Stat 6021: Linear Models for Data Science
Syllabus and Course Policies
Summer 2020

1 Overview

Instructor Name and Contact Information:
Jeffrey Woo, PhD
yjw4b@virginia.edu
434 924 7765
Office hours: Tuesdays and Thursdays: 9 to 10am for Blue group, 10 to 11am for Orange group; and discussion forum

Subject Area and Catalog Number: Stat 6021

Year and Term: Summer 2020

Class Title: Linear Models for Data Science

Level (Graduate or Undergraduate): Graduate

Credit Type: Graded

Class Description: An introduction to linear statistical models in the context of data science. Topics include simple and multiple linear regression, generalized linear models, time series, analysis of covariance, tree-based classification, and principal components. The primary software is R.

Required Text: Introduction to Linear Regression Analysis, Fifth Edition. Montgomery, Peck, and Vining. ISBN: 978-0-470-54281-1

Learning Outcomes: Upon successful completion of this course, you will be able to:

1. Statistical Methods
   (a) Use regression analysis to answer questions of interest in a wide variety of application environments.
   (b) Determine the most appropriate regression model for a given data set.
   (c) Identify the assumptions and limitations of a given regression model.
   (d) Diagnose and remedy common problems with the regression model found in real data.
2. Computing
   (a) Use the statistical software R for regression analysis.
3. Communication and Professionalism
   (a) State appropriate context-specific conclusions from an analysis.
   (b) Present and discuss orally and in writing, statistical ideas, methods, and results to lay and professional audiences.
   (c) Work in teams to demonstrate the skills of a professional statistician in organizing and managing projects.

4. Statistical Theory and Mathematics

   (a) Describe the mathematical framework of regression models.

   (b) Describe the importance of assessing the assumptions and limitations for a given regression model.

Assessment Components:

Assignment	Percent of Grade	Due
Guided Question Set Participation	10	Modules 4, 7, and 11
Homework Question Sets	30	Modules 3, 6, 10
Project 1	25	Module 7
Project 2	35	Module 12

Delivery Mode Expectations: This course is delivered in 12 learning modules through Collab. Each course module consists of an online lesson and a live lesson.

- Meetings (video and teleconference) for live lessons will be held on Mondays, Wednesdays, and Fridays via Zoom. Students are expected to actively participate in the live class discussions by asking questions and contributing to the discussion.

  Prior to the live meetings, students are expected to complete the textbook readings, additional notes, view the videos, R tutorial, take the ungraded learning check quiz, and start working on the guided question set. The full meeting schedule can be found in the Live Session Schedule section below.

- Prior to the live meetings, students are expected to complete the textbook readings, additional notes, R tutorial, and take the ungraded learning check quiz. The full meeting schedule can be found in the Live Session Schedule section below.

- Each live lesson will begin with a Q&A period in which students will have an opportunity to ask questions about the topics explored in the online lesson.

- In live meetings, students will complete the guided question set in randomly assigned groups. This small-group discussion will be followed by a large-group discussion of the guided question set. Students are expected to come to class prepared to participate in each week's live activity.

- After live meetings, students are expected to work on the homework questions. Students may work together on these, but each student's answers must be in their own words.

Required Technical Resources and Technical Components: The statistical software R will be used in this course and is available for free download.

Required Technical Resources and Technical Components: The statistical software R will be used in this course and is available for free download.

2 Class Specific Information

Class Instruction and Activities:
This course introduces commonly used linear regression models and includes detailed implementation of the models within real data examples using the R statistical software. The primary goal of the course is to develop an applied and intuitive (as opposed to purely theoretical or mathematical) understanding of the topics by focusing on real-life applications of statistical theory. Consequently, the learning methods include many hands-on exercises using real data sets. By the end of the course, students will be able to appropriately select from and apply regression models to examine relationships between multiple variables and interpret the results derived from such implementations. This course will give students the tools to use the techniques when faced with the need in practical research.

