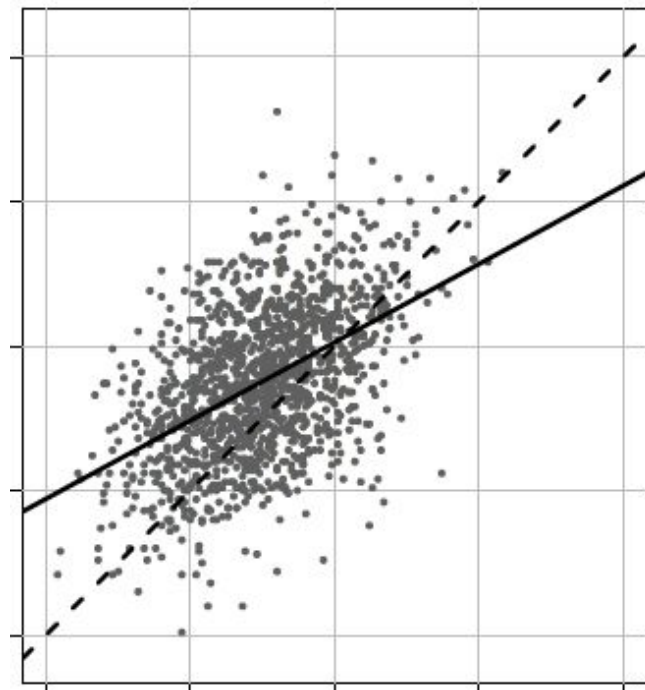STAT 3032 Section 001 Syllabus, Spring 2020

Regression and Correlated Data

Instructor: Austin Brown

TABLE OF CONTENTS:

Credit: The image was taken from the cover of Applied Linear Regression. [1]

Course Description :

What is Regression? Returning to the former (and often less developed) state? Nay, in statistics it takes on other meanings. Regression deals with the relationship between statistical quantities. It is a powerful tool that can help us extract information from the data. For example, we can study how much a professor's overall teaching quality is associated with the easiness of the course materials using the rating data from RateMyProfessors.com. Some regression analyses will reinforce our preconceptions about how things should be, while others challenge our common sense. In this class, we will study regression models that represent the mechanism

---

[1] Weisberg S. (2014). *Applied Linear Regression (4th Ed.)*. Hoboken, NJ: John Wiley & Sons

behind the cluttered raw data and constantly evaluate how our choices as analysts impact the validity of the conclusions.

Course Goals :

By the end of this course, you will be able to
- Fit the appropriate regression models to datasets and check the model assumptions.
- Make statistical inferences based on the regression models;
- Use the programming language R for computing, analyzing, and visualizing data.

GENERAL INFORMATION

Instructor: **Austin Brown**

 Email: brow5079@umn.edu

 Office: Ford Hall 495

 Office hours: MW 12:20-1:20pm and F 1:10-2:10pm
 at Ford 495 or by appointment at Ford Hall 495

TA: **Yuchen Yao**

 Email: yao00056@umn.edu

 Office hours: MW 1:30pm-2:30pm and

 Th 4:10-5:10pm at Ford 495 or

 by appointment at Ford Hall 495

Grader: **Zexian Lu**

 Email: lu000276@umn

 Office hours: by appointment at

 Ford Hall 495

Meeting Times:

Lecture (001): MWF 11:15am–12:05pm at Anderson Hall 270

Lab (002): T 10:10am–11:00am at Folwell Hall 28

Lab (003): T 11:15am–12:05pm at Appleby Hall 303

Lab (004): T 12:20pm–01:10pm at Folwell Hall 5

Prerequisites :

STAT 3011 or STAT 3021. Please make sure that you are familiar with the following topics: random variable, expectation (mean), variance, correlation, central limit theorem, basic probability distributions (such as Normal and t distribution), hypothesis testing, and confidence interval.

Textbook (Optional):

Sheather, S. J. (2009). *A Modern Approach to Regression with R*. New York, NY: Springer.

-You can download a pdf of the textbook for free from the University of Minnesota Libraries with your UMN account.
- You can also order a soft-cover copy for $24.99, according to the agreement between the University of Minnesota Libraries and the publisher (Springer).

