SYS 4021: Linear Statistical Models

Instructors:

Laura E. Barnes, Ph.D., Room 101B Olsson Hall, 924-1723, lbarnes@virginia.edu

Course Description:

What are the contributing factors to the severity train accidents? How do you predict if an e-mail is spam? How can you translate goal-directed problems such as these into actionable decisions and meaningful recommendations that can have vast societal implications? How can you harness multi-dimensional, heterogeneous data to analyze the problem? In this course, we will explore Evidence Informed Systems Engineering (EISE) practices and how they can be applied to difficult, open-ended problems.

The primary tools for EISE come from linear statistical models and the course demonstrates the use of these models for problem understanding, prediction, and control. Specific topics include multiple regression, principal components, analysis of covariance, logistic regression, time series analysis, bootstrapping, and response surfaces. In class, we will concentrate on the theory and practice of model construction, while the projects provide open-ended problem solving situations that illustrate the broad applicability of the methods.

Course Objectives:

- Apply evidence-informed systems engineering approaches to solve real-world problems.
- Assess the limitations of the information available to solve a problem.
- Identify the appropriate statistical modeling technique(s) appropriate for a specified problem.
- Derive an actionable recommendation with statistical confidence using the evidential reasoning process.
- Formulate meaningful, testable hypotheses from a real-world problem and associated data.
- Uncover bias, error, outliers, and influential observations in data using multi-dimensional visualization techniques.
- Communicate the application of the evidence-informed systems engineering process to a problem through a technical report directed to a client and / or practicing engineer.

Prerequisites:

SYS 3060, SYS 3034, APMA 3012 and major in systems engineering.

Required Software:

In this course, we will use R for data analysis. You can use the basic R (RB) or R-Studio (RS) interface. You may request permission to use other software, such as, S+ and SAS but the resulting models, analysis, and graphics in your reports must be comparable to the output we could obtain using R.

- (JN) Jupyter Notebook– `http://jupyter.org/`
- (RS) R– `https://www.rstudio.com/`
- (RB) R Project– `http://www.r-project.org/`

Course Communication Platform:

In this course, we will use Piazza (P) outside of class for course communications. Students should utilize this resource instead of e-mailing professors or TAs directly questions. Questions should be posted on the discussion board referencing the appropriate assignment. Questions will be responded to by a TA or instructors within 48 hours.

- (P) Piazza– `https://piazza.com/virginia/fall2016/16fsys4021/`

Recommended Texts:

- (MB) Maindonald, John and W. John Braun, **Data Analysis and Graphics Using R: An Example-Based Approach**, 3rd Ed., Cambridge Univ. Press: Cambridge, UK, 2010.
- (F1) Faraway, Julian, **Linear Models with R**, Chapman & Hall/CRC Texts:Boca Raton, FL, 2010.
- (V) Verzani, John, **simpleR**– *Using R for Introductory Statistics*,Chapman & Hall/CRC 2004, `http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf`.

Supplemental Readings:

- (F2) Faraway, J.J., *Practical Regression and ANOVA using R*, July, 2002, `http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf`.

Learning Resources:

- (WC) University of Virginia Writing Center– `https://virginia.mywconline.com/`.
- (CC) Coursera Course on R Programming–`https://www.coursera.org/course/rprog`
- (RB) R News & Tutorials–`http://www.r-bloggers.com/`

Nominal Class Outline:

Lectures	Topics	Projects	Readings
1	EISE	Train	MB 2
2-3	Visualization	Train	MB 1 & 4.5
4-5	Principal Components	Trains	MB 13.1, 13.2 & F1 9.1, F2 9.2
5-9	Multiple Regression	Trains	MB 6, 7.1 & F1 2-5, F2 1-3, 10
10-13	Generalized Linear Models	Spam	MB 8.1-8.3
14-17	Time Series	Spam	MB 9
18-21	Bootstrapping & Simulation	Health	MB 3.3
22-25	Advanced Topics	Health	MB 7.2-7.6
26	Summary		Review

Student Evaluation and Assessment

Evaluation

Grade Breakdown:

Midterm Exam	15%
Final Exam	25%
Weekly Labs	15%
Class Work	10%
Project Reports	30%
Class Participation	+

Learning Assessment Activities

Exams (40%):

Exams are based entirely on classroom notes and discussions, readings, projects, and laboratory assignments. Each exam will contain a closed book section with short answer questions and an open book section requiring analytical problem solving. The final examination is cumulative. Example questions for both exams will be provided before the exams.

