

The University of British Columbia
Faculty of Forestry
Forestry 531 -- Applied Multivariate Statistics
Course Outline for January through April, 2010

Calendar Description:

FRST 531 (3) Multivariate Statistical Methods. Multivariate analysis of variance, cluster, principal components, factor, canonical, and discriminant analysis. Theory and conceptual background are presented but emphasis is on selection of appropriate analysis and interpretation of results. Examples from forestry and related fields are analysed by computer programs available at UBC. Given in alternate years.

NOTE: An introductory course in the theory of probability and statistics is required. A course in the use of matrix algebra would be an asset. The preferred computer package for this course is SAS, although it is not expected that students will have had any exposure to this package.

Lectures:

Monday Wednesday Friday 12:00 to 13:00 a.m. FSC 1003

Instructor:

Dr. Valerie LeMay FSC 2039 604-822-4770 Valerie.LeMay@ubc.ca

Graduate Teaching Assistant:

Enrico Maria Goberti

Evaluation:

Assignments	35%
Final Exam	40%
Project	25%

NOTE: There will be six assignments for the course, plus additional exercises that you may choose to do (not graded, just for practice). Auditors must do the assignments, as this is part of the learning process, but do not do the project nor the final exam. One mark will be deducted for each day that an assignment is late.

NOTES:

- (1) The schedule and content of the course may be modified, depending upon the interests and background of students in the class.**
- (2) There will be no classes on April 2 and 5 (Easter Holidays), on April 8 and 9, nor during the reading weeks from February 14 though February 28.**

References:

Course Notes: You will need to pay for a set of course notes, as discussed in class.

Textbooks: There are two textbooks in the bookstore that are listed for this course, and are needed in order to gain an understanding of multivariate statistics. You **should buy at least one of these textbooks, as the concepts are difficult and you will need to use MORE than just the course notes.** The two textbooks are:

Manly, Bryan. 2005. Multivariate statistics: A Primer, 3rd edition. Chapman & Hall/CRC Press, New York. [mostly biological examples]

Tabachnick, B.C. and L.S. Fidell. 2001. Using multivariate statistics, 4th ed. Allyn & Bacon, Toronto. [The 5th edition has now been released, and may be the one purchased by the bookstore] [mostly social examples]

Some other references you may wish to consider are:

Ayres, F. 1962. Matrices. Schaum outline series. McGraw-Hill Book Company, Toronto. (This has examples with answers and is very useful for learning matrix algebra.)

Dillon, W.R. and M. Goldstein. 1984. Multivariate analysis, methods, and applications. John Wiley and Sons, Toronto. (excellent text for more on the details behind multivariate methods.)

Green, P.E. and J.D. Carroll. 1976. Mathematical tools for applied multivariate statistics. Academic Press, New York. (Includes the mathematics behind multivariate statistics based on both matrix algebra and geometric viewpoints.)

Hair, J.F., R.E. Anderson, and R.L. Tatham. 1987. Multivariate data analysis. 2nd ed. MacMillan Publishing Company, New York.

Kleinbaum, D. G., L.L. Kupper, K.E. Muller. 1988. Applied regression analysis and other multivariable methods. PWS-Kent Pub. Co., Boston, Mass.

e-books in the UBC library:

Der, G. and Brian S. Everitt. 2002. A handbook of statistical analyses using SAS. Chapman & Hall, Boca Raton, FL. Available from www.library.ubc.ca, then select “e-resources” and then “indexes and databases” and type in “STATSNETBASE” to access this full book. [biological including biomedical examples]

Härdle, Wolfgang and Léopold Simar. 2007. Applied Multivariate Statistical Analysis, 2nd ed. Springer, New York. Available from www.library.ubc.ca, then select “e-resources” and then “indexes and databases” and type in “Springer” to access this full book. [mostly social science examples]

Course Content and Schedule:

I. Introduction to multivariate statistics analysis. Classification of techniques and selection of topics for discussion. [Week 1]

II. Background math and statistics for multivariate statistics [Weeks 1, 2, and 3]

1. Types of variables; format of data
2. Matrix algebra basics, review.

Practice Exercise: Matrix algebra

3. Use of matrix algebra for multivariate analysis; covariance and correlation matrices.

Practice Exercise: Commonly used matrices

4. View of matrices and transformations of data via geometry
5. More matrix algebra and geometry including eigenvectors and eigenvalues (principle components analysis, PCA).

III. Commonly used Multivariate Techniques for one group of variables:

1. Applications of PCA. [Weeks 3, 4, 5] **NOTE: Some classes during the week of January 18 will be in FSC 1406 (computer lab), when you will be introduced to SAS, and use this package to get PCA output.**

Assignment 1: Principle Components Analysis, Due Wednesday, February 3

2. Factor analysis beginning with PCA. [Weeks 5 and 6]

Assignment 2: Factor Analysis, Due Wednesday, February 10

NOTE: Reading Weeks is after Week 6, Feb 14 through 28

3. Cluster Analysis [Weeks 7 and 8]

Assignment 3: Cluster Analysis, Due Wednesday, March 10

IV. Commonly used Multivariate Techniques for two groups of variables:

1. Discriminant analysis [Weeks 8 and 9]

Assignment 4: Multivariate Discriminant Analysis, Due Wednesday, March 17

2. Canonical Correlation Analysis [Weeks 9 and 10]

Assignment 5: Canonical Correlation Analysis, Due Wednesday, March 24

3. MANOVA [Weeks 10 and 11]

Assignment 6: Multivariate Analysis of Variance, Due Wednesday, March 31

V. Summary of methods covered and guide for selection of method to use. [Week 12]

VI. Exam Practice [Weeks 11 and 12]

VII Project Time [Week 13]

Project Due Date: Wednesday, April 21, by 4:30 pm.

Final Exam Date: To be announced, but will be within the exam period from April 19 to May 1 inclusive. Saturdays are included in the examination schedule.

Project Guidelines

Score: 5 (outline) + 15 (project) = 20 (maximum score)

Project Objective: The project part of the requirements for completion of this course is intended to give you the opportunity to use the tools presented in the course on a problem that is of interest to you. You must choose a project with a well-defined objective and data that you can use. Methods used in the course must be used in your project.

Outline: To obtain some early feedback on your chosen project, an outline for the project is due by Friday, March 5. In the outline, indicate:

- The objective of the research
- The data you will use
- The methods you might use

I will give you feedback on the project and let you know if this is too much for the project. The outline is worth 5 points towards a total maximum project score of 20 points.

PROJECT OUTLINE: Due Friday, March 5

PROJECTS DUE: Wednesday, April 21, 2010, by 4:30 pm (subject to the final exam schedule)

Project Report Specifications:

Length: The project should be from seven to 12 typewritten pages.

Topic: The subject of the project should be a research project that involves analysis of some data of interest to you using the techniques learned in the course.

Format: The project must be written as a formal research report with:

- a title page,
- introduction: including relevant background literature and the objectives for the paper
- materials and methods: often divided into “data” and “analyses”
- results: graphs and analyses formatted into tables and figures, that appear after they are introduced in the text. The reader should be guided to looking at important results)
- discussion: discusses your results and connects your results to previous research,
- conclusion: important conclusions and further research
- references cited: any format you wish, but MUST be consistently used)
- Appendices: These must be very well formatted into tables and graphs.

Relevant SAS outputs should be brought in as formatted figures and tables in the Results section, with captions (at the bottom for figures and at the top for tables) and these must be referenced in the text. Large outputs that are relevant should be put into formatted, labeled Appendices. This must be original work, with due credit to any reference material you have used.