COMMUNICATION

Class-wide Announcements: Class-wide announcements are made in class or through the Announcement function in Canvas. Please make sure to set up the Canvas notification preference so that you receive it in your email right away or daily, when an class-wide announcement is made in Canvas.

Email: You can send me or the TA emails through your UMN email or the Canvas Conversation (Inbox) function. Our goal is to respond to your email within 48 hours of receiving it (weekends excluded).

Last-minute emails: Last-minute emails are those sent to us after 5pm on the due dates. Please understand I have advised the TA to wait until the next day to respond to these emails. The purpose of this policy is to encourage you to seek out help early and not depend on our last-minute help.

Words of Advice (1):

This is a 4-credit course, which means that you are expected to have 12 hours of course-related work. In other words, you need to attend the lectures and the lab, and work for about 8 hours outside of class.

Contrary to the somewhat popular belief that lab sessions are optional, they are not!
Important material is covered during them! Lab sessions are also a great place to get hints on more challenging homework problems. Sometimes lab sessions will cover material we do not have time to cover in class. At other times they'll be used as a way to practice and test concepts**.**

GRADING GUIDELINES

Final Grade Cutoff

Percentage Cutoff	Letter Grade	Percentage Cutoff	Letter Grade
93%	A	73%	C
90%	A-	70%	C-
87%	B+	67%	D+
83%	B	63%	D
80%	B-	60%	D-
77%	C+	< 60%	F

Grade Components

Component	Contribution to final grade
Preparation Quiz	8%
Homework	32%
Exam1	20%
Exam2	20%
Exam3	20%
Extra credit	About 1%

The letter grades may be adjusted to your favor at the end of the semester.

Incompletes: We give out an "Incomplete" only if the following criteria are met
-The student has a documented case of hardship that prevents the completion of the course.
-The student has, up until the point of the request, been completing the coursework and exams.
-The student's average grade in percentage at the point of the request is 70% or above.

Please talk to us if you are considering requesting an "Incomplete".

Grade Concerns: We try to maintain fair and consistent grading. If you see mistakes in grading, please bring them to us and we are happy to make corrections or adjustments.

Extra Credit for Your Thoughts: During the semester, I will send out an anonymous survey to collect feedback about your experience in this course. If more than 80% of the class fills it out by the due date, every one of you gets 1% added to your final grade. By completing the survey, you are helping me adjust the course materials to your needs, and raising the grades of you and your classmates.

PREPARATION QUIZ

This assignment includes questions about the course policies and the basic statistics this class requires. It motivates you to read the syllabus thoroughly and to review the materials in the course prerequisite. The quiz is administered in Canvas. You have 3 attempts and the highest

score will be recorded. You can find it by clicking on "Assignment" in the course navigation bar. ***No collaboration is allowed for this assignment***.

Resources for review basic statistics:
- Khan Academy's course *Statistics and Probabilities*
  https://www.khanacademy.org/math/statistics-probability
- LinkedIn Learning (the UMN supported tool): search for courses of "Statistics Foundations". https://it.umn.edu/technology/linkedin-learning

HOMEWORK

General: There will be 9 homework assignments in total. Problems will be assigned one week before it is due. The lowest homework grade will be dropped. This is to cover situations when you cannot turn your homework in on time such as when you are sick, your computer breaks, your lost internet connection, too many tests, etc. Please use your assignment drop wisely. *There are no exceptions for the second drop.*

Submission: Please submit your homework as a pdf file in Canvas. You may include screenshots of handwritten answers in the pdf file, so long as they are readable. The emphasis is on understanding and explaining concepts and results, rather than number crunching. Points will be awarded not only to correct solutions, but also for clear and concise explanation and communication. ***You are responsible for making sure that Canvas has received your submission. No late assignments are accepted.***

Collaboration: We encourage you to do homework and solve problems in groups. You will realize that you have fully understood a concept once you can explain it to your peers. Hearing a different voice than mine may also help you understand some concepts. However, all work turned in should be your own. In practice, this means you are encouraged to discuss the problems, but you should write up your own solution and write explanations in your own words.

