STAT 330: Introduction to Regression

Winter 2019
Sec 002 Lecture MWF 11:00–11:50 in 2104 JKB
Sec 003 Lecture MWF 12:00–12:50 in 3106 JKB

Class Prep Lab 8–9
Open Lab MWF 1–2 in 244 TMCB

Dr. Scott D. Grimshaw
Office: 219 TMCB
Phone: 801–422–6251
Email: grimshaw@byu.edu
http://grimshawville.byu.edu/stat330.html
Office Hours: MWF 1–2

Course Catalog Description:

Regression, transformations, residuals, indicator variables, variable selection, logistic regression, time series, observational studies, statistical software.

Prerequisites:

Stat 230 and Math 112.

Learning Outcomes:

A student completing Stat 330 will be able to:

1. Fit a regression model with professional statistical software
2. Apply appropriate transformations to the response variable to improve agreement with regression assumptions
3. Use residuals and influence diagnostics to assess model fit, agreement with regression assumptions, and identify outliers and influential observations
4. Create sets of indicator variables for categorical explanatory variables
5. Apply stepwise selection to identify a subset regression model that selects the most significant explanatory variables from a large data set
6. Fit a logistic regression model with professional statistical software

Course Grade:

10%	Class Participation: Preparation (Theory Fridays not Included)
10%	Class Participation: Attendance (Theory Fridays not Included)
15%	Class Participation: Code
5%	Analysis Review (Strengths/Weaknesses) and Challenge
5%	Methods Homework Assignments (Groups Strongly Encouraged)
10%	Theory Homework Assignments (Individual Work, No 'Cheats')
5%	Project Non-Technical Report (2–3 pages)
10%	Midterm Exam 1
10%	Midterm Exam 2
10%	Midterm Exam 3
10%	Theory Exam (time scheduled by university for final)

	A = 94-100	A-= 90-93.9
B+= 85-89.9	B= 82-84.9	B-= 80-81.9
C+= 75-79.9	C= 70-74.9	C-= 65-69.9
D+= 62-64.9	D= 59-61.9	D-= 55-58.9
E= 0-54.9		

Grades may be lowered by the instructor if you fail to meet University standards.

Recommended Text:

Sheather (2009), *A Modern Approach to Regression with R.*
    The ebook is available through the BYU Library.

The author's website is `http://gattonweb.uky.edu/sheather/book/` and contains datasets and code from examples.

For Theory Fridays: *Intro to Linear Models* on class website

Authentic Data Experiences:

There has never been a better time to teach statistics since instructors can choose from an exploding set of real data applications in their courses. The common practice in books and data repositories to present a curated example where the students interact with a clean, well-organized, easily read file. Usually these are 'real data' according to the GAISE definition but there is a missing dimension of 'ossification' within that definition. Often the data acquisition and manipulation tasks aren't shared with the students. For example, one of the strengths of Sheather (2009) as a textbook is the many authentic data experiences which required significant data acquisition and manipulation work by the author, but all the datasets are presented to students as clean well organized files easily read from the book website. These 'static' datasets deprive students of the opportunity to learn and practice data management skills.

Authentic data experiences provide source data locations with sufficiently detailed instructions for the students to prepare the dataset for analysis. This is particularly important for applications that are time sensitive, where students analyzing data that was current at book publication will seem artificial and obtaining current data for the same problem requires basic data-related skills.

Well-written authentic data experiences identify applications with some, but not significant data acquisition and manipulation tasks, and providing clear instructions of what tasks are required by students to prepare the dataset for the analysis questions on the assignment. It is important to remember that practicing data-related skills is secondary and so the tasks are more modest in scope than would appear in a senior-level culminating experience. A rule of thumb was for students to spend 15-20 minutes to prepare the dataset for analysis. Code may occasionally be provided for parts the students may not have seen since there are no data course prerequisites or when the application required performing tasks that could take students longer than 15-20 minutes.

It is valuable for students to have many authentic data experiences. They provide an opportunity to demonstrate connections between data skills and statistical skills. The result is more practice for undergraduate statisticians that leads to more confidence as they prepare for careers working with data.

Class Participation:

Lectures in Stat 330 are a professional environment mimicking what I've experienced working with data to build models for business and scientific research questions. When we meet in class and take exams we try to do what makes sense in this universe. For lectures, come prepared as if you were at a business meeting where you are an important collaborator to getting the work done.

'Class Participation: Preparation' points are based on evidence the dataset was created and EDA performed (2 points). This file needs to be submitted through Learning Suite before the start of class, regardless of your intention to attend class. If you are not prepared by 8am on class day plan to attend the Class Prep Lab. TAs will randomly select (without replacement) 5 files to grade for each assignment (all other submitted files will receive 2 points).

'Class Participation: Attendance' expects attend class and participating in the analysis psuedo-code creation and communication. There are no excused absences, but five days of missing class will be dropped.

'Class Participation: Code' is due the day the analysis is completed in class and graded based on correctness, reproducibility, and comments. In case of an emergency that will impact your ability to meet the deadline, students must contact the instructor or TA as soon as possible with an alternative deadline.

Analysis Review (Strengths/Weaknesses) and Challenge:

Choosing the correct statistical model and analysis is one of the most difficult experiences for young statisticians. To prepare you for this and provide some creativity to reflect your interests, after completing an a class participation assignment you will summarize the statistical objectives of the research task and data characteristics and comment on the analysis strengths and weaknesses. You will also provide a research task and find data with the same characteristics as the assignment.

Methods Homework:

Produce **your own** assignment to turn in for grading, but working in groups and with the TA are strongly recommended. To receive full credit, answer all assigned questions completely, accurately, with important code (if applicable).

