



Department of Accounting & Law, School of Business
State University of New York at Albany

Acc 522. Statistical Methods for Business Decisions
Fall, 2006
J Gangolly

“All life is an experiment.”

"An ounce of action is worth a ton of theory..”

"Bad times have a scientific value. These are occasions a good learner would not miss."

"Build a better mousetrap and the world will beat a path to your door."

"Foolish consistency is the hobgoblin of small minds."

--Ralph Waldo Emerson

“He uses statistics as a drunken man uses lamp-posts—for support rather than illumination. “

-- Andrew Lang

“There is a certain embarrassment about being a storyteller in these times when stories are considered not quite as satisfying as statements and statements not quite as satisfying as statistics; but in the long run, a people is known, not by its statements or its statistics, but by the stories it tells.”

-- Flannery O’Connor

Welcome

Welcome to the exciting world of traditional as well as Bayesian statistics, exploratory data analysis and datamining. The emphasis in the course will be on the use of statistics and the powerful graphics provided by the object-oriented language S-Plus for the analysis and visualisation of data of special interest to auditors, information system auditors, computer security professionals, and other professionals involved in information systems risk management and forensics.

The course is fast paced, and at times rather formal in terms of statistical as well as programming constructs used.. It is therefore important that you keep up with the class at all times and not be left behind. Should you need help, seek it immediately. I am here to help you learn.

Enjoy!

Administrivia

Semester: Fall, 2006

Time: TTH: 4:15-5:35 PM

Room: BA 219

Instructor: Jagdish S. Gangolly

Office: BA 365A

Phone: (518) 442-4949

Fax: (518) 442-3944

Office Hours: TTH 2:30-4:15 PM. or by appointment

Course pages: <http://www.albany.edu/acc/courses/acc522.fall2006>

Newsgroup: groups.google.com (acc522)

Class Conduct:

The course consists of lectures, solution of problems, discussion of homework, discussion of research papers in the area, and book assignments. You are expected to do the readings well ahead of the class. Class time is to be used for the clarification of any doubts that you may have. Do not expect to merely listen to the instructor and gain knowledge. Applied statistics is a practical field backed by robust theory. A good understanding of the theory and its use in practice is essential to excel in the field. To some extent, this is a hands-on course, and you are required to demonstrate competence in the topics covered in order to receive an acceptable grade. *I shall be giving occasional homework assignments. I also shall be calling upon some of you to come to the board and discuss problems either in the textbooks, other sources, or homework assigned. You also will be required to discuss assigned research papers before the class and submit briefs.*

Software:

I shall be using the S-Plus system running under MS-windows. When using windows version in the class, I shall mostly be using command line mode and not the windows user interface, except for some graphics. The University has site licenses. You can obtain a free copy of the S-Plus software by filling in the form at <http://www.albany.edu/its/software/SWRequestForm.html>. The adventurous amongst you might like to download and play with the R language (the open source version of the S language).

I also will try to use the IMiner software that goes with S-Plus, if we are able to resolve license matters. Announcements on this will be forthcoming.

Newsgroup/e-mail:

We shall be using the class newsgroup (at <http://groups.google.com/group/acc522>. The address to send mail to the group is: acc522@googlegroups.com) extensively for making announcements regarding tests, homework, quizzes, added links to this course homepage, etc. In fact, the newsgroup will be the primary means of communication between us outside of the class. You should post to the newsgroup all your questions and doubts for clarification. You are strongly encouraged to answer queries posted by others, and such responses will count towards class participation points for grading. You should communicate with me via e-mail only for individual problems and questions. I have already added you to the group so you can start sending email to the group rightaway.

The Graduate Laboratory for Accounting Information Systems Access:

As a graduate student in the Department, you have access to the Accounting Information Systems Laboratory in the Accounting Department Suite on the third floor. All machines in the lab have S-Plus installed on them. When the license matters are settled, we will install IMiner software also on them. You will need to get the password from Ms. Lisa Scholz (442-4978) to enter the lab. Contact her in BA 365 as soon as possible.

Course Objectives:

- Understanding of *exploratory data analysis*
- Understanding of the *language S-Plus*
- Understanding of *belief functions*
- Understanding of *Bayesian analysis and networks*
- Understanding of the basics of *multivariate methods*

Catalog Description:

Extensive coverage of sampling techniques for decision making. Includes simple random sampling, stratified sampling, cluster sampling, treating unequal clusters, area sampling, imperfect frames, questionnaire design, and field operations.

Prerequisite: *Msi 220 or Mat 108 or equivalent.*

A Tentative Honest Description:

Exploratory descriptive data analysis using the language S-Plus. Basic graphics commands in S-Plus including trellis graphics. Descriptive data exploration and statistical modeling. Data preprocessing for Datamining. Classification: Decision trees, Association Rules in Large Databases. Multivariate Methods: Clustering & other multivariate statistical methods. Anomaly detection.

Textbooks

The Basics of S-Plus, by Andreas Krause, Melvin Olsen (Springer Verlag, 2005)

Introduction to Datamining, by P. Tan, M. Steinbach, V. Kumar (Addison Wesley, 2006)

I shall also be distributing reading materials as the semester progresses.

