STT 231: Statistics for Scientists | Introduction & Syllabus

To the student

STT 231 is an introduction to statistical analysis for students with zero- to one-year of prior experience in statistics or calculus. This audience includes undergraduate students in science, engineering, and medicine.

Our approach includes minimal statistical theory; notation is kept as simple as possible so that the material remains accessible to a broad audience. Instead, this course focuses on applications of statistics – its uses, misuses, and occasional abuses. The assumptions made by various models and procedures are investigated at length and criticism of procedures' performances under various conditions is encouraged. We want you to not only know how to "do statistics" but how to recognize when someone else might not! The idea of random variation, one of the most important and perhaps most misunderstood aspects of statistics, is made immediate and real through many exercises.

The relationship between statistics & the sciences

A widely recognized law of nature, proposed by Einstein in 1905, is $E = mc^2$, which relates $E = energy$ to $m = mass$ in terms of the constant c^2, the speed of light in a vacuum. The consequence of this equivalence between mass and energy cannot be understated – it has directly and indirectly led to many technological advances and improved the living standards of millions. There are many such laws describing the observed world around us. Generally, these laws express some exact relationship among two or more variables. Other examples include:

$A = \pi r^2$	The area of a circle, relating $A = area$ to $r = radius$.
$kT = PV$	The Ideal Gas Law relating the absolute temperature (T) of a confined gas to its volume (V) and pressure (O).
$d = \frac{1}{2}gt^2$	Newton's expression for the $d = distance$ traveled by free-falling object under the acceleration of gravity for a period of time (t).

Each of these equations is a triumph of science, representing a scientific phenomenon in a concise, useful way. To bastardize a famous quote of Ernst Mach (after whom 'Mach speed' is named), "[Science] is experience arranged in economical order." Every science can be situated in the pursuit of arranging and making sense of our lived experience. Some fields of science are particularly interested in understanding chemical reactions, others study biological processes, still others geological processes, and so on and so on. **Statistics** is not a science in this sense, because it does not focus on better arranging observed phenomena as do other scientific fields. Rather, statistics is more concerned with the process of arranging than the results themselves. Statistics as a field of study is best described as the science of reaching conclusions from observed data. More generally, statistics can be regarded as a body of methods for helping us learn from experience. In a world of uncertainty, statistics is (perhaps derisively) defined as the science of "making the least wrong decision" based on the data at hand.

It follows, then, that the material presented to you in this course *should make sense*. If it does not, you are likely misunderstanding something, insofar as statistics is often an extension of commonsense. Many statistical concepts and procedures require multiple passes before really making sense.

Finally, remember that statistics is *not* an avenue to 'the Truth' with a capital 'T.' Some data contain too little information to answer our questions. Perhaps they were poorly recorded or the number of observations were too few. Our hope is that you'll learn to recognize these situations. Some scientific questions can't yet be answered at all and their stubbornness cannot be remedied by some statistical formula or incantation. The fact is that to "prove" anything at all – in any way at all – is typically quite difficult (this is among the reasons the scientific laws above are such triumphs!). Statistics is often used as a way of telling us how much or how little can be learned from our observations and when we do arrive at a conclusion, just how certain or uncertain we should be about it. We're excited to be working with you this semester.

Syllabus

Learning in a pandemic

In response to the COVID-19 pandemic, this course is being delivered in a remote, asynchronous format. There are no required class meetings, nor attendance. Weekly lectures will be recorded and posted to D2L for students every Monday morning. Assignments are always due on Sundays and assessments always take place on Thursdays.

Even when we're learning apart from one another, we believe **strongly** in the power of collaborative learning when it comes to the field of statistics. Many assignments *will* require you to directly collaborate with your peers, either via video-conferencing app (such as Zoom or Facetime) or via document-sharing platforms (such as Google Docs, Microsoft OneDrive, etc.). This syllabus provides lots of detail regarding how the class will progress (and persist) through the COVID pandemic, whatever it may bring.

