

RES230 Data Science II

3rd and 4th quarters, Sophomore

Instructor	INOUE HIROYUKI
Style of Class	Lecture
Number of Credits	2
Day and Period	Thursday, period 3

Course Description

In recent years, a wide variety of data has become accessible online. It has also become possible to make use of large volumes of data, known as “big data.” It is crucial to be able to handle such data correctly, extract the necessary information and utilize them when addressing problems in local and international society, and this requires the capacity to use data effectively. This course is about how to handle and process data, centering on database technologies and data analysis methods. Students will acquire the ability to process data and interpret information from it. They will also study database fundamentals such as data expression methods, and acquire methods for extracting information from databases using SQL. To apply these skills, we create databases from actual government statistics and other data obtained online and extract essential information therefrom. Then we analyze the data statistically, focusing on multivariate analysis, and interpret it. Specifically, multiple regression analysis, discriminant analysis, cluster analysis, principal component analysis and factor analysis are applied to actual data in order to understand their characteristics and learn how to interpret and use data as information. We will also touch on data mining methods. All these things will be learned and acquired using actual computer software tools.

Course Objectives

Students will:

- (1) understand how to handle data,
- (2) acquire methods in extracting and interpreting information from databases,
- (3) research statistical data available online,
- (4) learn methods for creating databases of gathered data,
- (5) consider information that can be extracted from specific statistical data,
- (6) understand how to use tools of data analysis, and
- (7) experience gathering and analyzing data that they are interested in.

Prerequisites

Basic Computing, Fundamentals of Information Technology, Data Science I

Class Materials

Textbook:

Hayami Haruo, Miyazaki Nobuyoshi, Yamazaki Haruaki, *Databases* (Databases), Ohmsha, 2006.

Reference materials:

Handouts will be distributed

Course Method

This class takes the form of lectures and practical work on computers. The instructor gives lectures, and students develop their understanding of the lecture contents by doing computer-based exercises. Exercises will be set for each topic, and should be undertaken both within and outside of class.

Evaluation/Assessment

Submit exercises.

Grading

40% Exercises

Course Schedule

Week 1: Fundamentals of databases and data analysis

Orientation to content studied throughout the course; fundamentals and social necessity of databases; introduction to actual cases of data analysis

Week 2: Fundamentals of relational databases

Concepts of relational models, primary keys, and foreign keys, which provide the foundations of relational databases.

Week 3: Database design

Essential theories for database design, including first normal form, second normal form, and third normal form

Week 4: Extracting information using SQL (1)

Study of database manipulation language SQL using Microsoft Access, centering especially on the fundamentals of SELECT statements; learning while handling data in practice

Week 5: Extracting information using SQL (2)

Continued study of the fundamentals of SQL, including GROUP BY clauses in SELECT statements and sub-queries; learning while extracting actual information from databases

Week 6: Obtaining and creating databases out of government statistics, etc.

Methods for obtaining government statistics from the Ministry of Internal Affairs and Communications portal site; gathering necessary data and attempting to create databases

Week 7: Using software to search databases

Query generation functions in Microsoft Access; methods for extracting information from databases easily

Week 8: Exercises on databases using local data and international data

Practical exercises on databases learned thus far; data to be extracted from databases created from data gathered by students themselves, as well as exercises prepared by the instructor

Week 9: Data and multivariate analysis

Explanation of methods for processing various types of big data, to gain a deeper understanding of data analysis; special focus on understanding and using software for multivariate analysis

Week 10: Analyzing data through multiple regression analysis and discriminant analysis

Multiple regression analysis and discriminant analysis; exercises in multiple regression analysis and discriminant analysis using actual data

Week 11: Analyzing data through cluster analysis

Cluster analysis; exercises in cluster analysis using actual data

Week 12: Principal component analysis and factor analysis

Principal component analysis and factor analysis; exercises in principal component analysis and factor analysis using actual data

Week 13: Data mining methods (including text mining)

Methods used in data mining; specifically decision tree analysis, self-organizing maps, neural networks, etc., as well as touching on text mining

Week 14: Exercises in data analysis using local and international data

Practical exercises in data analysis methods learned thus far, centering on multivariate analysis; analysis conducted on data gathered by students themselves, as well as exercises prepared by the instructor

Week 15: Wrap-up, review, and final assignment

Review of Class content; work on final assignment

Preparation and Follow-up

Preparation: Research the keywords indicated in the content of each class. Read through the textbook and materials distributed in advance.

Follow-up: Revise class content and repeat exercises of the methods using data handled in class and related data.