Syllabus
STA 36-202 –*Methods for Statistical Data Science*
Summer 2018

Course Description

This course builds on the principles and methods of statistical reasoning that were developed in a first-semester intro statistics course, and will cover **regression analysis** (simple and multiple), **analysis of variance** methods, and **logistic regression**. [Additional topics as time allows.] The course will revisit in more detail the methods for examining the relationship between two variables and will also expand the methods to cases were there is more than one explanatory variable.

Prerequisites: 36-200 or 36-201 or 36-207 or 36-220 or 36-247 or 70-207.

Learning Objectives

A student who has successfully completed the course should be able to:

- Demonstrate **conceptual understanding** of the methods covered, and of the basic theory behind those methods;
- Show introductory-level **practical ability** with the methods covered in the course (e.g., to be able to choose the appropriate statistical methods, and to generate and properly interpret the results);
- Exhibit some **critical thinking** about statistics, including the 'validity' of the models applied, as well as the real-world meaning of the statistical results generated.

Course Staff

Instructor

Gordon Weinberg
3719 Wean
gordonw@andrew.cmu.edu
412-268-5496

Instructor Weinberg's office hours:
I will be in the office Mon through Thurs, 1:30PM – 2:30PM. Extra Friday office hours may be added (especially on days of the two projects), and will be announced on Canvas.
In addition to regular weekly office hours, individual arrangements are always possible; please contact me on an individual basis to make such arrangements.

Teaching Assistant

Natalia Lombardi de Oliveira
nlombard@andrew.cmu.edu

Natalia will be helping in the labs and with grading; she will also be sharing in the Shared Summer Statistics Office Hours, which will be posted on Canvas.

Course Requirements and Semester Grade

Your semester course grade consists of:

Labs	10% of grade [all labs counted, nothing dropped]
Homework	15% of grade [lowest hw dropped before calculating average]
2 Projects	30% (15% each) [both projects counted]
2 quizzes	20% (10% each) [both quizzes counted]
Final Exam	25% of grade [final exam not optional]

The scale that will be used for assigning end-of-semester letter grades is as follows:
A = 90 – 100; B = 80 – 89; C = 70 – 79; D = 60 – 69; R (fail) = 59 – 0.

Lecture is not graded; but lecture is probably the best way to learn the material that will be tested in the course.

Materials

▪ Optional Texts

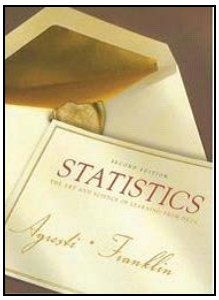
There is no required text for the course.

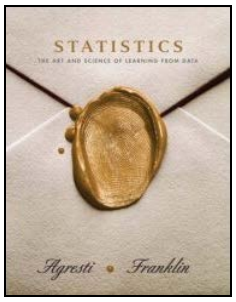
Daily outlines will be posted for download, to guide your note-taking in lecture; lab assignments will be posted; homework answer keys will be posted on a regular basis; and practice materials will be posted prior to each quiz. Most students in past semesters have found these materials to be sufficient for succeeding in the course.

But if you want to browse an optional text, some optional texts for the course are:

Statistics: The Art and Science of Learning from Data,
(1st or 2nd editions)
by Agresti and Franklin.

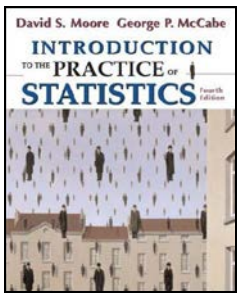
Copies are on reserve in Hunt library;
other copies may be shelved in libraries.

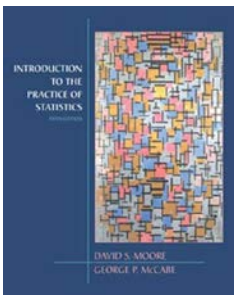


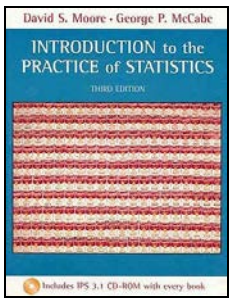


Introduction to the Practice of Statistics,
(various editions available)
by Moore and McCabe.

A PDF of the 6th edition is on Canvas;
copies are on reserve in Hunt library;
other copies may be shelved in libraries.