Additionally, the STT 231 instructional team will schedule many weekly office hour walk-in and appointment opportunities. We strongly encourage you to take advantage of these opportunities to directly meet with your instructors and teaching assistants. Scheduling information below:

Instructor & TA Office Hours*

Instructor: John Keane, stt231team@msu.edu
Zoom join code:
Times:

Instructor: Dr. Pramod Pathak, stt231team@msu.edu
Zoom join code:
Times: MWF, 3:00pm – 4:00pm EST

Instructor: Dr. Aklilu Zeleke, stt231team@msu.edu
Zoom join code:
Times:

GTA 1: name, stt231team@msu.edu
Zoom join code:
Times:

GTA 2: name, stt231team@msu.edu
Zoom join code:
Times:

GTA 3: name, stt231team@msu.edu
Zoom join code:
Times:

*The statistics learning center (SLC) offers additional office hours for introductory statistics courses via Zoom. This service is not specific to STT 231, but is offered by the STT department and is designed to supplement a student's learning and comprehension of classroom material. Note that the SLC is a supplement, not a substitute, for instruction. Links to the SLC along with its schedule can be accessed from the course homepage on D2L.

Description and prerequisites

Calculus-based course in probability and statistics. Probability models and random variables. Estimation, confidence intervals, tests of hypotheses, and simple linear regression with applications in sciences.
Prerequisites*:* MTH 124 or MTH 132 or MTH 152H or LB 118

Course Materials

- **Required – WebWork subscription:** The WebWork platform will be used for most homework assignments and many assessments. Students will be required to purchase a subscription via CashNet at a cost of $50.00. WebWorK will accept payments from 9/25/2020 to 10/2/2020.
- **Required** – **R and RStudio**: This course uses a free statistical programming language ('R') and a convenient interface ('RStudio') as an aid in exploring statistical concepts and practicing various procedures.
- **Optional – Textbook:** This course uses *Introductory Statistics with Randomization and Simulation*, David M Diez, Christopher D Barr, Mine Çetinkaya-Rundel, 2014 as a guide for many of its lectures. It is a free, open source textbook available as a tablet-friendly PDF online.

Grading Policy

Component	Peer Writing Assignments	RStudio Assignments	WebWorK Homework	Assessments	Total
Percentage	20%	20%	25%	35%	100%
Policy	Drop 1 lowest	Drop 1 lowest	Drop 1 lowest	Drop 1 lowest	-

Your grade will be based on a weighted scale. Your percentage score on each component of the course will contribute to your final class standing according to the breakdown provided in the table above. Grades are rounded to the nearest percentage before being converted to a GPA-equivalency. If your final percentage is 89.49%, your final grade will be a 3.5. If your final percentage is 89.50%, your final grade will be a 4.0. Many people end up close to cutoffs, but moving cutoffs only creates more people being close to the new cutoffs, etc. *Email requests to round grades or change the cutoffs at the end of the semester will not receive a response.*

GPA	Range
4.0	100% - 90%
3.5	85% - 89.99%
3.0	80% - 84.99%
2.5	75% - 79.99%

GPA	Range
2.0	70% - 74.99%
1.5	65% - 69.99%
1.0	60% - 64.99%
0	0% - 59.99%

An I-Incomplete may be given only when: the student (a) has completed at least 6/7 of the term of instruction, but is unable to complete the class work and/or take the final examination because of illness or other compelling reason; and (b) has done satisfactory work in the course; and (c) in the instructor's judgment can complete the required work without repeating the course.

Graded course components

Peer Writing Assignments (20%)

At the beginning of the semester, you will be assigned to a group of three other students as part of a collaborative writing team. Many weeks, your group will be given a writing prompt on Monday mornings and tasked with creating a group essay between 1 and 2 pages in length responding to this prompt. These group essays will be submitted via D2L. The following week, you'll receive another group's essay and be asked to conduct a peer review / response to their work. The purpose of these assignments is two-fold:

1. **We stress good writing as a method of learning statistics.** We regard writing-to-learn as an important aspect of statistical thinking. These peer writing assignments call on you to write interpretations and explanations of your statistical findings and to respond to your peers' submissions. Moreover, they emphasize that, while statistics is a branch of mathematics, it is also a discipline that encourages certain *types* of thinking. It is not enough to be able to 'do statistics,' one must also be able to interpret, evaluate, and critique them.
2. **We promote collaborative learning among students.** These assignments provide a natural occasion for encouraging you to work in groups, allowing you to collaborate and learn from one another as much as you do from course lectures, if not more. We believe that engaging with your peers is particularly important in remote learning contexts.

