a}} = 1$.
- **Angle** between vectors $\mathbf{a}$ and $\mathbf{b}$: $\cos(\theta) = \mathbf{a}'\mathbf{b}/(L_{\mathbf{a}}L_{\mathbf{b}})$;
  $\mathbf{a}$ and $\mathbf{b}$ are **perpendicular/orthogonal** if $\mathbf{a}'\mathbf{b} = \mathbf{0}$
- **Projection** of vector $\mathbf{a}$ onto $\mathbf{b}$: $(\mathbf{a}'\mathbf{b})\mathbf{b}/(\mathbf{b}'\mathbf{b})$

Example: Let $\mathbf{x} = (2, 1, 3)'$, $\mathbf{y} = (1, 3, 2)'$. Then we have: Length $L_\mathbf{x} = L_\mathbf{y} = 3.74$; Angle between $\mathbf{x}$ and $\mathbf{y}$: $\theta = 0.79$; Projection of $\mathbf{y}$ onto $\mathbf{x}$: $\mathbf{z} = (1.57, 0.79, 2.36)'$



Outline

Rank and Inverse

- **Linear dependence** of vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$: if there exist constants $c_1, \dots, c_n$, not all zero, s.t. $\sum_{i=1}^n c_i \mathbf{a}_i = 0$. Otherwise the vectors are **linearly independent**.
- **Rank** of matrix $\mathbf{A}$: number of linearly independent rows (or equivalently, columns) of $\mathbf{A}$.
- **Full rank**: if the rank of $\mathbf{A}_{n\times p}$ is the smaller of n and p.
- **Nonsingular** matrix $\mathbf{A}$: if $\mathbf{A}$ is square and of full rank. If $\mathbf{A}$ is square of of less than full rank, then $\mathbf{A}$ is **singular**.
- **Inverse** of nonsingular square matrix $\mathbf{A}$: a matrix $\mathbf{A}^{-1}$ s.t. $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$
- **Orthogonal** matrix $\mathbf{C}$: if $\mathbf{C}\mathbf{C}' = \mathbf{C}'\mathbf{C} = \mathbf{I}$ or $\mathbf{C}^{-1} = \mathbf{C}'$.

Properties of **rank** and **inverse**:

- $\text{rank}(\mathbf{AB}) \leq \min(\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B}))$.
- $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$.
- $\text{rank}(\mathbf{AA}') = \text{rank}(\mathbf{A}'\mathbf{A}) = \text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}')$.
- $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$.
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
- If $\mathbf{Ax} = \mathbf{Bx}$ for all possible values of $\mathbf{x}$, then $\mathbf{A} = \mathbf{B}$.
- If $\mathbf{B}$ is nonsingular, then $\mathbf{AB} = \mathbf{CB}$ implies $\mathbf{A} = \mathbf{C}$.

Determinant

Properties of **determinant** of $n \times n$ square matrix $\mathbf{A}$ ($|\mathbf{A}|$):

- $|c\mathbf{A}| = c^n|\mathbf{A}|$ where c is a constant.
- $|\mathbf{A}| = 0$ if $\mathbf{A}$ is singular.
- $|\mathbf{A}| \neq 0$ if $\mathbf{A}$ is nonsingular.
- $|\mathbf{A}| > 0$ if $\mathbf{A}$ is positive definite.
- If $\mathbf{A}$ and $\mathbf{B}$ are square matrices, then $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$.
- If $\mathbf{A}$ is partitioned by $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$, then

$$|\mathbf{A}| = \begin{vmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{vmatrix} = |\mathbf{A}_{11}||\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}| = |\mathbf{A}_{22}||\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}|$$

Trace

Trace of $n \times n$ matrix $\mathbf{A} = (a_{ij})_{n\times n}$: $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$