EXAMS

Date & Time:

Exam 1 — Wed, Feb 26 at the regular lecture time.
Exam 2 — Wed, Apr 1 at the regular lecture time. (This is not a joke!)
Exam 3 — Mon, May 4 at the regular lecture time.

General: Exams are closed-book and closed-notes. ***No calculators are allowed.*** For each exam, you may take one (8.5 by 11 inches; both sides) paper as a cheat sheet. The cheat sheet needs to be *hand-written* and you may write down anything you would like to include.

Exam 2 and Exam 3 are not comprehensive, and will cover the material since the last exam. However, some concepts from the previous exam are foundational to what we learn next and may be mentioned or used in the later exams.

> **Words of Advice (2):** *DO NOT CHEAT. It's not worth it.*
>
> If we discover that a student has taken unfair advantage or misrepresented someone else's work as their own, he/she will be assigned a penalty appropriate for the level of offence, which could be receiving a zero for the assignment or failing the course. In some cases, we will need to refer the student responsible for academic dishonesty to the Office for Community Standards, formerly known as the Office for Student Conduct and Academic Integrity. Again, your integrity is worth so much more… If you have any questions regarding the expectations for a specific assignment or exam, ask.

Missed Exams: If you know in advance that you will miss an exam due to activities that constitute excused absences according to the University Policy, please reach out to me as early as possible. Unfortunately, illnesses, death in the family or other traumatic events are part of life. I understand how difficult these times are. If you contact me within 24 hours of the event and provide documentation, I will be happy to give you a make-up exam.

Exam Grading: Exams will be graded on a platform called Gradescope, of which the University has paid for an institutional license. The purpose is to make grading more efficient and consistent across multiple graders. If you have concerns about the usage of this platform for grading exams, please let me know.

PROGRAMING

This course involves some programming, as we will be performing data analysis in R or RStudio. R is a free, open-source, high-quality statistical software with millions of users worldwide. RStudio provides an environment for working with R.

> **Words of Advice (3):**
>
> Like learning a new language, it takes time to feel natural about the way R works. Practice helps a lot! The more you use R, play with R, and wrangle with R, the more fluent you will get.
>
> Tutorials are also very helpful. LinkedIn Learning (a UMN supported tool) has many video tutorials on R which you could check out. Once you sign in, search with the key word "R". https://it.umn.edu/technology/linkedin-learning

CLASS RESOURCES

-**Canvas**: This is where we will be posting the handouts, lecture slides, homework solutions, as well as other study materials.

> **Words of Advice (3):**
>
> These materials are for your study only, not for your friends who may take STAT3032 in the future, nor for the entire online community.
>
> Posting the instructors' course materials on other websites without prior permission is strictly prohibited. Please see University Copyright Policy for more information. We've had an unpleasant incident where an instructor's notes ended up on a certain online platform without their permission. Please respect our decision to limit the access of these study materials to only you and your classmates for now.

-**Office Hours:** Please drop by, on your own or with other students, to talk about the course, study skills, homework, school in general, etc. If you are unable to meet during the scheduled office hours, please email us so that we can set up a time that works for you.

-**Canvas Discussion Board:** This is where you can post a question about the course materials and assignments. One of your classmates may just have the answer for it. In addition, some of your classmates may have the same question and they will benefit from your post, and vice versa. We (instructor, TA, and grader) will be monitoring the discussion board every day, to review the posts, evaluate the responses, and provide revisions/corrections/answers.

It is the fastest way to get help if you get stuck (for example, in a homework question). If you email us with a question on homework or general course materials, you will receive the reply: "Please post your question/issue on the discussion board." Please don't feel discouraged. Consider that you will have access to more resources and more helpers on the Canvas discussion board.

ADDITIONAL COURSE POLICIES

Mental Health and Stress Management:

As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance and may reduce your ability to participate in daily activities. University of Minnesota services are available to assist you. You can learn more about

the broad range of confidential mental health services available on campus via the Student Mental Health Website: http://www.mentalhealth.umn.edu.