RStudio Assignments (20%)

Many weeks, you will be individually assigned an "RStudio Assignment" via D2L on Monday mornings. To complete these assignments, you will download a word document from D2L along with its associated data set(s). Thereafter, you'll be asked to complete a series of exercises analyzing that data set using statistical methods recently covered in lecture and directly type your responses to the word doc file. Once finished, you'll submit your edited copy back to D2L by the associated deadline. The purpose of these assignments is two-fold:

1. **We want to use technology as a tool.** Students should regard technology as an invaluable tool both for analyzing data and for studying statistical phenomena. Technology performs calculations and presents the visual displays necessary to analyze genuine datasets that are often large and cumbersome. Additionally, we believe RStudio offers students the opportunity to use simulations to visualize and explore the long-term behavior of sample

statistics under repeated random sampling, a concept that underpins most of the statistical procedures we explore in STT 231.

2. **We want to look at real data.** Analyzing genuine data not only exposes students to what the practice of statistics is all about, it also prompts them to consider the wide applicability of the methods we explore.

WebWorK Homework (25%)

- There will be homework assignments due via Webwork most weeks throughout the semester. Typically, assignments will open on Mondays at 12:01 AM EST and remain open until exactly 11:59 PM EST the following Sunday.
- Most exercises in each homework assignment offer you multiple attempts to earn full credit. After each attempt, some will provide feedback on portions you should consider revising and others will not. Be mindful to use your attempts carefully, as many WebworK exercises are nit-picky regarding answer precision and rounding.
- When a HW assignment is posted, you will have seen *most* corresponding material. Start HW early when it opens. Past students have found out too late that starting HW a few hours before it is due means too little time to think about the questions, review material, do statistical analyses, or ask questions. *Do not wait until the night before.*
- HW assignments must reflect your own work — you can talk to others [in fact, we encourage it!], but calculations and final answers must be your own.
- Once the HW due date/time has passed, the solutions will become immediately available online through WebWork. Your HW will be graded over the weekend using a combination of automated and by-hand methods.
- **No late HW will be accepted** (solutions are provided immediately), we do know things can come up. Your one lowest HW score will be dropped at the end of the semester before computing the HW part of your semester grade.
- The purpose of these web-based homework assignments is two-fold:

1. **We provide ample opportunities to repeatedly practice statistics**. Many statistical concepts require two or three exposures to truly digest and internalize. Webwork allows students the opportunity to make mistakes and learn from them through immediate feedback.
2. **We think assessments should never be a surprise.** All Webwork questions are based on the formatting and scope of our assessment questions. These homework assignments offer you a built-in opportunity to always be studying for upcoming assessments.

Assessments (35%)

You will have several evenly weighted assessments throughout the semester. Each assessment will emphasize the content covered since the previous one, and roughly follow the tentative course schedule listed at the end of this syllabus. This is to say that no assessment will be cumulative. [Be aware that concepts in the course *are* cumulative and that some questions indirectly make use of concepts from previous learning objectives]. **All assessments are timed, open-book, open-notes, and administered via WebWorK.** Students will have an 18-hour window to take an assessment on the Thursday of the week there are assigned. Once a student begins, they will have 90 minutes to complete all exercises therein.

There are **no make-up assessments, so check your calendars against the tentative course schedule now**. Students with conflicts or documented accommodations for extended time on tests must email stt231team@msu.edu, include details, and turn in any documentation by Wednesday, September 16th, 11:59 PM EST.

Communication amid COVID

In remote learning environments, it is critical to communicate clearly and frequently with your instructors, especially if you encounter a prolonged period where you are unable to engage in course content. We have anticipated that some students may encounter interruptions to their studies for any number of reasons (e.g., illness, need to provide medical or child care, sustained loss of internet, etc.) and have plans in place accordingly. To provide built-in flexibility to

students as they progress through their coursework this semester, the lowest score of each category of assignments (Peer Writing Assignments, RStudio Assignments, Webwork, and Assessments) will be dropped at the end of the semester before grades are tabulated.

In extreme cases where a student cannot reliably progress through course content for more than one week, they must **immediately inform their instructors** of their situation via stt231team@msu.edu so that individualized accommodations can be made. If a student is unable to attend class and can provide documentation of the obstacle they face, we intend to work hard to accommodate their situation with empathy.

In general, members of the instructional team make every effort to respond to emails within one business day.

