Math 242: Applied Statistics

Fall, 2016

Course Description: We live in the age of data. Thanks to the ubiquity of computers, data on almost any topic is at your fingertips. But how can you turn all this data into intelligent conclusions? This course aims to teach you this skill, as well as how to work with real-world data sets and how to think probabilistically. Part of this goal will be understanding common fallacies and incorrect statistical reasoning; in this course skepticism is encouraged! You will learn how to run statistical tests, how to measure the reliability of the inferences you draw from the data, and how to design experiments to maximize the reliability of their results. You'll learn how to determine if quantities are correlated (i.e. vary together), and how to find the best fitting curve to explain their relationship. You'll learn the value of simulation, learn how to manipulate data sets using the statistical package R, and conduct a final project using a data set of your choosing. No prior experience with probability, statistics, or programming is required.

Course Goals:

(1) Practice how to break a complex problem into simple pieces. Develop your analytic thinking skills and problem-solving skills. Learn to mold theory-based approaches to solve messy real-world problems.

(2) Develop a probabilistic way of thinking, and the ability to quantify the extent of your (un)certainty in statements you make. Find flaws in others' arguments and develop intelligent skepticism.

(3) Learn to manipulate data sets in both R and Excel. Come to appreciate the power of simulation. Learn skills of data visualization, expectation, estimation, inference, and prediction.

(4) Develop group work skills, test-taking skills, and writing skills. Become proficient at re-writing and communicating technical content to a non-technical audience.

(5) Develop time management skills, metacognitive skills, and the habit of thinking intentionally about your learning and your goals.

Instructor:	David White
Office:	Olin 202
Extension:	6644
Email:	david.white@denison.edu
Class Meetings:	9:30-10:20 MWF and 9:00-9:50 Th, in Olin 217
Office Hours:	1:30pm-2:20pm M, 11:20-12:20pm W, 12:20-1:20 F **and by appointment**, Olin 202
TA Session	12-1pm Tuesday, Olin 216, **and by email** to Gege Tian (tian_g2)
Final Exam:	Monday, December 19, 9-11am
Web Resources:	http://personal.denison.edu/~whiteda/math242fall2016.html

Course Overview This course will be broken roughly into four pieces:

- Descriptive Statistics - Graphical and Numerical Summaries of Data
- Probability - Randomness, Expectation, and Probability Distributions
- Inferential Statistics - Sampling, Estimation, and Hypothesis Testing
- Regression and Analysis of Variance

Text and Materials. The textbook that you will need for this course is Statistical Modeling: A Fresh Approach by Daniel Kaplan, 2nd edition.

Your Final Grade. Your final grade in this course will be calculated according to the following:

Labs:	25%	Homework and Reading Notes:	10%
Semester-long Project:	10%	Midterm Exams	30%
Participation and Quizzes:	10%	Final Exam	15%

Keys to Success

- Prior to exams, be able to solve every quiz and homework problem under time pressure.
- Review the material from class the same day it is given. Find a way to attach this new knowledge to things you already understand.
- Read the textbook slowly and carefully, at your desk, with a notebook nearby to write down questions. Have R open while reading. Type code in, play around, and experiment to figure out what R can do. Pay particular attention to examples, and try to work them out yourself before reading the solution.
- Start homework and projects early. Give yourself time to get stumped and to get past these difficulties. Focus first and foremost on making sure you understand the data set and can compute on it. Finish the statistical analysis part of the lab a full day before it is due so as to have time to write the lab report.
- Study statistics a bit every day rather than in bursts just before an exam.
- Keep a list of key definitions, built-in functions, and syntax, and commit them to memory throughout the course. Test your memory each week.
- Have a perfect, hand-written copy of each homework and quizzes within one day after the problems have been discussed in class. Use this to study for exams.

Homework will be due approximately once per week, and features problems from the textbook, small writing exercises, and exercises in R and Excel. In addition, you are expected to keep up with the reading, and there may be surprise quizzes. In this class, the expectation is that you will have done a careful read of the book,

and completed the Reading Questions, before class begins. We'll start class with a chance to ask questions on the reading, before the quiz. You should expect to spend about 10-12 hours on coursework outside of class per week. Exercise numbers and pages to read will be posted daily on the course webpage.

Collaboration on homework is strongly encouraged, but you should write up your own homework yourself, and it should be well written and readable, showing all your work. I am happy to answer questions in my office hours, and I encourage you to come if you are confused about anything. You will get the most out of this time if you attempt the homework first and come with questions already prepared. I will post homework solutions the day the homework is due, so late submissions will not be accepted. In consideration for sickness, personal emergencies, etc. I will drop the lowest homework grade.

Labs are larger scale projects featuring real-world data sets and requiring written lab reports demonstrating a deeper understanding than the homework. No late labs will be accepted.

