

# Syllabus

## STA 36-202 –*Statistics & Data Science Methods*

### Spring 2018

#### Course Description

This course builds on the principles and methods of statistical reasoning that were developed in 36-201 (or an equivalent first-semester intro statistics course), and will cover **regression analysis** (simple and multiple), **analysis of variance** methods, and **logistic regression**. The course will revisit in more detail the methods for examining the relationship between two variables and will also expand the methods to cases where there is more than one explanatory variable.

Prerequisites: 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

### ▪ Instructors

Gordon Weinberg  
3719 Wean  
[gordonw@andrew.cmu.edu](mailto:gordonw@andrew.cmu.edu)  
412-268-5496

Instructor Weinberg's office hrs:  
Tuesdays, 10:30AM – 12:30PM.

In addition to regular weekly office hours, individual arrangements are always possible; please contact me on an individual basis to make such arrangements.

Rebecca Nugent  
232-C Baker  
[rnugent@andrew.cmu.edu](mailto:rnugent@andrew.cmu.edu)  
412-268-7830

### ▪ Teaching Assistants

Various TAs will lead the lab sections and will share in the grading.

Some of the TAs will hold office hours (times and locations to be posted on Canvas when available).

## Course Requirements and Semester Grade

Your semester course grade consists of:

Labs	10% of grade [all labs counted, nothing dropped]
Homework	15% of grade [two lowest hw's dropped before calculating average]
2 Projects	20% (10% each) [both projects counted]
2 hourly exams	30% (15% each) [all exams counted, nothing dropped]
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.

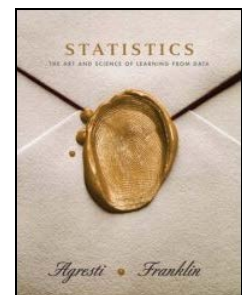
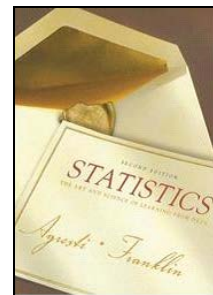
Daily outlines will be posted for download, to guide your note-taking in lecture; lab assignments will be posted and hardcopy lab materials will be available in each lab; homework answer keys will be posted on a weekly basis; and practice materials will be posted prior to each exam. Most students in 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.

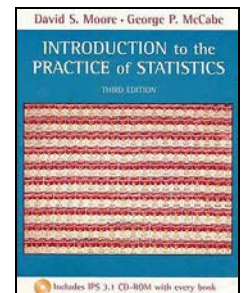
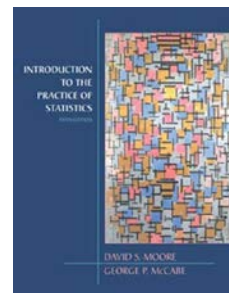
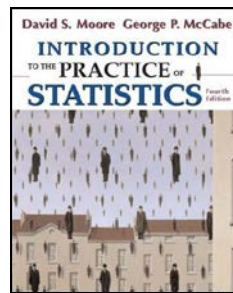
**Copies are on reserve in Hunt library;**  
other copies may be shelved in libraries.



---

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

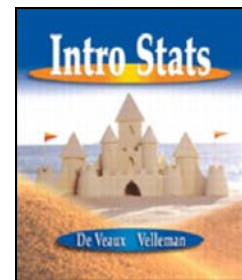
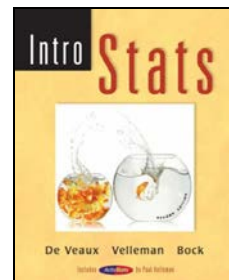
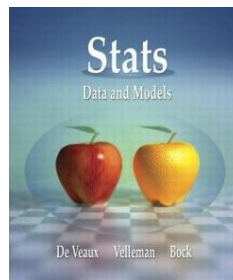
**A PDF of the 6th edition is on Canvas;**  
**copies are on reserve in Hunt library;**  
other copies may be shelved in libraries.



---

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 first two midterms, a calculator that can at least do square root is required; for the third midterm and 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.

- **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.

On exams, 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.

