

Syllabus of Data Analysis*

Course Name/Title: Data Analysis* **Course code:** 63924

Course Type: Basic Course, Optional Course

Total Teaching Hours: 48 (Classroom Hours:36 Laboratory Hours:12)

Course Credit: 3

I Course Objective

The purpose of this course is to enable the students to master the ability of data collection and statistical analysis for the data collected in the investigation by the basic theory and methods of data analysis on the basis of the basic course probability theory and mathematical statistics. Master the operation of professional statistical software SAS from the establishment of data files, basic analysis to the relevant regression analysis and multivariate analysis of the entire process of operation; to enable students to master a practical and highly professional skills; the same time to lay the foundations for other professional .

- To provide students with a good understanding of the concepts and methods of linear regression and multivariate statistical analysis, described in detail in the syllabus.
- To help the students to study large amounts of data in practical problems, and to make reasonable statistical inferences and make accurate forecasts from data.
- To help the students solve problem of random phenomena by using the basic principles of mathematical statistics, to explore the problems from multiple angles.
- To help the students to improve the ability of information access, information storage, information extraction and information application.

II Correlations between Course Objectives and Graduation Requirements

Graduation Requirements	Graduation Requirements Index Point	Course Objectives
Graduation Requirement 1		Course Objective 1,2
Graduation Requirement 2		Course Objective 2,3

(Note: Basic courses and specialized courses must correlate with the graduation requirements as specified in the Program outline. The correlated graduation requirement index point must be put before the descriptive phrases or sentences. General courses are exempted from this rule.)

III Course Content

Theoretical Teaching Content

1. Descriptive statistics
 - 1.1 Numerical characteristics of the data
 - 1.2 The distribution of data
 - 1.3 Numerical characteristics and correlation analysis of multivariate data
2. Regression analysis
 - 2.1 Linear regression model
 - 2.2 Stepwise regression

- 2.3 Logistic regression model
- 3. Principle component analysis and factor analysis
 - 3.1 Principle component analysis
 - 3.2 Factor analysis
- 4. Discriminant analysis
 - 4.1 Mahalanobis distance discriminant analysis
 - 4.2 Bayes discriminant analysis
- 5. Cluster analysis
 - 5.1 Distance and similarity coefficient
 - 5.2 Cluster algorithm
- 6. Introduction of SAS procedures of commonly used data analysis methods
 - 6.1 Introduction of the SAS system
 - 6.2 SAS procedures of commonly used data analysis methods

Experimental Teaching Content

Data analysis experiments adhere to the "student-centered" principle, around the basic content of data analysis, so that students make full use of computer statistical software to experience and understand the basic concepts and conclusions of mathematics, so that students access to the use of computers to solve a variety of mathematical problems. Through the data analysis experiment, students should be able to independently and skillfully use the software for statistical calculations, so that theory is clear, hands-on, and write the experimental report. Teachers introduce a little thought, ask questions and tasks, and let the students try to discover and summarize the laws through their own hands and observe the experimental results.

Experimental Project 1: Descriptive statistics. Master the use of SAS to create and display data files. Master SAS procedure of UNIVARIATE and CAPABILITY.

Experimental Project 2: Regression analysis. Master linear regression analysis by using SAS software for parameter estimation, analysis of variance, parameter hypothesis test, etc. Master SAS procedure of REG .

Experimental Project 3: Principle component analysis and factor analysis. Master the use of SAS software for principal component analysis and factor analysis. Master SAS procedure of PRINCOMP and FACTOR.

Experimental Project 4: Discriminant analysis. Master the use of SAS software for discriminant analysis. Master SAS procedure of DISCRIM.

Experimental Project 5: Cluster analysis. Master the use of SAS software for cluster analysis. Master SAS procedure of CLUSTER.

IV Practice Plan and Requirement

Theoretical course or theoretical teaching content class allocation table

T. Hours Teaching Content	T. Methods	Theoretical teaching	Exercise class	subtotal
Descriptive statistics		6	0	6 hours

Regression analysis	8	1	9 hours
Principle component analysis	5	1	6 hours
Factor analysis	3	0	3 hours
Discriminant analysis	7	1	8 hours
Cluster analysis	4	0	4 hours
total	33	3	36 hours

Experimental class, curriculum design, case and other practical courses or practical teaching content class allocation table

Experimental Project	Content summary	Equipment and experimental environment	h o u r s	Number of students	property	requirement
Descriptive statistics	Master the use of SAS to create and display data files. Master SAS procedure of descriptive statistics	Computer and SAS software	3	1	basic	compulsory
Regression analysis	Master linear regression analysis by using SAS software for parameter estimation, analysis of variance, parameter hypothesis test, etc.	Computer and SAS software	3	1	basic	compulsory
Principle component analysis and Factor analysis	Master the use of SAS software for principal component analysis and factor analysis.	Computer and SAS software	3	1	basic	compulsory
Discriminant analysis and Cluster analysis	Master the use of SAS software for discriminant analysis and cluster analysis.	Computer and SAS software	3	1	baisc	compulsory

V Suggestions for Teaching

1. Teaching method

The use of heuristic and case-based teaching methods to strengthen the exchanges between teachers and students, to guide students to think independently, to strengthen the training of

scientific thinking.

2. Teaching means

In the teaching of data analysis course, multimedia technology can be used to assist teaching, including: writing content is too difficult to use blackboard to express the content, form, data, mathematical figures, history of mathematics, preface class, review class.

VI Suggestions for Course Learning

Extracurricular exercises are an important part of guiding students to learn and examining the teaching effect. It is also a symbol that reflects the requirements of the course. The selection of the exercises should focus on the teaching of the basic exercises and reduce the requirements of the technical exercises. It should try to select the students' solve problems, consolidate the knowledge, but also close to the application of the actual students can stimulate interest in the operation. Generally every 3 hours of a homework, you can use the exercises on the text, but also self-compiled exercises set. In addition can also be arranged special papers and reading notes and extra-curricular reading and other rich content.

To improve students' ability to learn and acquire knowledge is an important part, it is necessary to properly arrange the content of self-taught students. In order to extend the limitations of the space-time field of classroom teaching, students can use the network classroom to self-study, review and personalize the study through the use of the rich teaching resources provided by the student's campus network. In addition, the website platform also provides online work, Q & A system, test system, communication platform function, in order to improve the autonomy of students learning and enhance the exchange between teachers and students.

VII Assessment

(Assessment content and methods must correlate with course objectives.)

Assessment Methods or Approaches	Assessment Requirements	Assessment Weighting	Evaluation of Course Objectives
classroom performance	attendance and so on	10%	
conventional assignments	5 times	30%	
final exam	open-book	60%	

VIII Textbooks and References

1. Applied Multivariate Statistical Analysis, Third Edition, Wolfgang Karl Härdle, Léopold Simar, 2012, Springer.

2. Applied Linear Regression, Third Edition, Sanford Weisberg, 2005, Wiley Series in Probability and Statistics.
3. Data Analysis, Fan Jincheng, Mei Changlin, 2010, Science Press.

Written by: Chen Youyou

Reviewed by: Yang Jianping

Date: 2017 年 3 月 14 日