- $\mathrm{tr}(\mathbf{A} \pm \mathbf{B}) = \mathrm{tr}(\mathbf{A}) \pm \mathrm{tr}(\mathbf{B})$
- $\mathrm{tr}(c\mathbf{A}) = c \cdot \mathrm{tr}(\mathbf{A})$
- $\mathrm{tr}(\mathbf{B}^{-1}\mathbf{A}\mathbf{B}) = \mathrm{tr}(\mathbf{A})$
- $\mathrm{tr}(\mathbf{A}\mathbf{B}) = \mathrm{tr}(\mathbf{B}\mathbf{A})$ ($\mathbf{A}$ and $\mathbf{B}$ may not be square.)
- $\mathrm{tr}(\mathbf{A}'\mathbf{A}) = \mathrm{tr}(\mathbf{A}\mathbf{A}') = \sum_{i=1}^{n} \sum_{j=1}^{p} a_{ij}^2$, where $\mathbf{A} = (a_{ij})_{n\times p}$

Eigenvalues and Eigenvectors

Eigenvalue λ_i and Eigenvector $\mathbf{x}_i$ of square matrix $\mathbf{A}_{n\times n}$: scalar λ_i and nonzero vector $\mathbf{x}_i$ (normalized so that $\mathbf{x}_i'\mathbf{x}_i = 1$) s.t. $\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i$, or $(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{x}_i = 0$, $i = 1, \ldots, n$.

- **Characteristic equation**: $|\mathbf{A} - \lambda_i \mathbf{I}| = 0$
- $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i$, $|\mathbf{A}| = \prod_{i=1}^n \lambda_i$.
- If $\mathbf{A} > 0$, then all $\lambda_i > 0$.
- If $\mathbf{A} \geq 0$, then all $\lambda_i \geq 0$, and the number of positive λ_i's equal to the rank of $\mathbf{A}$.
- If $\mathbf{A}$ is symmetric, then $\mathbf{x}_i$'s are mutually orthogonal.
- $1 \pm \lambda_i$ is the eigenvalue of $\mathbf{I} \pm \mathbf{A}$; λ_i^2 is the eigenvalue of $\mathbf{A}^2$; $1/\lambda_i$ is the eigenvalue of $\mathbf{A}^{-1}$ if $\mathbf{A}$ is nonsingular. The corresponding eigenvectors are still $\mathbf{x}_i$.

Spectral Decomposition

For a symmetric $n \times n$ matrix $\mathbf{A}$ with the eigenvalue and normalized eigenvector pairs $(\lambda_i, \mathbf{x}_i)$, $i = 1, \ldots, n$, the **spectral decomposition** of $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{CDC}' = \sum_{\mathbf{i=1}}^{\mathbf{n}} \lambda_{\mathbf{i}} \mathbf{x_i} \mathbf{x_i}',$$

where $\mathbf{C} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is orthogonal, and

$$\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Spectral decomposition can be used to compute the following:

- Square root of symmetric and positive definite $\mathbf{A}$:
  $\mathbf{A}^{1/2} = \mathbf{C}\mathbf{D}^{1/2}\mathbf{C}'$, where $\mathbf{D}^{1/2} = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$.
- Square of square matrix $\mathbf{A}$:
  $\mathbf{A}^2 = \mathbf{C}\mathbf{D}^2\mathbf{C}'$, where $\mathbf{D}^2 = \text{diag}(\lambda_1^2, \ldots, \lambda_n^2)$.
- Inverse of nonsingular matrix $\mathbf{A}$:
  $\mathbf{A}^{-1} = \mathbf{C}\mathbf{D}^{-1}\mathbf{C}'$, where $\mathbf{D}^{-1} = \text{diag}(1/\lambda_1, \ldots, 1/\lambda_n)$.

