

Multivariate Analysis

Chapter 1: Introduction

Shaobo Jin

Department of Mathematics

What is Multivariate Analysis?

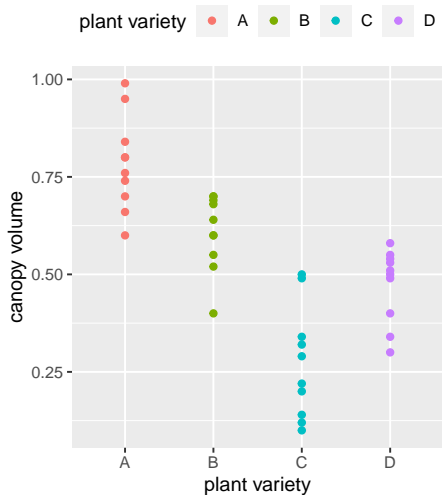
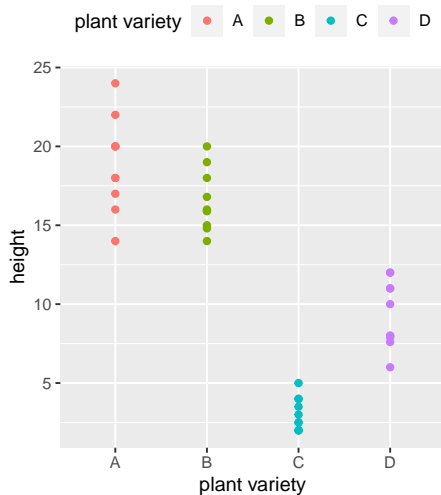
- Multivariate analysis deals with simultaneous measurements on many variables.
- Analysis of each variable separately can miss any interesting patterns in the data.
- Multivariate analysis includes many topics:
 - test mean and covariance structure (e.g., MANOVA),
 - regression analysis,
 - principal component analysis,
 - factor analysis,
 - canonical correlation analysis,
 - discriminant analysis and classification,
 - cluster analysis.

Intended Learning Outcome

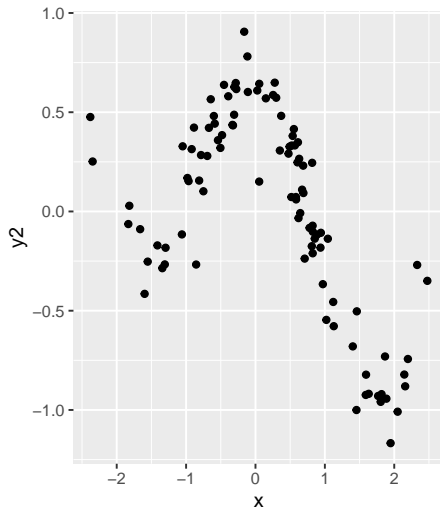
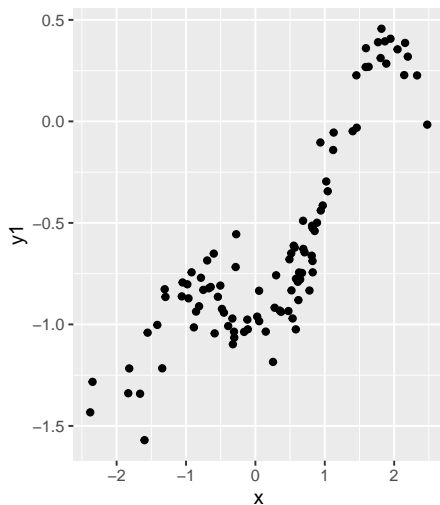
On completion of the course, you should be able to

- ① give an account of some methods of visualizing multivariate data sets
- ② give an account of and use multivariate normal distribution
- ③ perform statistical tests of the mean value vector of a multivariate normal distribution
- ④ perform statistical tests of two or several populations of a multivariate normal distribution
- ⑤ give an account of methods and techniques for validation of multivariate normal distribution
- ⑥ use principal component and factor analysis for typical problems
- ⑦ use canonical correlation analysis
- ⑧ use classification techniques
- ⑨ give an account of and use methods for multivariate cluster analysis
- ⑩ present mathematical statistical arguments to others.

