The University of British Columbia
Faculty of Forestry

Forestry 531 Applied Multivariate Statistics

Course Outline for January through April, 2015

Calendar Description:

FRST 531 (3) Multivariate Statistical Methods. Multivariate analysis of variance, cluster, principal components, factor, canonical, and discrimiant 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 analyzed 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. We will use SAS in this course, although it is not expected that students will have had any exposure to this package. An optional workshop using R to do the same analysis is being planned for April, 2015.

Lectures:
Monday Wednesday Friday 1200 to 1300 (noon hour) FSC 1003

Office Hours:
Monday Wednesday (to be announced)

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

Graduate Teaching Assistant:
Omnia Ibrahim, PhD Candidate

Evaluation:

Assignments	35%
Project	25%
Final Exam	40%

NOTES on Evaluation:

One mark will be deducted for each day that an assignment is late. No marks will be given for a late assignment if the assignments have already been marked and returned to other students. Auditors must do the assignments, as this is part of the learning process, but only comments will be provided not a grade. Also, Auditors do not do the project nor the final exam.

NOTES on the Schedule:

(1) There will be no classes on Feb. 9 (family day), on April 3 and 6 (Easter Holidays), nor during the reading week from February Feb. 16 to 20. Also, there will be no classes March 9 to 13.
(2) Final Exam Date: To be announced, but will be within the exam period from April 14 to April 29 inclusive. Saturdays are included in the examination schedule.

References:

Course Notes: **You will need to pay a fee for the course notes that will be handed out in class. The fee will be $20**, and you can pay this fee at the Forest Resources Management Department office, FSC 2045 (Scheyla). I will monitor the payment of the fee.

Strongly Recommended Textbooks:

I strongly recommend that you purchase a textbook on multivariate statistics will examples on your research area. **In terms of textbooks you may consider buying, here are some recommendations that past students have indicated that they do like. These have been placed on hold in the Woodward library (hopefully – they are very popular and tend to "disappear")**. You can borrow these for 1 day to read more on the topics we cover and to help you decide on a textbook to buy.

1. "Multivariate statistical methods: A primer" by Bryan F.J. Manly

This is an excellent book with examples from natural sciences that is published by Chapman & Hall/CRC Press. It is paperback -- so not too expensive to buy.

2. "Using multivariate statistics" by Tabachnick, Barbara G, Fidell, Linda, and others.

This has examples from social sciences and is published by Pearson/Allyn & Bacon.

Some other references you may wish to consider are:

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

Bartholomew, D.J., F. Steele, I. Moustaki, and J.I. Galbraith. 2008. Analysis of multivariate social science data. 2nd ed. CRC Press, New York.

Dillon, W.R.and M. Goldstein. 1984. (*and more recent editions*) Multivariate analysis, methods, and applications. John Wiley and Sons, Toronto. (Excellent text for more on the math and other 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. (*other editions available*). Applied regression analysis and other multivariable methods. PWS-Kent Pub. Co., Boston, Mass.

McGarigal, K. S. Cushman, and S. Stafford. 2000. Multivariate statistics for wildlife and ecology research. Springer-Verlag, New York.

Learning SAS (and R):

There are many books on using SAS (using R). I recommend these two to start learning:

1. "A Handbook of Statistical Analyses using SAS (using R)" by Geoff Der and Brian S. Everitt. (not expensive – both are paperback books).
2. "An introduction to Applied Multivariate Analysis with R" by Brian S. Everitt (ebook through UBC library).

Course Content and Schedule:

I. Introduction to multivariate statistics analysis. [Jan 5]
Overview of multivariate methods and this course.

II. Background math and statistics for multivariate statistics [Jan 7 to 16]

1. Types of variables; format of data.

2. Matrix algebra basics, review. Use of matrix algebra for multivariate analysis: covariance and correlation matrices.

Practice exercise: Matrix algebra and commonly used matrices.

4. View of matrices and transformations of data via geometry and how this relates to Principle Components Analysis.

III. Introduction to SAS [January 19 and 21 will be in FSC 1406, the computer lab]

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

1. Principle components analysis (PCA). **[Jan 16 to 26]**

Exercise 1: Principle Components Analysis

2. Factor analysis beginning with PCA. **[Jan 26 to Feb 2]**

Exercise 2: Factor Analysis

3. Cluster Analysis **[Feb 4 to 13]**

Exercise 3: Cluster Analysis

Reading Week: *[Feb 16 to 20]*

V. Commonly used Multivariate Techniques for two groups of variables

1. Multivariate discriminant analysis **[Feb 23 to 27]**

Exercise 4: Multivariate Discriminant Analysis

2. MANOVA **[March 2 to 6]**

Exercise 5: Multivariate Analysis of Variance

Project Time: *[March 9 to13]*

3. Canonical Correlation Analysis **[March 16 to 20]**

Exercise 6: Canonical Correlation Analysis

4. Logistic models **[March 23 to 27]**

Exercise 7: Logistic Models (optional to hand in)

VI. Review, Final Exam, and Project Time [March 30 to April 10]

PROJECT GUIDELINES

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

Individual OR Group of Two Persons

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 have the right to use. Methods used in the course must be used in your project. This must be original work, with due credit to any reference material you have used. [See UBC information on Plagiarism, handed out in class]

Outline: To obtain some early feedback on your chosen project, an outline for the project is due by Friday, March 6. 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, Feb 27, by 4:30 pm
PROJECTS DUE: Monday, April 13 by 4:30 pm (i.e., prior to final exams schedule)

Project Report Specifications:
Length: The project should be from seven to 12 typewritten pages, 1.5 line spacing, and 11 to 12 pt font.

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: divide this into "data description" and "analyses"
- Results: graphs and analyses formatted into tables and figures with text before each table and figure indicating to the reader what is important in the table and/or figure.
    - Figure captions go below the figure
    - Table captions go above the table.
- 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 for all references listed.
- 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 first introduced and referenced in the text. Large outputs that are relevant should be put into formatted, labeled Appendices.