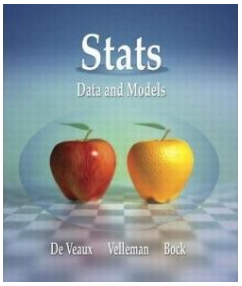


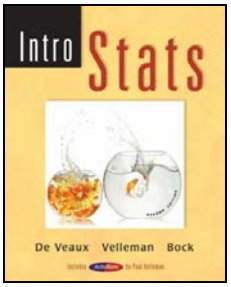


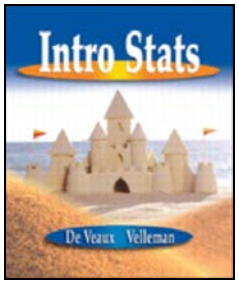


Stats, or Intro Stats,
(various editions available of either)
by DeVeaux, Velleman, and Bock.

Copies are on reserve in Hunt library;
other copies may be shelved in the libraries.







▪ Supplementary review material

If you want to review any of the prerequisite material, in addition to the texts listed above, some other options are listed on Canvas.

Materials, continued

Calculator

For the quizzes, a calculator that can at least do square root is required; for the final exam the calculator needs to also have exponential (e^x) and logarithm functionality. Note that cellular communication devices will not be allowed as calculators during quizzes/exams.

R Software

In the computer lab portion of the course, you will be introduced to the R statistical software package*. The use of R will be reinforced on homeworks and projects.

> * So named because it's the letter before S ["S" being the name of an older Statistics programming language. The commercial paid package based on it became "S plus," so the free/open source package, rather than being named "S minus," was named "R."]

You will use R in an online browser-based utility. Nothing else is required.

On quizzes and the final exam, you will not be required to know any computer commands; but you will be expected to understand and interpret simple computer output or computer-generated graphs.

Class Day and Time

M, W, F, 10:30–11:50AM, Porter Hall 125-B;
T, Th, 10:30–11:50AM, Baker 140 E & C computer clusters
[note, for lab please go to the 140 E room; the class should be small enough that we won't need to split into the other room]

Daily Lecture Outlines

Each day's lecture will use prepared outlines which will contain material like graphs, examples, or datasets that the instructor will refer to throughout the lecture, and which will also contain spaces for you to take notes. For each lecture, you will need a copy of that day's outline to view in lecture (either in hardcopy or else in electronic format), and you will also need note-taking materials.

Daily lecture outlines will be posted electronically on Canvas:

https://www.cmu.edu/canvas/ → "files"

Each lecture outline will generally be available online a day in advance.

You do not need to bring any textbook to lecture.

Homework

How and When Homework is Assigned and Collected

Homework assignments will be posted electronically on Canvas:

https://www.cmu.edu/canvas/ → "files"

Waitlisted students will be given temporary 'observer' Canvas access, and are expected to submit homework with the rest of the class.

Homeworks will be submitted electronically, and generally due by 10AM on Tuesdays and Thursdays. [Assignments will be posted at least a few days prior.]

Late homework will not be accepted for credit.

Homework Content and Purpose

Homework will emphasize the course material covered in lecture, but may also contain exercises that extend the course material beyond lecture. Homework will also help to practice and solidify skills in creating presentation-quality output through R markdown.

You should generally be spending a few hours on each homework.

Missed Homework Policy

Homework will not be eligible for any credit after the posted due date/time.

One homework score is dropped at the end of the semester to account for illness or other emergency reasons for missing a homework submission. This policy is chosen instead of 'extensions' so that that all homeworks are graded together (which ensures grading uniformity), and to avoid the need to evaluate requests for extensions (which are inevitably subjective and thus potentially unfair to students).

Homework Grading

Homework will be graded by the course Teaching Assistant. Partial credit will be given where appropriate.

Lowest Homework Score

Your lowest one (1) homework score will be dropped before tallying your semester homework average. This is intended to account for illness or other emergency.

Computer Labs

Purpose of Lab

Twice-weekly computer lab assignments will give you practical experience analyzing real data, using the R programming environment.

R is the standard research-level package for statistical computing*. The lab portion of the course will introduce you to R and some of its capabilities relevant to the course material. You will also use R to complete the homeworks and projects. (On quizzes and the final exam, you will not be required to know any computer commands; but you will be expected to understand and interpret simple computer output or computer-generated graphs.)

> * So named because it's the letter before S ["S" being the name of an older Statistics programming language. The commercial paid package based on it became "S plus," so the free/open source package, rather than being named "S minus," was named "R."]

Each lab day, there will be a lab assignment posted electronically on Canvas:

https://www.cmu.edu/canvas/ → "files"

The course staff will be available to help you.
Each lab assignment will generally be based on recent lecture material, so lab will also serve as preparation for the upcoming homework assignment on that same material.
You can take your notes to lab to help you.

Lab Credit

Labs will be worth 2 points apiece, as follows:

1 pt for attending lab and working for the hour;
1 pt for submitting the completed lab online by 7PM.

The labs will be completed in a utility running on the departmental server; therefore, server log info will indicate whether you were present and working for the lab hour. Paper attendance may be used in lab to supplement as needed.

