

**COURSE SYLLABUS
FOR FULL-TIME UNDERGRADUATE PROGRAMS**

(Issued under Decision No.1380/QĐ-ĐHKTQĐ on 15/8/2016 by the University President)

1. COURSE NAME: Multivariate Statistical Analysis 2

Code: TOKT1126

Number of Credit: 02

2. DEPARTMENT IN CHARGE OF INSTRUCTION

Office: Faculty of Mathematics for Economics

Office Hours: Working hours, the working day

Office Telephone: (84) 04 3628 3007

3. PRE-REQUISITE

Multivariate Statistical Analysis 1

4. COURSE DESCRIPTION

This module provides deep specialist knowledge for training programs in Applied Mathematics in social economics.

Factor analysis methods with specific models as the main component analysis, analyzing corresponds; Canonical correlation analysis; The method of classification and discriminant analysis.

All contents will be presented with skills using specialized statistical softwares according to international standards (SPSS and Stata). Empirical analysis with data in Vietnam, the World and the region.

5. COURSE OBJECTIVES

By the end of this course, students should be able to :

- Understand the methodology of socio-economic research with statistical models
- Understand the theoretical basis of formation and solving problems multivariate statistical analysis
- Understand the scope and conditions of use of the models and methods of analysis
- Develop good analytical skills with the statistical model and using specialized softwares

- Gain deep insights into the socio-economic situation of Vietnam and possibilities, the use of statistical analysis model through the use of empirical data sets in Vietnam

6. COURSE CONTENTS

TENTATIVE SCHEDULE

<i>No</i>	<i>Contents</i>	<i>Total hours</i>	<i>In details</i>		<i>Notes</i>
			<i>Theory</i>	<i>Practice, Discussion, Exams</i>	
1	Chapter 7	9	5	4	<i>Students need to practice more on the computer</i>
2	Chapter 8	5	5	0	
3	Chapter 9	0	self-study	0	
4	Chapter 10	4	3	1	
5	Chapter 11	6	4	2	
6	Chapter 12 practical exercise	6	4	2	
	Total	30	21	9	

CHAPTER 7– FACTORS ANALYSIS - METHOD PRINCIPAL COMPONENTS ANALYSIS

This is the opening chapter to a class model using the knowledge of higher mathematics. The theoretical basis for the following chapters is formed in this chapter. Principal components method is used effectively will form the foundation for the next chapter.

In addition to theoretical knowledge being fully presented, analytical skills with the support of softwares are also presented in detail with experimental data in Vietnam.

7.1. Factor analysis problem

7.2. The problem and the basic idea

7.2.1. Examples

7.2.2. Problem analysis factor in 2 and 3 dimensional space

7.2.3. The problem in p -dimensional space

7.2.4. Problem analysis variables in R^n

7.2.5. Contact between R^p and R^n analysis

7.2.6. Analysis of individual and additional variables

- 7.3. Description, summary data
 - 7.3.1. Data and Features
 - 7.3.2. Individual space
 - 7.3.3. Space variables
- 7.4. Create variables and projections in linear space
 - 7.4.1. Create a new variable
 - 7.4.2. Projection
- 7.5. The principal component analysis
 - 7.5.1. Projection onto the subspace
 - 7.5.2. Principal axe, the principal component and the principal factor
 - 7.5.3. Define λ and select number of principal components
 - 7.5.4. Redisplay the data
- 7.6. In case of using the covariance matrix measurement
- 7.7. Analyze and evaluate the results of applying the method of the principal component of factor analysis
 - 7.7.1. The general tests
 - 7.7.2. Correlation of the principal components and the original variables
 - 7.7.3. The reflection coefficient of the individual contacts and the principal components
 - 7.7.4. Analysis of the variables in R^n
 - 7.7.5. Relations R^p points and lines in the column points in R^n
 - 7.7.6. Analysis of additional individuals
 - 7.7.7. Illustrations analysis on R^p and R^n
 - 7.7.8. For example, by the direct calculation
- 7.8. The rate on the super space explains projector
 - 7.8.1. Explains general level (total)
 - 7.8.2. Rate parts
- 7.9. Criteria for selecting the principal component analysis
 - 7.9.1. Criteria theory
 - 7.9.2. Criteria Kaiser
 - 7.9.3. Criteria ratio variance
- 7:10. Model principal component regression
 - 7.10.1. Multicollinearity and principal component regression
 - 7.10.2 The impact of the independent variables on the dependent variable
 - 7.10.3. For example, the model efficiency business assets
- 7:11. Stata and SPSS to analyze the principal components
 - 7.10.1. Stata to analyze principal components
 - 7.10.2. SPSS to analyze the principal components
- 7:12. The role of these variables and how to select variables
 - 7.12.1. Determine the purpose of analysis and select the original variable