Semester-long Project: Early in the semester you and a partner of your choosing will obtain data on a topic you'll choose in consultation with me. Each time we learn something new you'll interpret it or test it on your pet data set. You'll come up with questions about this topic you hope to answer, and may need to obtain more data as the course progresses. At the end of the semester you'll hand in a longer paper answering your questions and give a presentation of your findings. You topic must exhibit sufficient complexity to cover all topics from chapters 1-9 and at least one advanced topic (see course topics below), so will need to contain several variables. I have several data sets you can use.

Exams: There will be a comprehensive 2 hour final exam during exam period. To help you consolidate what you have learned we'll have three exams during the semester, on Monday, October 10, some time between 4 and 10pm, on Wednesday, November 16, between 4 and 10pm, and on Friday, December 16 in class. Each exam should take about an hour, but having the night blocks will give you more time if you need it. You will be permitted to use R during the exams. Understanding the homework problems, the recommended problems (bottom of course webpage), the quiz problems, and the labs is the best form of preparation.

Course Format. The course meets 4 days per week, for 50 minutes each day. Class will begin with a Q & A forum where I will attempt to clear up any confusion you may have. Please take advantage of this and come with questions prepared. We're all in this together and don't want to leave anyone behind. Most classes will contain opportunities to work on problems in groups. This will give you a chance to identify things which may cause confusion on the homework and ask for clarification while we are all in the same room. Thursday will often be a lab day, so you can sharpen your skills in R. I will often give handouts and worksheets in class, so **please obtain a 3-ring binder you can use to store and organize these handouts**. It should be at least an inch thick.

Communication. It cannot be stressed enough how essential communication is to succeeding in this course. After identifying topics that may be giving you trouble, please communicate this information to me. There's no such thing as a bad or unwelcome question. Additionally, please communicate with each other. In a course going this fast, collaboration is almost a necessity to keep up with the material. Explaining concepts and examples to each other is a great way to learn. It is my goal to create a comfortable environment best conducive for learning.

Participation. Since working with other students is a major part of this course, it is important that everyone participate. Class attendance is therefore essential. Each day you'll be graded on a scale from 0-3, with 0 signifying an absence, 2 attending attentively, and 3 active participation such as asking or answering a question.

Grading Scale: A standard 10% grading scale will be used: 60% is required to pass the class, 70% will be a C-, 80% will be a B-, and 90% will be an A-.

Disability: Any student who feels he/she may need an accommodation based on the impact of a disability should contact me privately as soon as possible to discuss his/her specific needs. I rely on the Academic Support & Enrichment Center in Doane 104 to verify the need for reasonable accommodations based on documentation on file in that office.

Academic Integrity: Academic dishonesty is, in most cases, intellectual theft. It includes, but is not limited to, providing or receiving assistance in a manner not authorized by the instructor in the creation of work to be submitted for evaluation. This standard applies to all work ranging from daily homework assignments to major exams. Students must clearly cite any sources consulted, including classmates who have been collaborators on the homework and online sources of aid. Neither ignorance nor carelessness is an acceptable defense in cases of plagiarism.

I expect that you will all abide by the honor code in this course. Please do not use resources outside of me, your fellow students, the tutors, and the textbook. Students are not allowed to copy code from online sources, and should not search for solutions online. Collaboration on homework and projects is permitted, but you should acknowledge in your write-up when you gave or received help on the assignment. Collaboration on quizzes and exams is not permitted. Violations of the honor code will be reported, and violations may result in failure in the course, suspension, or expulsion.

Appropriate Use of Course Materials: The materials distributed in this class, including the syllabus, exams, handouts, study aides, and in-class presentations, may be protected by copyright and are provided solely for the educational use of students enrolled in this course. You are not permitted to re-distribute them for purposes unapproved by the instructor; in particular you are not permitted to post course materials or your notes from lectures and discussions online. Unauthorized uses of course materials may be considered academic misconduct.

Email: I will frequently contact you via email. Please check your email regularly. I will also check my email regularly, but often not after 8pm.

Topics: A rough schedule of topics follows, but is subject to change. The schedule on the course webpage supersedes the schedule here:

Week 0: Course Overview, Statistical Models (only 2 days)

Week 1: Types of Data, Variance, Data Visualization

Week 2: Categorical Modeling, sampling distribution, confidence intervals

Week 3: models as functions, choosing and interpreting models

Week 4: Fitting linear models, correlation, residuals, QQ plots, interaction

Week 5: Geometry of least squares regression, multicollinearity

Week 6: Exam 1, nested models

Week 7: Confidence intervals for regression, prediction intervals, fall break

Week 8: Probability, Monte-Carlo simulations

Week 9: Hypothesis testing, effect size, statistical vs. practical significance, types of errors

Week 10: Inference for regression

Week 11: Cleaning data, exam 2

Week 12: Thanksgiving break

Week 13: Classical statistics

Week 14: Causation and Experiments

Week 15: Student project presentations, exam 3