Submitting a lab assignment electronically *without* attending lab will not be eligible for any credit.

The submitted lab is generally graded leniently, but if you submit a mostly incomplete lab (or a lab with considerable errors), part of the submission point may be docked.

No Dropped Labs

No labs will be dropped. All lab scores will count towards your semester lab average.

Lab Attendance

Please be on time to lab. Severe or chronic tardiness risks lower lab grade.

Waitlisted students should attend lab.

Data Analysis (DA) Projects

There will be 2 data projects. The first project will be due at 7PM Friday July 20; the second project will be due 7PM Friday August 3. Each project will be assigned at least several days in advance.

For the projects, you will get to choose a real dataset (we will provide a selection) to analyze, using the course concepts and methods taught up to that point. In the projects, you will execute the entire data analysis workflow, from articulating the real-world motivation for the question, to the exploration of the data, then building and validating the appropriate statistical model, and using that model to answer the question of interest. Appropriate guidance and prompts will be provided.

Projects will be done 'report style' (you should think of them like a report that you would submit to a statistical consulting client or to an academic research journal). Using the software capabilities, you will learn how to achieve publish-quality data analysis reports either in HTML or in PDF format.

Projects will be graded by the course staff. Grading rubrics will be provided with project instructions.

Quizzes/Exams

Dates/Time/Location/Coverage

Quiz 1: Friday, July 13, first half hour of lecture (Porter 125-B).
Covers regression material up to that point. [Practice material will be provided.]

Quiz 2: Friday, July 27, first half hour of lecture (Porter 125-B).
Covers ANOVA material up to that point. [Practice material will be provided.]

Final Exam: Friday, August 10, during regular lecture time (10:30AM-11:50AM), in the regular lecture room (Porter 125-B). Coverage cumulative.

Quiz/Exam Format

Quizzes and the final exam will generally be a combination of multiple-choice and 'work-out' exercises; some exercises may also ask for short-answer interpretation.

Required / Allowed Materials on Quizzes and Final Exam

Exams will be closed-book. Tables of values or computer output will be provided as needed.

For the first quizzes, a calculator that can at least do square root is required; for the final exam the calculator needs to also have exponential (e^x) and logarithm functionality. Note that cellular communication devices will not be allowed as calculators during exams.

One (1) standard (8 ½ " by 11 ") sheet of notes, front and back, will be allowed on each quiz; two (2) standard (8 ½ " by 11 ") sheets of notes, front and back, will be allowed on the Final Exam. Your notesheets may contain any information you like, and may be produced in whatever format you choose (written, typed, printed, photocopied, etc.). No other notes will be allowed.

Exam Policies, continued

Missed Quiz Policy

Absence from quizzes may be excused at the discretion of the instructor. If the absence is excused, the missing quiz grade will be replaced with the grade on the final exam grade (or on the parts of the final exam specific to the missing quiz material).

Pre-arranged absence from quizzes (e.g., planned trip, extracurricular participation) must be discussed with the Instructor a reasonable amount of time prior to the quiz.

Verifiable documentation may be requested where appropriate. Oversleeping is not a valid excuse.

No makeup quizzes will be given. Unexcused absence from quizzes or exams will merit a zero on the quiz or exam.

Academic Honesty on Exams

During quizzes/exams, all non-related material, including cell phones, must be stored out of reach. Appearance of giving or taking unauthorized assistance during exams will be subject to penalties under the University cheating/plagiarism policy.

Special Accommodations

Eligibility for special accommodations on exams is determined by the Carnegie Mellon Office of Equal Opportunity Services (EOS).

If you are eligible for special accommodations on exams, it is your responsibility to notify the instructor in a reasonably timely manner.

Picking-up / Retaining Graded Materials

Graded quizzes will be made available the following Mondays. Students are expected to collect their graded materials in a timely manner.

Students are expected to retain graded exams and all other returned materials for the duration of the semester, to study from, and also in case of online gradebook problems. In case of emergency problem with the online gradebook system, you may be requested to hand-back some graded materials in order to verify the grade.

Academic Honesty

Meaningful and vibrant discussion on the course material is one of the best ways to learn. We hope that you will become so engaged in the course material that you will be discussing, debating, and collaborating with other students. We want to incentivize honesty and meaningful effort.

Therefore, if you work with someone else on an assignment (for instance, with a class partner or a study group of students from the class), or if you get assistance on an assignment from someone (even if they're not in the class), you should disclose that fact, for instance by writing, "**I worked on this assignment with … [and then the complete list of names of everyone you worked with]**" at the top of your assignment. This is akin to the proper academic practice of listing all the collaborators on a scholarly article.