- 7.12.2. Correlation analysis and initial variable filter
- 7.12.3. Selection of analytical models
- 7.12.4. Factor analysis method principal components

References of the chapter:

- 1 - Ngo Van Thu, 2015, Statistics practice with the help of Winstata and SPSS, Published NEU, Chapter 7.
- 2 - T.W. Anderson, 1971, An Introduction to Multivariate Statistical analysis, Wiley. Chapter 11.
- 3 - Benzecri, 1990, L'Analyse des donnees en sociologie, IdP Universitaires de France.
- 4 - Gilbert Saporta, 1993, Analyse des donnees, L'INSEE. Chapter 1.
- 5 - Allen Webster, 1992, Applied statistics for Business and economics, Irwin
- 6 - Michel Volle, 1980, Analyse des donnees, Economica. Chapter 5-6.

CHAPTER 8 – CORRESPONDENCE ANALYSES

Student self-guided readings

Introduction class problem analysis leading to corresponds models and the application of factor analysis to analyze corresponds one-dimensional.

- 8.1. Corresponding problem analysis
- 8.2. Contingency table and the metrics for cloud data
 - 8.2.1. Descriptive geometry a contingency table
 - 8.2.2. Describe the structure matrix and math
- 8.3. When measurement distance squared
- 8.4. Analyzing the Principal component in the projection space
 - 8.4.1. The principal component analysis and cultural center not trivial factor according to primary data lines and columns
 - 8.4.2. Analyzing the principal component is not the center of the projection of the cloud
 - 8.4.3. Conversion formula
 - 8.4.4. Trace matrix and the reproduction of data
- 8.5. Performed simultaneously
 - 8.5.1. The projection of the line and the column on the same super space
 - 8.5.2. Description and the contribution of the line point, the column for the principal factors
- 8.6. Apply corresponds analysis for canonical analysis for both quantitative variables
 - 8.6.1. Description quantify qualitative variables
 - 8.6.2. Quantifying and analyzing qualitative variables corresponds
 - 8.6.3. Analyzing the two groups switch with indicator

- 8.6.4. Performing simultaneous optimization of the individual signs
- 8.6.5. Description and signs
- 8.6.6. Description simultaneously in two metrics
- 8.7. Corresponding analysis on SPSS - An example
- 8.8. Stata corresponding analysis - An example

References of the chapter:

- 1 - Ngo Van Thu, 2015, Statistics practice with the help of Winstata and SPSS, Published NEU, Chapter 8.
- 2 - T.W. Anderson, 1971, An Introduction to Multivariate Statistical analysis, Wiley. Chapter 12.
- 3 - Benzecri, 1990, L'Analyse des donnees en sociologie, IdP Universitaires de France.
- 4 - Gilbert Saporta, 1993, Analyse des donnees, L'INSEE. Chapter 2.
- 5 - Allen Webster, 1992, Applied statistics for Business and economics, Irwin.
- 6 - Michel Volle, 1980, Analyse des donnees, Economica. Chapter 8-9.

CHAPTER 9 – MULTIPLE CORRESPONDENCE ANALYSIS

This chapter equips students with knowledge corresponding general analysis on the basis of expanding the model in Chapter 8. It introduces using different models to meet the kind of problem analysis with different variables in the analysis. It also develops skills to use softwares and how to analyze answers.

- 9.1. Describing and encoding data
 - 9.1.1. Data
 - 9.1.2. Description Order data
- 9.2. Corresponding multivariate analysis
 - 9.2.1. Case $p = 2$
 - 9.2.2. General case ($p > 2$)
- 9.3. Corresponding analysis on SPSS multivariate
 - 9.3.1. Direct analysis tables X
 - 9.3.2. Analysis on the frequency table

References of the chapter:

- 1 - Ngo Van Thu, 2015, Statistics practice with the help of Winstata and SPSS, NEU Press, Chapter 9.
- 2 - T.W. Anderson, 1971, An Introduction to Multivariate Statistical analysis, Wiley. Chapter 12.
- 3 - Benzecri, 1990, L'Analyse des donnees en sociologie, IdP Universitaires de France.

- 4 - Gilbert Saporta, 1993, Analyse des donnees, L'INSEE. Chapter 3.
- 5 - Allen Webster, 1992, Applied statistics for Business and economics, Irwin
- 6 - Michel Volle, 1980, Analyse des donnees, Economica. Chapter 12.