Cholesky Decomposition

A positive definite matrix $\mathbf{A} = (a_{ij})_{n\times n}$ can be factored into

$$\mathbf{A} = \mathbf{T}'\mathbf{T},$$

where $\mathbf{T}$ is a nonsingular upper triangular matrix. One way to obtain $\mathbf{T} = (t_{ij})_{n\times n}$ is the **Cholesky decomposition**:

$$\begin{aligned}
t_{11} &= \sqrt{a_{11}}, \quad t_{1j} = \frac{a_{1j}}{t_{11}} && 2 \le j \le n, \\
t_{ii} &= \sqrt{a_{ii} - \sum_{k=1}^{i-1} t_{ki}^2} && 2 \le i \le n, \\
t_{ij} &= \frac{a_{ij} - \sum_{k=1}^{i-1} t_{ki} t_{kj}}{t_{ii}} && 2 \le i < j \le n, \\
t_{ij} &= 0 && 1 \le j < i \le n.
\end{aligned}$$

Cholesky Decomposition: Example

Consider the 3×3 matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & -3 \\ 0 & 6 & 3 \\ -3 & 3 & 6 \end{pmatrix}$$

Then by the aforementioned formulas, or by $>$*chol*(A) in R,

Cholesky Decomposition: Example

Consider the 3×3 matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & -3 \\ 0 & 6 & 3 \\ -3 & 3 & 6 \end{pmatrix}$$

Then by the aforementioned formulas, or by $> chol(A)$ in R,

$$\mathbf{T} = \begin{pmatrix} \sqrt{3} & 0 & -\sqrt{3} \\ 0 & \sqrt{6} & \sqrt{1.5} \\ 0 & 0 & \sqrt{1.5} \end{pmatrix}$$

Outline

Matrix Computation in R

- Addition and substraction:
```
A=matrix(1:12, nrow=4, ncol=3)
B=matrix(seq(3, 12, length=12), 4, 3)
A+B # matrix addition
A-B # matrix substraction
```
- Transpose: Use the command "`t()`". For example,
```
A=matrix(1:12, nrow=4, ncol=3)
t(A) # calculate the transpose of A
```
- Multiplication: Use the command "`%*%`". For example,
```
A=matrix(1:12, nrow=4, ncol=3)
B=matrix(seq(2, 8, length=9), 3, 3)
A %*% B
```
`A*B` is not the usual matrix multiplication, it will multiply two matrices of the same shape elementwisely. Try
```
A=matrix(1:12, nrow=4, ncol=3)
B=matrix(seq(2, 8, length=9), 3, 3)
C=matrix(seq(2, 8, length=12), 4, 3)
A * B # will give error message
A * C # will give a 4x3 matrix whose entries are the product of the
      # corresponding entries of A and C
```

- Inverse: Use the command "`sovle()`". For example,

```
A=matrix(rnorm(9), 3) # create a 3x3 matrix whose entries are iid N(0, 1)
solve(A) # calculate the inverse of the matrix A
```

- Determinant: Use the command "`det()`" to calculate the determinant of a square matrix. For example,

```
det(matrix(seq(1,3, length=9), 3))
```

- Eigen-pairs and Spectral Decomposition:
  You can use the command "`eigen()`" to find the eigen-pairs of a symmetric matrix, and hence the spectral decomposition of the given matrix. For example,

```
> A=matrix(rnorm(9), 3)
> A=(A+t(A))/2 # to make sure we have a symmetric matrix
> eigen(A) # do the spectural decomposition of A
$values       # eigenvalues of A
[1] 1.9157871 1.0658594 0.2605430

$vectors      # P matrix in the Spectural Decomposition
           [,1]       [,2]        [,3]
[1,]  0.5790448 -0.3205481  0.74963728
[2,] -0.4593717 -0.8878969 -0.02483444
[3,] -0.6735612  0.3299819  0.66138279
```

Summary and Take-home Messages

About multivariate analyses:

- Why do we need them?
- What topics do they contain?
- What is the typical data structure?

About matrices and vectors:

- Review of their basic definitions, operations and characteristics
- Implementation of the matrix computation in R