## Lecture Day and Time

M, W, F, 9:30 – 10:20AM, GHC 4401 (Rashid Auditorium).

- **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.

## Lab Day and Time

Section A:	Thurs, 9:30–10:20AM, Hunt Library basement “near” and “far” clusters
Section B:	Thurs, 12:30–1:20PM, Baker Hall, 140 C and 140 E clusters
Section C:	Thurs, 3:30 – 4:20PM, Hunt Library basement “near” and “far” clusters
Section D:	Fri, 10:30–11:20AM, Wean 5202 (“Collaborative Teaching Cluster”)
Section E:	Fri, 11:30AM–12:20PM, Wean 5202 (“Collaborative Teaching Cluster”)
Section F:	Fri, 1:30AM–12:20PM, Hunt Library basement “near” and “far” clusters

The Teaching Assistants will lead the lab sections.

You are encouraged to take your notes, calculator, and previous graded materials to lab.

## Waitlists and Registration Issues

Note that registration for 36-202 is handled by lab section. If you drop a lab section, you will start at the ‘back’ of any waitlist for whichever section you then add, even if it is the same section you originally dropped. Conduct drop/add with caution.

Note that the TAs are not empowered to make registration decisions. If you have registration issues or waitlist questions, communicate with the instructors.

**While you are waitlisted, you should attend lecture, you should attend the lab section you are waitlisted for, and you should still submit homework along with the rest of the class.**

Waitlisted students will be given temporary ‘observer’ Canvas access in order to access lecture outlines, lab materials, and homework assignments. If you are waitlisted but have not yet been given ‘observer’ Canvas access, please let the instructor know as soon as possible.

We ask that you be removed from the waitlist if you are still not officially registered by the ‘semester course drop deadline.’

If you can't get into the course, note that it is also offered in upcoming Summer and Fall semesters.

## 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.

Homework assignments will generally be posted on Wednesdays.

Homeworks will be submitted electronically, and generally due by 11:59PM on Wednesday of the week following the posting of the assignment.

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 homework each week.

### ▪ Missed Homework Policy

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

Two homework scores are 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 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), as well as to avoid the potential loss of homeworks (which is an especial risk in a course as large as 36-202).

### ▪ Homework Grading

Homework will be graded by the course Teaching Assistants, with different TAs grading different exercises, to ensure uniformity and fairness. Partial credit will be given where appropriate.

### ▪ Lowest Homework Scores

Your lowest two (2) homework scores will be dropped before tallying your semester homework average. This is intended to account for occasional illness or other emergency.

## Computer Labs

- **Purpose of Lab**

Weekly computer lab assignments will supplement other aspects of the course by giving you practical experience analyzing real data, in a more intensive and dynamic way than lecture can provide.

Each week, there will be a lab assignment posted electronically on Canvas. The Teaching Assistants will be available to help you.

Each week's 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 should take your notes to lab.

Computer lab is also where graded exams will be returned. It is the student's responsibility to collect graded materials in a timely manner throughout the semester.

- **Lab Credit**

You will get attendance credit for attending lab. If you attend lab and then submit the completed lab electronically by 11:59PM of the same day as your lab, the lab will be scored for secondary credit.

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

- **No Dropped Labs**

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

## Lab Policies, continued

### ▪ Lab Attendance

Please attend only the lab section you are registered (or waitlisted) for.

Attending a different lab section may cause physical space problems (registration forbids ‘over-packing’ the rooms), and will cause grading difficulties (your lab scores may get lost).

Requesting to attend a lab section different from the one you are registered for is potentially unfair to the course staff (because such requests are inevitably subjective), and also unfair to those students who are on waitlists.

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

Waitlisted students should attend lab.

### ▪ Make-up Labs

Teaching Assistant contact for make up labs will be posted on Canvas. Make up lab sessions are primarily intended for students who miss a lab due to prearranged absence such as for extracurricular activity. Make-up labs may be granted on a case-by-case basis, at the discretion of the Teaching Assistant or Instructors.

Requests for make up labs must be made by the weekend directly following the missed lab. Verifiable documentation of reason for absence should be furnished when possible.

