ST 635 Intermediate Statistical Modeling for Business

Dominique Haughton

Course Information

- Course
  - ST 635 Intermediate Statistical modeling for Business
  - Spring 2011
  - Two sections: Wednesdays and Thursdays 5-7:20 pm (you need only to attend one)
  - Centra hybrid and room TBA
- Professor Dominique Haughton
  - Office: Morison 380
  - Extension: 2822
  - Email: dhaughton@bentley.edu
- Text and Software
  - *Multivariate Data Analysis 7th ed.* by Hair et. al. 2010 available as e-book at
  http://www.coursesmart.com/9780138132323?__professorview=false&__instructor=1173314
  - *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner* by Shmueli, Patel and Bruce (SPB), inclusive of copy of **XLMiner** software
  - Statistical Software
    - **SAS** (Statistical Analysis System)
    - **SPSS** (Statistical Package for the Social Sciences), inclusive of the **Tree** tool (Decision Trees)
    - **XLMiner**
  - SAS ebook reference
    - 'Learning SAS by Example' by Ron Cody can be accessed through the library at http://libcat.bentley.edu/search/X?SEARCH=learning+sas+by+example&SORT=D&b=24x7&b=safar&b=netl

Course Information

- Course Description
  - This course considers statistical modeling situations involving multiple variables, as is commonly the situation with many business applications. The issues we will address include, for example: How do we predict who is more likely to respond to a direct mail offer? How can we identify important segments in our customer base? How do we summarize large sets of variables?
  - A central objective of the course is for participants to become knowledgeable consumers as well as users of common multivariate statistical models. The topics covered are logistic regression, clustering, factor analysis, decision trees, and other multivariate topics as time permits.

- Course Materials
  - Copies of all course materials including lecture notes, handouts, assignments, data sets and any supplemental material will be available on the course Blackboard site
  - Information on how to obtain SAS and SPSS will be given at the beginning of the semester; XLMiner is provided along with the data mining book.

- Statistical Computing
  - Facility with statistical analysis software is a required skill for the course. SAS will be the primary software used in the course; we will also use SPSS and XLMiner.

Course Information

- Course Objectives
  - Create an understanding of multivariate statistical methods that are most commonly used in business applications
    - Develop a general understanding of how each method works
    - Recognize which method is appropriate to a particular application
    - Understand how to conduct the analysis using the appropriate software
    - Learn how to interpret the statistical results in the business context
    - Understand general statistical principles well enough to facilitate learning additional methodologies beyond those covered in the course
  - Promote statistical thinking as an intellectual framework for problem solving
    - Develop hypotheses
    - Identify data with which you can test your hypotheses
    - Understand variability and the components of variation
  - Have FUN learning
  - Caveats
    - While understanding the theory behind the statistical methods is **not** the objective of the course, it is important to understand enough of the assumptions behind the techniques to ensure they are not used inappropriately (it is important to understand what you don't know!)
    - Fun is not synonymous with "easy". The material can be subtle and sometimes challenging. Some of the greatest mathematical minds in history labored to develop these ideas and techniques and to make them accessible. You will succeed if you invest the time to understand the material rather than treat it as a collection of formulas.

Course Information

- Evaluation
  - Homeworks   35%
  - Class participation   40%
  - Centra proficiency   5%
  - Final project   20%
    - The class will be divided into study groups and each group will choose an interesting problem or hypothesis that they will explore using the types of analyses that are discussed in class. Each group will perform their analyses and present their results to the class.
- Academic Integrity
  - The Bentley College Honor Code formally recognizes the responsibility of students to act in an ethical manner. The written homework in this course is meant to be an individual exercise. Students will, naturally and appropriately, talk about the problems (this is encouraged) but the final analysis and write up **must be a student's own work in its entirety.** This includes all calculations. Establishing a solid ethical foundation is an important part of your Bentley education and will enhance the value of your degree.

ST 635 and analytics careers

- ST 635 is key to access to essentially any analytics job.

- SAS is also key, notably Base Programmer certification.

- SPSS and R are also of great value.

- Analytics skills obtained need to listed on your resume; they are important buzz words, for example linear and logistic regression analysis, cluster analysis, factor analysis, decision trees.

- The set of courses ST 625, ST 635, MA 611, MA 710 and CS 605 is the best way to train effectively for analytics jobs at Bentley. These five courses plus SAS Base Programmer certification define Superanalytics status.

- If possible I would like to meet individually and as soon as possible anyone who is hoping to pursue an analytics career to help with plans to ensure the best training possible.

Syllabus

- Topics may be moved slightly from week to week depending on how long coverage actually takes.

Class meeting	Topics
1	Review course information, syllabus and project definition. Introduction to SAS Introduction to multivariate statistical analysis. Reference: Hair et. al. Chapter 1
2	Logistic regression. Handouts. SPB Chapter 10. Reference: Hair et al. Chapter 6.
3	Logistic regression
4	Logistic regression
5	Decision Trees. SPB Chapter 9.
6	Decision Trees

Syllabus

Class meeting	Topics
7	Analytics Career night
8	Hierarchical cluster analysis. SPB Chapter 14 – Reference: Hair et. al. chapter 9
9	Non-hierarchical cluster analysis. SPB Chapter 14 – Reference: Hair et. al. chapter 9
10	Principal components analysis. SPB Chapter 4 – handouts
11	Principal components analysis
12	Factor analysis – Reference: Hair et. al. chapter 3.
13	Factor Analysis
14	Additional data mining topics
15	Final presentations and party