Because of the team-teaching approach implemented in STT 231, we do not anticipate any issues arising if one or more instructors are absent for an extended period of time. Many assignments are graded automatically, and those that are graded by hand are evaluated by a team of individuals. Thus, students can expect most grades to be posted to their D2L gradebook within a week of the due date passing. If there are any changes to our teaching roster, students will be notified via email and D2L announcement.

Academic Integrity

Michigan State University affirms the principle that all individuals associated with the academic community have a responsibility for establishing, maintaining, and fostering an understanding and appreciation for academic integrity. Academic integrity is the foundation for university success. Learning how to express original ideas, cite works, work independently, and report results accurately and honestly are skills that carry students beyond their academic career. In addition, the Statistics and Probability department adheres to the policies on academic honesty as specified in General Student Regulations 1.0, Protection of Scholarship and Grades; the all-University Policy on Integrity of Scholarship and Grades; and Ordinance 17.00, Examinations. (See Spartan Life: Student Handbook and Resource Guide and/or the MSU Web site.)

Therefore, unless authorized by your instructor, you are expected to complete all course assignments, including homework, projects, and exams, without assistance from any source. You are expected to develop original work for this course; therefore, you may not submit course work you completed for another course to satisfy the requirements for this course. Also, you are not authorized to use the www.allmsu.com Web site to complete any course work in this course. Students who violate MSU academic integrity rules may receive a penalty grade, including a failing grade on the assignment or in the course. Contact your instructor if you are unsure about the appropriateness of your course work. (See also the Academic Integrity webpage.)

Limits to confidentiality

All conversations and course materials submitted for this class are generally considered confidential pursuant to the University's student record policies. However, students should be aware that University employees, including instructors and TAs, may not be able to maintain confidentiality when it conflicts with their responsibility to report certain issues to protect the health and safety of MSU community members and others. As the instructor, I must report the following information to other University offices (including the Department of Police and Public Safety) if you share it with me:

- Suspected child abuse/neglect, even if this maltreatment happened when you were a child,
- Allegations of sexual assault or sexual harassment when they involve MSU students, faculty, or staff, and
- Credible threats of harm to oneself or to others.

These reports may trigger contact from a campus official who will want to talk with you about the incident that you have shared. In almost all cases, it will be your decision whether you wish to speak with that individual. If you would like to talk about these events in a more confidential setting you are encouraged to make an appointment with the MSU Counseling Center.

Inform Your Instructor of Any Accommodations Needed

From the Resource Center for Persons with Disabilities (RCPD): Michigan State University is committed to providing equal opportunity for participation in all programs, services and activities. Requests for accommodations by persons with disabilities may be made by contacting the Resource Center for Persons with Disabilities at 517-884-RCPD or online. Once your eligibility for an accommodation has been determined, you will be issued a Verified Individual Services Accommodation ("VISA") form. Please present this form to the instructor at the start of the term and/or two weeks prior to the accommodation date (test, project, etc.). Requests received after this date may not be honored.

Commercialized Lecture Notes

Commercialization of lecture notes and university-provided course materials is not permitted in this course.

Build Rapport

If you find that you have any trouble keeping up with assignments or other aspects of the course, make sure you let your instructor know as early as possible. As you will find, building rapport and effective relationships are key to becoming an effective professional. Make sure that you are proactive in informing your instructor when difficulties arise during the semester so that we can help you find a solution.

Tentative weekly schedule

DISCLAIMER: This schedule is subject to change. Please check the D2L announcements tab for any amendments before taking action based on this schedule. In general, **Peer Writing**, **RStudio**, and **Webwork** assignments are made available on Monday mornings and are due on Sundays at 11:59 PM EST. **Assessments** are open for 18 hours on Thursdays starting at 6:00 AM EST and ending at 11:59 PM EST. 'NA' means that there is *no* assignment for a particular category due during a given week.