Evaluation Standards and Assessments

- **Guided Question Set Participation (10%)** During the live lesson, students will work in randomly assigned groups to discuss the guided question set. Students are expected to come to the live lesson on time and prepared to fully participate in the discussion. At the end of module 4, 7, and 11, each student will be graded on his or her participation in the small-group discussion of the guided question set. This grade will be based on a self- and peer-evaluation of each student's contribution to the group discussions.
- **Homework Question Sets (30%)** There will be three homework sets. The homework will be composed of applied problems. Homework will not be accepted late. Students are encouraged to collaborate on the homework assignments, but each student must prepare their solutions individually. Working together is a great way to learn the material but copying others' work and submitting it as your own is in direct violation of the Honor Code and will be treated as such.
- **Project 1 (25%)** Project 1 is due July 30. In this group project, students will be given a data set with a number of predictor variables and a response variable. Using R, they are to determine the "best" regression model and in the process to diagnose and remedy problems. Each group will submit a written report to present their analyses and answer questions of interest contextually.
- **Project 2 (35%)** Project 2 is due Aug 7. In this group project, students will be given a data set with a number of predictor variables and a response variable. Using R, they are to determine the "best" regression model and in the process to diagnose and remedy problems. Each group will submit a written report and record a slide presentation to present their analyses and answer questions of interest contextually.

Course Topics:

Module	Topic	Assignments
1	Introduction to Simple Linear Regression (SLR)	Guided question set 1, homework set 1
2	Inference with SLR	Guided question set 2, homework set 1
3	Model Diagnostics and Remedial Measures in SLR	Guided question set 3, homework set 1
4	Introduction to Multiple Linear Regression (MLR)	Guided question set 4, homework set 2
5	Sums of Squares and Multicollinearity	Guided question set 5, homework set 2
6	Categorical Predictors	Guided question set 6, homework set 2, Project 1
7	Model Selection	Guided question set 7, homework set 3
8	Model Diagnostics and Remedial Measures in MLR	Guided question set 8, homework set 3
9	Logistic Regression Part One	Guided question set 9, homework set 3
10	Logistic Regression Part Two	Guided question set 10, homework set 3, Project 2
11	Time Series Regression and Resampling Methods	Guided question set 11
12	Shrinkage Methods	

Live Session Schedule: Live sessions will be held on Monday, Wednesday, and Friday mornings. Below is a list of specific dates.

Date	Time
7/13/2020	9:00-9:50 am EDT (Blue Group)
	10-10:50 am EDT (Orange Group)
7/15/2020	9:00-9:50 am EDT (Blue Group)
	10-10:50 am EDT (Orange Group)
7/17/2020	9:00-9:50 am EDT (Blue Group)
	10-10:50 am EDT (Orange Group)
7/20/2020	9:00-9:50 am EDT (Blue Group)
	10-10:50 am EDT (Orange Group)
7/22/2020	9:00-9:50 am EDT (Blue Group)
	10-10:50 am EDT (Orange Group)
7/24/2020	9:00-9:50 am EDT (Blue Group)
	10-10:50 am EDT (Orange Group)
7/27/2020	9:00-9:50 am EDT (Blue Group)
	10-10:50 am EDT (Orange Group)
7/29/2020	9:00-9:50 am EDT (Blue Group)
	10-10:50 am EDT (Orange Group)
7/31/2020	9:00-9:50 am EDT (Blue Group)
	10-10:50 am EDT (Orange Group)
8/03/2020	9:00-9:50 am EDT (Blue Group)
	10-10:50 am EDT (Orange Group)
8/05/2020	9:00-9:50 am EDT (Blue Group)
	10-10:50 am EDT (Orange Group)
8/07/2020	9:00-9:50 am EDT (Blue Group)
	10-10:50 am EDT (Orange Group)

Communication & Student Response Time: Discussion forums are set up so students are encouraged to pose questions on the discussion forums. Classmates are encouraged to help each other.