Late homework should be avoided. In case of an emergency that will impact your ability to meet homework deadlines, students must contact the instructor or TA as soon as possible with an alternative deadline.

Theory Homework:

To receive full credit, answer all assigned questions completely, accurately, and showing all work. These problems should be done independently without searching for solutions or using mathematical/computational 'cheats' or asking others or asking the TA. You should anticipate grades less than '100%' on these problems. Please complete all Methods Homework before spending time on these problems.

Late homework should be avoided because of the negative impact on your grade. In case of an emergency that will impact your ability to meet homework deadlines, students must contact the instructor or TA as soon as possible with an alternative deadline.

Midterm Exams:

You will be assigned to a group of three for each exam. The expectation is that group work will be **collaborative**. All group members will receive the same grade.

A minimum level of performance on the material for the midterm is required to be assigned to a group. You will be notified at the group availability announcement if you are underperforming.

Those who take the exam as individuals (no group) will be penalized 50%.

Theory Exam:

The Theory Exam will follow the University Final Exam Schedule and be given in the classroom.

Professional Statistics Software:

This course will use R exclusively, but this is a class on statistical modeling and not an R class. Everything we cover can be done in SAS (possibly offered as a Stat 381 section next Fall), and selections of what we cover can be done in Python and SPSS and Stata and other software.

R is free and can be downloaded by following the instructions at `http://cran.r-project.org/` See the links under "Download and Install R" to choose the correct version for your operating system. Several summaries of R commands are available online (see e.g., `http://www.personality-project.org/r/r.commands.html`). Some students prefer the RStudio overlay on R (installed after R is installed) which unites different features in a single console.

Be thoughtful in organizing your work. Each class participation authentic data experience should be it's own file with an appropriate name. Don't create a junky RStudio console. If you use R Markdown the code you submit must run without user intervention and contain expected level of comments.

TAs:

TA office hours will not be used to redo a lecture. They will be used to answer specific questions about the class participation assignments, methods homework, and project report. This implies that students will have worked on the assignment before coming into office hours. If you miss class, please consult with other students in the class or a tutor to teach you the analysis that was discussed during your absence.

Project Non-Technical Report

In their role interpreting data across disciplines, statisticians must be excellent communicators who are comfortable interacting with a wide range of statistical consumers. The intended audience for this report is non-statisticians. Think about how much detail the audience wants, try to avoid jargon, and be careful falling into 'process writing.'

The project will be assigned from the collection of authentic data experiences. The top 5 papers in the section will receive an A grade. All others are expected to be at least an A-level, and papers below that will have required revision.

Good 'data journalism' writers to read and model for writing about technical material for a smart non-technical audience: Carl Bialik, Nate Silver, Walter Hickey, Mona Chalabi, Rachel Bachman, Jo Craven McGinty, Jim Pagels, Regina Nuzzo, Andrew Gelman, David Spiegelhalter, David Salsburg, Jim Albert, Paul Sabin — and if there is someone you like to read please share!

Students With Disabilities

BYU is committed to providing reasonable accommodation to qualified persons with disabilities. If you have any disability that may adversely affect your success in this course, please contact the University Accessibility Center (UAC) at 801-422-2767. Services deemed appropriate will be coordinated with the student and instructor by that office.

Mental Health Services

Barriers to learning are created by stress, anxiety, family and relationship concerns, and personal crises. If stressful life events or mental health concerns are inhibiting your ability to participate in daily activities or leading to diminished academic performance, please contact the BYU Counseling and Psychological Services (CAPS; 1500 WSC, 801-422-3035, `caps.byu.edu`). CAPS provides individual, couples and group counseling to students. These services are confidential and are provided by the university at no added cost to you. Professional psychologists and counselors who specialize in helping college students are available 24-hours a day to assist students in crisis; if you have an emergency during non-business hours (5pm- 8am), please contact BYU Police Dispatch (801-422-2222) who will put you in touch with a counselor.

Honor Code, Dress, and Grooming Standards

The Honor Code helps make BYU a unique university where the spirit of the Lord can quicken learning, and where we can study, work, and live among people who share our values. I personally support all aspects of the Honor Code, and expect that its principles, particularly the tenets of Academic Honesty and Dress and Grooming, will be maintained in this course. In short, do not cheat or plagiarize; and be modest and appropriate in your manner of dress, and civil in your behavior.

Preventing & Responding to Sexual Misconduct

In accordance with Title IX of the Education Amendments of 1972, Brigham Young University prohibits unlawful sex discrimination against any participant in its education programs or activities. The university also prohibits sexual harassment—including sexual violence—committed by or against students, university employees, and visitors to campus. As outlined in university policy, sexual harassment, dating violence, domestic violence, sexual assault, and stalking are considered forms of "Sexual Misconduct" prohibited by the university.

University policy requires all university employees in a teaching, managerial, or supervisory role to report all incidents of Sexual Misconduct that come to their attention in any way, including but not limited to face-to-face conversations, a written class assignment or paper, class discussion, email, text, or social media post. Incidents of Sexual Misconduct should be reported to the Title IX Coordinator at `t9coordinator@byu.edu` or (801) 422-8692. Reports may also be submitted through EthicsPoint at `https://titleix.byu.edu/report` or 1-888-238-1062 (24-hours a day).

BYU offers confidential resources for those affected by Sexual Misconduct, including the university's Victim Advocate, as well as a number of non-confidential resources and services that may be helpful. Additional information about Title IX, the university's Sexual Misconduct Policy, reporting requirements, and resources can be found at `http://titleix.byu.edu` or by contacting the university's Title IX Coordinator.