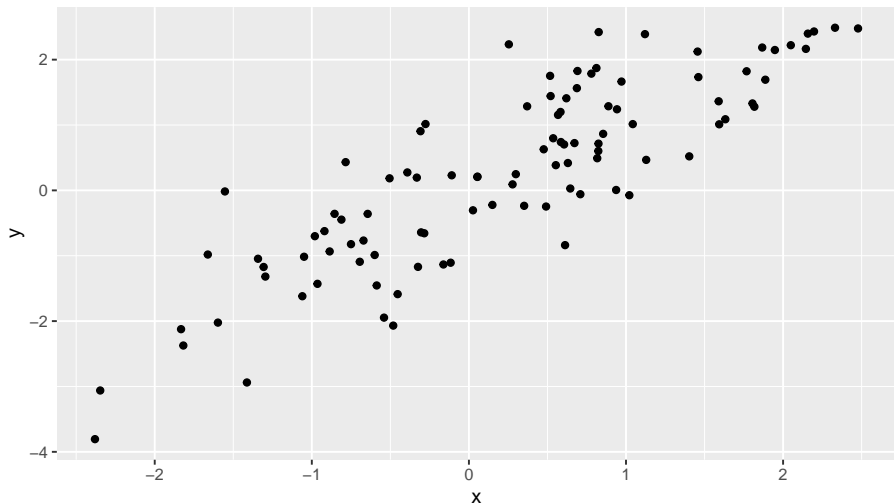
Applications of Multivariate Analysis: MANOVA



Applications of Multivariate Analysis: Regression



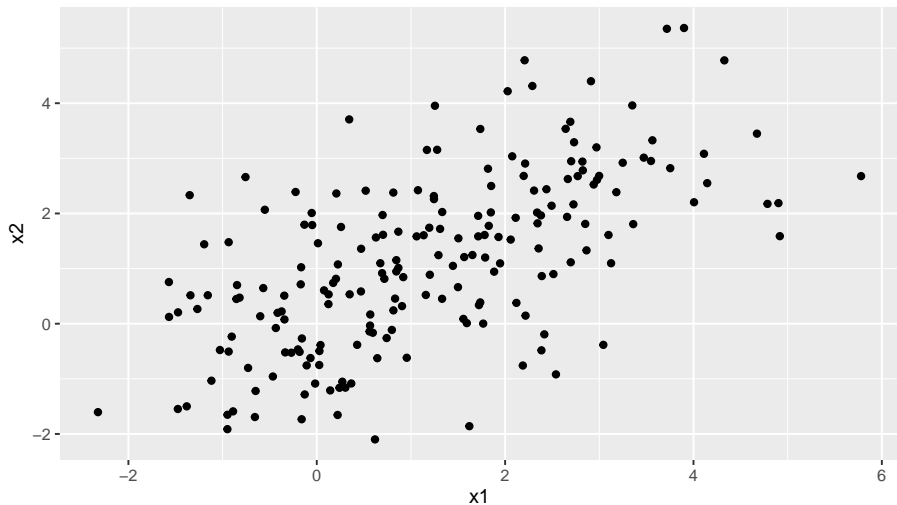
Applications of Multivariate Analysis: PCA



Applications of Multivariate Analysis: Factor Analysis

	--	-	-/+	+	++
1. I am a 'worrier'	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. I make friends easily	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. I have a vivid imagination	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. I trust others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. I complete tasks successfully	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. I get angry easily	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. I really enjoy large parties and gatherings	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. I think art is important	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Applications of Multivariate Analysis: Clustering



Applications of Multivariate Analysis: Classification



An Example: Prices of Game Consoles

Website	Nintendo Switch	Xbox	PS5
Webhallen	4290	6449	7190
MediaMarkt	4290	6595	7590
Elgiganten	4290	7295	7190
Amazon	4099	6752	6929
Inet	4290	6790	7190
CDON	3597	6746	7195

Data

	variable 1	...	variable k	...	variable p
item 1	x_{11}	...	x_{1k}	...	x_{1p}
item 2	x_{21}	...	x_{2k}	...	x_{2p}
\vdots	\vdots		\vdots		\vdots
item j	x_{j1}	...	x_{jk}	...	x_{jp}
\vdots	\vdots		\vdots		\vdots
item n	x_{n1}	...	x_{nk}	...	x_{np}

x_{jk} = the measurment of the k th variable on the j th item

n = number of observations/items/individuals/units

p = number of variables

Data Array/Frame/Matrix

We often collect our data as a matrix:

$$\mathbf{X}_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix}.$$

x_{jk} = the measurement of the k th variable on the j th item

$\mathbf{X}_{n \times p}$ = data array with n rows and p columns

n = number of observations/items/individuals/units

p = number of variables

\mathbf{x}_j = a $p \times 1$ vector, representing the j th observation

Notation

A random vector is a vector whose elements are random variables.
 random matrix is a matrix

	Random	Realization
Scalar variable	X	x
Vector (bold)	\mathbf{X}	\mathbf{x}
Matrix (bold)	\mathbf{X}	\mathbf{X}

Data Visualization

- ① Box plot
- ② Scatter plot
- ③ Violin plot