Emails can be sent to the instructor. Be sure to have our class name and number labeled clearly in the subject heading and allow for 1-2 business days for a response.

Electronic Submission of Assignments: All assignments must be submitted electronically through Collab by the specified due dates and times. It is crucial to complete all assigned work - failure to do so will likely result in failing the class.

3 Policies

SDS Grading Policies: The standing of a graduate student in each course is indicated by one of the following grades: A+, A, A-; B+, B, B-; C+, C, C-; D+, D, D-; F. B- is the lowest satisfactory grade for graduate credit.

Attendance: Students are expected to attend all class sessions. Instructors establish attendance and participation requirements for each of their courses. Class requirements, regardless of delivery mode, are not waived due to a student's absence from class. Instructors will require students to make up any missed coursework and may deny credit to any student whose absences are excessive. Instructors must keep an attendance record for each student enrolled in the course to document attendance and participation in the class.

University Email Policies: Students are expected to check their official UVa email addresses on a frequent and consistent basis to remain informed of University communications, as certain communications may be time sensitive. Students who fail to check their email on a regular basis are responsible for any resulting consequences.

Mid-Term and End-of-Class Evaluations: Students may be expected to participate in an online mid-term evaluation. Students are expected to complete the online end-of-class evaluation. As the semester comes to a close, students will receive an email with instructions for completing this. Student feedback will be very valuable to the school, the instructor, and future students. We ask that all students please complete these evaluations in a timely manner. Please be assured that the information you submit online will be anonymous and kept confidential.

Academic Integrity: The School of Data Science relies upon and cherishes its community of trust. We firmly endorse, uphold, and embrace the University's Honor principle that students will not lie, cheat, or steal, nor shall they tolerate those who do. We recognize that even one honor infraction can destroy an exemplary reputation that has taken years to build. Acting in a manner consistent with the principles of honor will benefit every member of the community both while enrolled in the School of Data Science and in the future. Students are expected to be familiar with the university honor code, including the section on academic fraud `https://honor.virginia.edu/academic-fraud`.

Each assignment will describe allowed collaborations, and deviations from these will be considered Honor violations. If you have questions on what is allowable, ask! Unless otherwise noted, exams and individual assignments will be considered pledged that you have neither given nor received help. (Among other things, this means that you are not allowed to describe problems on an exam to a student who has not taken it yet. You are not allowed to show exam papers to another student or view another student's exam papers while working on an exam.) Sending, receiving or otherwise copying electronic files that are part of course assignments are not allowed collaborations (except for those explicitly allowed in assignment instructions). Assignments or exams where honor infractions or prohibited collaborations occur will receive a zero grade for that entire assignment or exam. Such infractions will also be submitted to the Honor Committee if that is appropriate. Students who have had prohibited collaborations may not be allowed to work with partners on remaining homework assignments.

Special Needs: It is my goal to create a learning experience that is as accessible as possible. If you anticipate any issues related to the format, materials, or requirements of this course, please meet with me outside of class so we can explore potential options. Students with disabilities may also wish to work with the Student Disability Access Center to discuss a range of options to removing barriers in this course, including official accommodations. Please visit their website for information on this process and to apply for services online: `https://www.studenthealth.virginia.edu/sdac`.

If you have already been approved for accommodations through SDAC, please send me your accommodation letter and meet with me so we can develop an implementation plan together.

4 Technical Support

Technical Specifications: Computer Hardware

- Operating system: Microsoft Windows 8.1 (64-bit) or Mac OS X 10.10
- Minimum hard drive free space: 100GB, SSD recommended
- Minimum processor speed: Intel 4th Gen Core i5 or faster
- Minimum RAM: 8GB

Technical Support Contacts

- Login/Password: `https://in.virginia.edu/helpdesk`
- UVaCollab: `collab-support@virginia.edu`