Makeup labs are intended for absence due to extracurricular participation, or rare emergencies. Chronic abuse of the privilege may be penalized.



---

## Data Analysis (DA) Projects

There will be 2 data projects, each due the week following the two hourly exams.

Details will be posted.

## Exams

### ▪ Exam Coverage

- Exam I: Monday, February 26 (in lecture).  
Covers regression models (the material from approximately the first 5 weeks).  
Will be held during the regular lecture time in the regular lecture room.
- Exam II: Monday, April 9 (in lecture).  
Covers ANOVA models (the material from approximately week 6 to 12).  
Will be held during the regular lecture time in the regular lecture room.
- Final Exam: Date, time, and location to be determined by University Registration.  
*You should avoid making end-of-semester travel plans until the Final Exam date has been made available.*  
Coverage: Cumulative, with coverage approximately in proportion to lecture  
(including material not on exam II).

### ▪ Exam Format

Exams will generally be a combination 'work-out' exercises, multiple-choice, fill-in-the-blank, and short-answer interpretation.

Exams will represent exercises, concepts, formulations, and terminology from throughout the course.

### ▪ Required / Allowed Materials on Exams

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

For the first two midterms, a calculator that can at least do square root is required; for the third midterm and 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 midterm; 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 Exam Policy

Absence from exams may be excused at the discretion of the instructor. If the absence is excused, the final exam grade will replace the missing hourly exam.

Pre-arranged absence from exams (e.g., planned trip, extracurricular participation, or court appearance) must be discussed with the instructor a reasonable amount of time prior to the exam.

All excuses must be accompanied by verifiable documentation. Notification, such as phone message, that is unaccompanied by supporting documentation does not constitute valid 'documentation'; and oversleeping is not a valid excuse.

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

### ■ Academic Honesty on Exams

During 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 on Exams

Eligibility for special accommodations on exams (due to disability or special needs) 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, and **it is your responsibility to make separate exam arrangements with the EOS office prior to each exam** (no later than one week prior to each).

### ■ Picking-up / Retaining Graded Exams

Graded exams will be made available in lab, usually the same week the exam is taken. 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. **For assignments in 36-202 that are graded individually and not as group projects, 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: <http://www.cmu.edu/policies/documents/Academic%20Integrity.htm>

- **Penalties**

Course penalties can range from a zero on the item to an R ("fail") in the course.

In addition to course penalties, each cheating or plagiarism violation is also formalized by a letter, describing the incident, which is sent to the Office of Academic Affairs, as required by University regulations (described on <http://www.cmu.edu/policies/documents/AcadRegs.html>), and which becomes part of your Carnegie Mellon record. A total of two such letters accumulated during your undergraduate tenure at CMU can result in a suspension hearing.

## 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 at least a day before each lecture and storing them in a folder to bring to class; or else by downloading them to your laptop or tablet computer, recharging the battery, and then bringing the laptop or tablet computer to class.
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** (which are designed for you to practice each week's recent material, as preparation for the following week's homework), **be engaged** (try to envision how each week's topics fit into a larger overall picture of the course material; teach and debate the lab questions with your lab partner; and formulate thoughtful questions to ask the lab TAs), and **complete any unfinished labs on your own time** (you are encouraged to discuss labs in office hours for more feedback).
6. **Re-write 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 weekly homework assignment** (probably at the end of each week); 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-write if necessary before submitting.
8. **Attend office hours** every week, 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 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)

**Schedule in Brief**

(subject to modification as necessary)

Lecture Weeks	Topics
1 – 5	Review: Exploratory Data Analysis, and the basics of statistical inference. Simple Linear Regression. Non-linear regression: transformations and polynomial regression. Multiple regression. Multicollinearity. Including categorical variables in the regression model (the Partial F-Test). Model Selection: Stepwise regression versus the “Best Subset” procedure.
6 – 12 [includes spring break]	One-way Analysis of Variance. Multiple comparisons. Two-way Analysis of Variance (Interaction). A single factor repeated measures design. The two-factors (between subjects) repeated measures design. The mixed design.
13 – 15	Simple logistic regression Multiple logistic regression Nominal logistic regression Ordinal logistic regression

Welcome, and good luck!