Weekly Labs (15%):

Each week laboratory notes will be posted on the course Collab site outlining R knowledge we cover in class. On Sunday night, weekly laboratory assignments based upon course and laboratory notes will be posted via the *Tests & Quizzes* feature on the course Collab site. These assignments provide exercises in R programming that supplement the material covered in class and provide the foundation for the project reports. Each lab assignment requires students to program in R and analyze a supplied data set. The assignments are designed assess your knowledge on statistical modeling techniques and their mechanics. These assignments must be done individually, include an honor pledge, and be completed by Friday night at midnight. While there is no time limit for these assignments, they are designed not to take more than 50 minutes. Laboratory sessions are excellent practice for exams and real-world analysis under time constraints. At the end of the semester, the lowest grade on the lab assignments will be dropped. Additionally, if you are at the top 10% of the dropped lab assignments, you will get extra credit.

Class Work (10%):

Class work will consist of pop quizzes, writing responses, and group R work based on course material (readings, videos, lecture notes).

Projects (30%):

The class has three project exercises on various real-world topics and data sets. These projects provide a real-world context for what we learn and

are open-ended problem solving experiences that illustrate the concepts of evidence-informed systems engineering. Hence, they provide the opportunity to demonstrate understanding of class material in with real data to solve a goal-directed problem. Projects are designed to teach students how to perform a detailed analysis as well as proficiently communicate your results as you would in a technical publication or client report.

Project Reports:

Each project exercise requires you to complete a technical report with your assigned group. The instructor will form project teams using data from the CATME platform. Your group must strictly adhere to the format requirements for the report. Be sure to reference all material used to complete the report and to include the report template and project assignment. Additionally, students will complete a peer review of their teammates upon submission of each project. An example of the *reference style* is shown below.

[1] L.E. Barnes, Project Assignment 1: Train Accidents, Fall 2015, Charlottesville, VA, 2015.

[2] J.J Faraway. (2002, July). *Practical Regression and ANOVA using R*, [Online]. Available: http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf.

[3] I.M. Smart, M.E. Too, and E. VonSmarter, Personal conversation regarding preparation of box plots for project 1 on train accidents, August 2015.

[4] Federal Railroad Administration, Report to Congress: Positive Train Control: Implementation Status, Issues, and Impacts, August 2012.

[5] www.sys.virginia.edu 2016, Systems and information engineering web page, last file access on September 2, 2015.

Project Report Submission:

On the day a project report is due each group member must submit an electronic copy **in pdf (NOT doc or docx, etc.)** on the class web-site and pledge the submission.

Class Participation:

While class participation is not a part of the core evaluation structure, you can earn up to 2% on your final grade for active class participation.

Course Policies

Honor Policy:

For exams and laboratory quizzes you may only consult the instructors and teaching assistants. You may not work with others on these assignments.

For project reports, you will be assigned a group of 3 or 4 team members to collaborate with on analysis and prepare the report. You are also encouraged to consult others in the class. However, the work you submit must be your own group's product. You may not copy the work of anyone else. Cited material must be properly referenced in your reports. If you have questions, consult the course instructors or teaching assistants.

If you cheat, copy from someone else on an exam or quiz, or copy someone else's work on a project that you do not cite then you will fail the course.

Every report, quiz, and exam must include acknowledgment that you abided by the honor code. Failure to include this acknowledgment will mean that your submission will not be graded.

Late Work Policy:

The due dates and times for all work will be announced and posted on the class web-page. Only the instructor can grant exceptions to the due dates and this will normally be done in response to requests made prior to the due date or for true emergencies. Assignments submitted late will be penalized if request is not made in advance. In many cases you will do better to submit an incomplete assignment rather than a late one.

Learning Accommodations:

All students with special needs requiring accommodations should present the appropriate paperwork from the Student Disability Access Center (SDAC). It is the student's responsibility to present this paperwork in a timely fashion and follow up with the instructor about the accommodations being offered. Accommodations for test-taking(e.g., extended time) should be arranged at least **5 days** before an exam.

The SDAC is located in the Department of Student Health and can be contacted at 243-5180/5181.