Disclosing your study partner(s) is a necessary step in academic openness. However, disclosure is not sufficient to indicate individual effort. **Assignments in 36-202 are graded individually and not as group projects, therefore each student's assignment should demonstrate individual effort**. Identical or suspiciously-similar work will still be docked for credit if it indicates lack of meaningful individual effort. If you work on an assignment with a study partner, you should re-write your assignment with your own words and your own work before submission.

A good recommendation for getting the most out of collaboration while still ensuring academic honesty and genuine individual learning is to use the '**one-hour whiteboard rule**,' a technique akin to something used in other departments, in which students may discuss the assignment together using a whiteboard, *but you are not allowed to copy anything down while you are looking at the whiteboard*. You must then go somewhere else away from the whiteboard, wait a reasonable amount of time (like an hour), and then each student writes up their assignment individually *without any further collaboration during the write up of the assignment*. This technique forces you to check your own understanding of what you are writing, and it helps to ensure that no two students should have the identical work or words.

Cheating / Plagiarism

▪ Definition

Cheating and plagiarism are defined by University policy which is available online: https://www.cmu.edu/policies/student-and-student-life/academic-integrity.html

▪ Penalties

Course penalties can range from a zero on the item to an R ("fail") in the course. Additional penalties are possible beyond the course grade, as described on: https://www.cmu.edu/student-affairs/ocsi/academic-integrity/documents/academic-disciplinary-actions-overview-for-undergraduate-students.2013.pdf

What to Bring to Lecture

Each day's lecture will use prepared outlines which will contain material like graphs, examples, or datasets that the instructor will refer to throughout the lecture, and which will also contain spaces for you to take notes directly on the outline.

Daily lecture outlines will be posted electronically on Canvas:

https://www.cmu.edu/canvas/ → "files"

Each lecture outline will generally be available online a day in advance.

You do not need to bring any textbook to lecture.

Lecture Attendance and Behavior in Lecture

Lecture attendance is not graded in 36-202. However it is to the student's benefit to attend each lecture. The reason we have lecture in the first place (rather than just a textbook) is because most people find it much easier to learn from a live person.

Some studies at other universities have suggested that skipping lecture may be associated with a lower course average by a few percentage points for each lecture missed.

While in lecture, please be respectful to those around you. Make sure your cell phone is off during lecture; please don't leave until lecture is finished, and please don't engage in non-course-related chat during lecture (note that sound carries well in the lecture hall).

Please also be respectful of the room, by keeping food or drink to a minimum, and picking up after yourself before you leave.

Recommended Weekly Habits for Success

1. First, **review the previous lecture's notes** before each lecture (to have the recent material fresh in your mind), and **read the upcoming topics in an optional text if you feel the need** (to familiarize yourself with the topics that will be covered, and so that there is time for potential questions to occur to you, which you might want to ask in lecture).

2. Prepare to view a copy of the outlines in lecture, by printing them the night before each lecture; or else by downloading them to your laptop or tablet computer.

3. Then, **attend lecture, and be engaged** while in lecture (try to anticipate answers to examples as they are presented in lecture, and ask questions in lecture); note that lecture outlines will be generally made available, but they will only be topic outlines and therefore won't be very useful to you if you don't attend lecture.

4. Next, after each lecture, **re-write your lecture notes** 'in your own words' (to test if you can explain the ideas to yourself without any logical gaps and to practice appropriate statistics terminology), and re-do any lecture examples (to take the time to go through each step carefully on your own).

5. Attend labs and **be engaged**: Try to envision how each lab's topics fit into a larger overall picture of the course material; teach and debate the lab questions with your peers; and formulate thoughtful questions to ask the lab TAs (you are encouraged to discuss labs in office hours for more feedback).

6. Review your lab answers after lab; and complete the lab assignment on your own if you did not have time to complete it during lab hour.

7. Then, **do the homework assignments**; spend at least a few hours alone on the homework and formulate an answer on your own for each exercise; then ask about the homework in office hours, and re-do if necessary before submitting.

8. Attend office hours, even if you don't have questions on a current homework assignment (ask questions on lecture or reading or lab; and have the TA or instructor go over previous graded material with you); gaps in your understanding that you might not have realized you had can be identified through one-on-one discussion.

9. Get **a good night's sleep** prior to each quiz/exam. A clear head is more important than cramming.

Personal Stress Care and Psychological Assistance

Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep, and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefits from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, the University strongly encourages you to seek support. Consider reaching out to a friend, faculty, or family member.

If you or someone you know is feeling suicidal or in danger of self-harm, call someone immediately, day or night:

- **CMU Counseling and Psychological Services (CaPS)**

  412-268-2922
  http://www.cmu.edu/counseling/

- **ReSolve Crisis Network**

  888-796-8226

- **CMU Campus Police**

  412-268-2323

  (or dial 911 for emergency if off campus)

Welcome, and good luck!