CHAPTER 10 – CANONICAL CORRELATION ANALYSIS

This program supplements the extensive knowledge of math class correlation analysis and regression. The content of this chapter is theoretical. With the approach of analyzing the canonical people solve theoretical problems of data analysis. Regression multivariate analysis, corresponding analysis, analysis of discriminant (Chapter 12) are assumed to be specific cases of a canonical correlation analysis. The analytical skills are mainly targeted for undergraduate students.

- 10.1. Canonical correlation analysis with two groups of variables
 - 10.1.1. Problem analysis with two canonical variables
 - 10.1.2. Matrix description of the problem and the solution
 - 10.1.3. Explain geometry of the problem solution
 - 10.1.4. Find all the pairs of canonical variables
 - 10.1.5. The testing
 - 10.1.6. Descriptive geometry canonical analysis results
 - 10.1.7. Analysis on SPSS and STATA rules
- 10.2. Canonical correlation analysis overview

References of the chapter

- 1. Benzecri, 1990, L'Analyse des donnees en sociologie, Presses Universitaires de France Imprimerie des.
- 2 - Gilbert Saporta, 1992, Analyse des donnees, Ecole national de la Statistique et de l'administration économique.
- 3 - Dale J. Poirier, 1995, intermediate statistics and Econometrics.
- 4 - David W. Hosmer, Stanley Lemeshow, 1989, Applied Logistic Regression, Wiley.
- 5 - Jean-marie bouché, Gilbert Saporta, 1998, L'Analyse des donnees, Que sais - Je.
- 6 - Helene Erkel-Rousse, 1992, Analyse et programmation des donnees Applications. L'ensae.
- 7 - L. SIMAR, 2003, An introduction to multivariate data analysis, diffusion Cica Universitaire.
- 8 - Michel Volle, 1994, Analyse des donnees. Economica.
- 9 - IBM SPSS Modeler Algorithms Guide 15, 2009.
- 10 Joseph F. Hair, Jr., Rolph E. Anderson, Ronald L. Tatham and William C, 1998, Multivariate Data Analysis, 5th edition. Copyright © Prentice Hall, Inc., Chapter 8.

11 - Ngo Van Thu, 2015, The statistical practice, NEU Press, Chapter 10.

CHAPTER 11 – CLUSTER ANALYSE

Introducing statistical classification model with some different approaches. How to choose the model and solution method. The algorithm can be used and some softening algorithms on dedicated software. Skills analysis and evaluation of the results. This chapter and Chapter 12 focus on theoretical issues, practice and and skill training.

11.1. Some basic concepts

11.1.1. The distance and the difference

11.1.2. Combinatorial problems in subgroups

11.2. The method of dividing the class

11.2.1. The method of dynamic cloud type

11.2.2. Layered with binary variables

11.3. Layered hierarchy

11.3.1. Hierarchical classification allowed

11.3.2. The combined measure based on the spread (difference)

11.3.3. Standards and methods inertia Ward

11.3.4. Layered with non-Euclidean distance

11.4. Subclass for variable

11.4.1. class variables - split according to overall research perspective

11.4.2. Distance measures for quantitative variables

11.4.3. Distance measures for qualitative variables

11.4.4. Lerman and algorithmic approach based on the correctness of the contact

11.5. The problem of classification on SPSS and Stata

11.5.1. Procedure K - Means Cluster

11.5.2. Hierarchical subclass instances

11.5.3. Hierarchical layered variables

11.5.4. The subclass procedure in Stata for individuals

References of the chapter:

1 - Ngo Van Thu, 2015, Statistics practice with the help of Winstata and SPSS, Published NEU, Chapter 11.

2 - Nguyen Quang Dong: Textbook of Advanced Economic Publishing House KHKT.2007, Chapter 3

3 - T.W. Anderson, 1971, An Introduction to Multivariate Statistical analysis, Wiley. Chapter 6.

- 4 - Benzecri, 1990, L'Analyse des donnees en sociologie, IdP Universitaires de France.
- 5 - Gilbert Saporta, 1993, Analyse des donnees, L'INSEE. Chapter 4.
- 6 - Allen Webster, 1992, Applied statistics for Business and economics, Irwin
- 7 - Michel Volle, 1980, Analyse des donnees, Economica. Chapter 12-15.