Requirements

The classes will consist of lectures, solution of problems, discussion of papers, and programming exercises. I shall be dividing the class into groups of 2 each, balanced in terms of skills in accounting, programming, facility with computers, mathematical maturity, needs of the projects selected, and other such attributes. The groups will work through out the semester on all assignments, each group member taking turns to be the lead on the assignments.

Grading

The final course grade is dependent on the following factors:

100 points: Test (In class open book/notes. Details will be announced in the class)

75 points: Class Group presentations and written briefs

100 Points: Group Project & written report

25 points: Class participation

300 points: Total points (max)

The final course grade is strictly relative, based on the total points scored.

The grades, once assigned cannot be changed except in case of errors in grading. Under no circumstances is it possible to do extra credit work to improve the grade.

About the Instructor:

Jagdish S. Gangolly is currently an Associate Professor of Accounting and of Management Science & Information Systems, in the School of Business, and the Director of the Ph. D Program in Information Science at the Faculty of Informatics, College of Computing & Information at the State University of New York at Albany. He is also an affiliate and advisor at the Institute for Informatics, Logic & Security Studies at SUNY Albany. He was the Interim Director of the New York State Center for Information Forensics & Assurance (CIFA) during 2003-5. CIFA is a partnership between the University, New York State Police, and the New York State Office of Cyber Security and Critical Infrastructure Coordination (NY-CSCIC).

He holds a Bachelor's degree with a major in Mathematical Statistics, a master's degree with a major in Operations Research, and a Ph. D degree in Business Administration (Accounting). He is also a Certified Internal Auditor. He has previously taught at the University of Pittsburgh, University of Kansas, Claremont McKenna College & the Claremont Graduate School, and California State University at Fullerton. He has worked in senior executive positions in management services in the pulp & paper industry as well as in soft drink franchising in India.

His papers have appeared in Journal of Accounting Research, ACM Journal of Educational Resources in Computing, International Journal of Information Management, Auditing: Journal of Practice & Theory, Journal of the Operational Research Society, Critical Perspectives on Accounting, Expert Systems with Applications: An International Journal, Artificial Intelligence in Accounting & Auditing, International Journal of Digital Accounting Research, and the New Review of Applied Expert Systems & Emerging Technologies.

In 1989, he was the guest editor of Advances in Accounting; and he currently he serves on the editorial boards of the American Accounting Association's Journal of Emerging Technologies in Accounting, and the journal International Journal of Digital Accounting Research. He is also an Associate editor of the e-Services Journal. He serves on the Technical Committee Working Group 11.1 on Information Security Management, of the International Federation for Information Processing (IFIP).

His current research activities are primarily in the areas of conceptual information organisation, and the formal specification of control in accounting information systems. He also has collateral research interest in the relationships between Accounting and Legal Philosophy.

Department of Accounting & Law
State University of New York at Albany

Acc 522. Statistical Methods for Business Decisions (Fall, 2005)
J Gangolly

Tentative Schedule

September 7, 2006

Theme: *Introduction to Descriptive Statistics, Datamining, and S-Plus*

Readings: TSK: Ch.1, 2, 3.

September 14, 2006

Theme: *Data & Graphics in S-Plus I*

Topics: *Matrices, Frames, Arrays, and their manipulation; Basic graphics commands in S-Plus; Trellis graphics*

Readings: *S-Plus Manual*, TSK: Ch.1, 2, 3. KO: Ch 3 – 6.

September 21, 2006

Theme: *Data & Graphics in S-Plus II*

Topics: *Matrices, Frames, Arrays, and their manipulation; Basic graphics commands in S-Plus; Trellis graphics*

Readings: KO: Ch 3 – 6.

September 28, 2006

Theme: *Descriptive Data Exploration & Statistical Modeling*

Topics: *Use of S-Plus Graphics; Regression.*

Readings: KO: Ch 3 – 6.

October 5, 2006

Theme: *Classification: Decision Trees and Model Evaluation*

Topics: *General approach to solving classification problems, Algorithm for decision tree induction, Model overfitting and evaluating performance.*

Readings: TSK: Ch 4.

October 12, 2006

Theme: FLOAT

Readings:

October 19, 2006

Theme: Association Rule Mining I

Topics: *Frequent itemset generation, Rule generation, Compact representation of frequent itemsets, Evaluation of association patterns*

Readings: TSK: Ch 6.

October 26, 2006

Theme: Association Rule Mining II

Topics: *Frequent itemset generation, Rule generation, Compact representation of frequent itemsets, Evaluation of association patterns*

Readings: TSK: Ch 6.

November 2, 2006

Theme: *Classification: Clustering I*

Topic: *Hierarchical & Partitioning methods for Clustering, and their use in S-Plus.*

Readings: TSK: Ch. 8

November 9, 2006

Theme: *Classification: Clustering II*

Topic: *Hierarchical & Partitioning methods for Clustering, and their use in S-Plus.*

Readings: TSK: Ch. 8

November 16, 2006

TEST

November 23, 2006

No Class (Thanksgiving)

November 30, 2006

Theme: *Multivariate Methods III: Other Methods*

Topics: *Multi-Dimensional Scaling, etc.*

Readings: *S-Plus online manuals.*

December 7, 2006

Class Presentations of projects

Updated on September 8, 2005 by Jagdish S. Gangolly (j.gangolly@albany.edu)