Disability Accommodations:

The University of Minnesota views disability as an important aspect of diversity, and is committed to providing equitable access to learning opportunities for all students. The Disability Resource Center (DRC) is the campus office that collaborates with students who have disabilities to provide and/or arrange reasonable accommodations.

If you have, or think you have, a disability in any area such as, mental health, attention, learning, chronic health, sensory, or physical, please contact the DRC office on your campus (UM Twin Cities - *612.626.1333*) to arrange a confidential discussion regarding equitable access and reasonable accommodations.

Sexual Harassment:

"Sexual harassment" means unwelcome sexual advances, requests for sexual favors, and/or other verbal or physical conduct of a sexual nature. Such conduct has the purpose or effect of unreasonably interfering with an individual's work or academic performance or creating an intimidating, hostile, or offensive working or academic environment in any University activity or program. Such behavior is not acceptable in the University setting. For additional information, please consult Board of Regents Policy:
http://regents.umn.edu/sites/regents.umn.edu/files/policies/SexHarassment.pdf

In my role as a University employee, I am required to share information that I learn about possible sexual misconduct with the campus Title IX office that addresses these concerns. This allows a Title IX staff member to reach out to those who have experienced sexual misconduct to provide information about the personal support resources and options for investigation that they can choose to access. You are welcome to talk with me about concerns related to sexual misconduct. Within the requirements of my job, I will be as responsive to your requests for confidentiality and support as possible. You can also or alternately choose to talk with a confidential resource that will not share information that they learn about sexual misconduct. Confidential resources include The Aurora Center, Boynton Mental Health and Student Counseling Services.

Other university policies and procedures can here found on this page:
https://policy.umn.edu/education/syllabusrequirements-appa

COURSE SCHEDULE

	Date	Topics (Lecture and Lab)	Assignments/Deadlines (on Thursdays, unless otherwise specified)
1	Jan 21-24	-Course overview -Review of statistics	-Release: Practice questions on the basic statistics -Assigned: Preparation Quiz
2	Jan 27-31	-Simple linear regression models & assumptions -R practice	-Due: Preparation Quiz -Assigned: HW1
3	Feb 3-7	-Regression coefficient estimators and their distributions -Prediction	-Due: HW1 -Assigned: HW2
4	Feb 10-14	-Intervals of predicted response variable value -ANOVA	-Due: HW2 -Assigned: HW3
5	Feb 17-21	-Model diagnostics for simple linear regression -Power transformation -Dummy variable	-Due: HW3 -Released on Wednesday: Study guide and practice exam of Exam1
6	Feb 24-28	-Review -EXAM1 on Wed -Polynomial regression	-Assigned: HW4
7	Mar 2-6	-Multiple linear regression	-Due: HW4

		-Partial F test	-Assigned: HW5
	Mar 9-13	Spring Break	
8	Mar 16-20	-Model fitness vs. complexity -Variable selection	-Due: HW5 -Assigned: HW6
9	Mar 23-27	-Collinearity -Transformation and diagnostics of multiple linear regression models	-Due: HW6 -Released on <u>Wednesday</u>: Study guide and practice exam of EXAM2
10	Mar 30-Apr 3	-Review -EXAM2 on Wed -Logistic regression models	-Assigned: HW7
11	Apr 6-10	**-MONDAY: Last day to cancel class with W on the transcript** -Estimate logistic regression model coefficients -Compare logistic regression models (Deviance and Chi-squared test)	-Due: HW7 -Assigned: HW8
12	Apr 13-17	-time series & weak stationarity -AR models	-Due: HW8
13	Apr 20-24	-Prediction -ACF plot and PACF plot	-Assigned: HW9

14	Apr 27-May 1	-Differencing -MA models	<mark>-Due: HW9</mark> -Released on <u>Monday</u>: Study guide and practice exam of EXAM3
15	May 4	-EXAM3 on Mon (Last day of instruction)	

Note:
- The course schedule is subject to change.
- The <mark>highlighted items</mark> are the deadlines. All assignments are due **in Canvas** at **11:59 pm** of the due dates, unless otherwise specified.
- Generally speaking, when an assignment is due, a new one is posted.
- The practice exam and the study guide are posted a week before each exam.