CHAPTER 12 – DISCRIMINANT ANALYSIS

This chapter equips students with a different approach of the problem classification individuals and variables. It introduces technical analysis and the current popular applications of this model class, expands in the experimental model of the credit rating companies and the financial institutions, and outlines the results applied in Vietnam.

- 12.1. A specific illustration of the discriminant analysis
 - 12.1.1. Layered cloud data in a given qualitative variables
 - 12.1.2. Layering a cloud data no classification criteria
- 12.2. The geometric method
 - 12.2.1. Data and concepts
 - 12.2.2. Variance analysis
 - 12.2.3. two groups
 - 12.2.4. Geometric rules for additional elements
- 12.3. Probabilistic methods
 - 12.3.1. Bayes Rule
 - 12.3.2. Multidimensional standard model
 - 12.3.3. The force of the grouping rules
 - 12.3.4. Logistic regression
- 12.4. Analysis with SPSS and Winstata discriminant
 - 12.4.1. Analysis on SPSS discriminant
 - 12.4.2. Analyze discriminant in Stata

References of the chapter:

- 1 - Ngo Van Thu, 2015, Statistics practice with the help of Winstata and SPSS, Published NEU, Chapter 12.
- 2 - Nguyen Quang Dong: Textbook of Advanced Economic Publishing House KHKT.2007, Chapter 3
- 3 - T.W. Anderson, 1971, An Introduction to Multivariate Statistical analysis, Wiley. Chapter 11.
- 4 - Benzecri, 1990, L'Analyse des donnees en sociologie, IdP Universitaires de France.
- 5 - Gilbert Saporta, 1993, Analyse des donnees, L'INSEE. Chapter 5.
- 6 - Allen Webster, 1992, Applied statistics for Business and economics, Irwin

7 - Michel Volle, 1980, Analyse des donnees, Economica. Chapter 9.

7. REQUIRED TEXTBOOK & COURSE MATERIALS

1 - Ngo Van Thu, 2015, the statistical practice, Publishing NEU

2 - Database (as required by the lecturer)

- Survey data 2002-2004-2006-2008 Living Standards
- Enterprise Survey data 2000-2009
- Data on the stock market from 2000 to 2010.
- Data on a number of thematic surveys.

3- Statistical software: SPSS, STATA, EXCEL

8. RECOMMENDED TEXTS & OTHER READINGS

1 - Agresti, Alan, 2007, An introduction to categorical data analysis, ISBN 978-0-471-22618-5. A John Wiley & Sons.

2 - Allen Webster, 1992, Applied Statistics for Business and Economics, 1992 Irwin.

3 - Dale J. Piorier, 1995, intermediate statistics and Econometrics.

4 - Douglas A, Lind, William G.Marxhal, Robert D.Mason, 2001 in Business & Economics Statisstical Techniques, McGraw-Hill.

5 - J.K. Lindsey, 2007, Generalized Linear Models Applying, Limburgs Universitair Centrum, diepenbeek.

6 - K.M.Ramachandran, Chris P.Tsokos, 2009, Mathematical statistics with Application. Elsevier, Chapter 8.

7 - Russell Davidson, James G. MacKinnon, 1993, Estimation and inference in econometrics, Oxford University.

8 - S. Moore, P.McCabe, A.Craig, 2009, Introducton to the Practice of Statistics, Freeman W-H, Chapter 10, 11, 14

9 - Thomas H. Wonnacott, Ronald J. Wonnacott, 1990, statistics for business and economics Introductory, Wiley, Chapter 11-15.

10 - Le Van Phong, Tran Trong Nguyen, 2011, Theory of Probability.

11 - Nguyen Minh Thang, 1987, Survey Sampling, Statistical Publishing House.

12 - Hoang Dinh Tuan, 2010, Mathematical Economic, Science and Technology Publishing House.

13 - Ngo Van Thu, 2015, the statistical practice, NEU Press, Chapter 6.

14- Nguyen Quang Dong, 2002, Advanced Econometrics, Published KHTK.

15 - Nguyen Cao Van, Tran Thai Ninh, Ngo Van Thu, 2011, the Curriculum Theory of Probability and Statistics, Published NEU.

9. ASSESSMENT & GRADING POLICY

- Scale (point) : 10
- Structure of points:
 - + Discussion point: 10%
 - + The exercise, check out: 30%
 - + The final examination period: 60%
- Conditions of the final exam:
 - + Must attend at least 80% of courses

Hanoi, 2016

HEAD OF DEPARTMENT

PRESIDENT

(signed)

(signed)

PhD. Nguyen Manh The

Prof.Dr. Tran Tho Dat