Module 1 (Data Distributions): Week of August 31
Learning objectives: 1. Given a data set, identify observational units (cases) and variables. Classify variables by their type. 2. Load cleaned data sets into RStudio. 3. Define a distribution as an object describing the behavior of a variable. 4. Proficiently describe categorical data distributions using (relative) frequency tables and summarize them using percentages, rates, proportions, and odds.
Peer Writing Prompt: NA
RStudio assignment: NA
Webwork: NA
Assessment: NA

Module 1 (Data Distributions): Week of September 7
Learning objectives: 5. Proficiently describe quantitative data distributions with respect to shape, center, and spread; create and interpret histograms and boxplots, compute and interpret means, medians, standard deviations, and quantile rankings. Create these statistics/plots in RStudio. 6. Compute and interpret standardized scores as a measure of distance an observation falls from the mean. 7. Evaluate relationships between a categorical & quantitative variable with side-by-side boxplots and comparison of numerical summary statistics. Describe and compare quantitative data distributions in RStudio.
Peer Writing Prompt: NA
RStudio assignment: NA
Webwork: HW1
Assessment: NA

Module 1 (Data Distributions): Week of September 14
Learning objectives: 8. Use contingency tables and marginal and conditional distributions within them to investigate relationships between two categorical variables. Create mosaic plots as visual representations of these tables in RStudio. Define and evaluate independence / association between categorical variables. Informally evaluate statistical independence. 9. Create scatterplots in RStudio and compute correlation between two variables. Interpret the correlation as the approximate average of $z_{xi} * z_{yi}$. Informally evaluate statistical independence.

10. When investigating relationships between two variables, identify and describe the effect of a third variable on that relationship (e.g., Simpson's paradox, confounding variables).

Peer Writing Prompt: Analyzing a manufacturing process.

RStudio assignment: Contrasts and relationships.

Webwork: HW 2

Assessment: NA

Module 2 (Intro to inference): Week of September 21

Learning objectives:

11. Construct statistical models of a random process. Define parameters including base prevalence, specificity, and sensitivity. Informally apply Bayes' Theorem to compute conditional probabilities.
12. Define parameters (p, μ, σ) as quantities that govern some aspect of a random process. Define sampling distributions as the distribution of a statistic for all possible samples of the same size from a given random process. Describe them in terms of center and spread.
13. Construct and interpret empirical sampling distributions from a finite population in RStudio through repeated random sampling. Use the empirical standard error as the standard deviation of a statistic.
14. Describe the effect of the sample size on attributes of a statistic's sampling distribution.

Peer Writing Prompt: Responding to your peers.

RStudio assignment: Introductions to sampling distributions.

Webwork: HW 3

Assessment 1: Objectives 1 through 10

Module 2 (Intro to inference): Week of September 28

Learning objectives:

15. Represent random processes with statistical models governed by parameters. Express research questions in terms of competing hypotheses about unknown model parameters and conduct randomization-based tests of these hypotheses. Describe randomization-based null distributions as a type of sampling distribution that constrains model parameters. Use null distributions to compute p-values and draw conclusions regarding a null claim.
16. Interpret p-values as the likelihood of sample results (or results more extreme) computed under a statistical null model. Identify common misinterpretations of p-values.

Peer Writing Prompt: Evaluating COVID tests.

RStudio assignment: Simple randomization-based hypothesis tests.

Webwork: HW 4

Assessment: NA

Module 2 (Intro to inference): Week of October 5

Learning objectives:

17. Apply normal models appropriately. Compute proportions and quantiles of a normal density and compare relative positions within / between data sets using RStudio.
18. Determine when conditions for the CLT are met such that a normal density can be responsibly applied to model for a sampling distribution (i.e., success-failure condition for binomial distributions, population normality / sample size for distribution of sample means, and independence of observations).

Peer Writing Prompt: Responding to your peers.

RStudio assignment: NA

Webwork: HW 5

Assessment: NA

Module 3 (Parametric Inference): Week of October 12

Learning objectives:

19. Conduct parametric hypothesis tests for a single model proportion p by hand or in RStudio. Interpret test statistic, p-value, and estimated effect size.
20. Justify applications of the t-distribution by citing the added variability of standardized statistics associated with estimating σ with s.
21. Conduct parametric hypothesis tests for a single model mean μ by hand or in RStudio. Interpret test statistic, p-value, and estimated effect size.
22. Construct confidence intervals for population proportions & means by hand or in RStudio. Interpret intervals as providing a range of parameter values that are compatible with the collected data (i.e., produce two-tailed p-values ≥1-α). Recognize common misinterpretations of confidence intervals.

Peer Writing Prompt: NA

RStudio assignment: Breaking the Central Limit Theorem

Webwork: HW 6

Assessment 2: Learning objectives 11 - 18

Module 3 (Parametric Inference): Week of October 19

Learning objectives:

23. Explain the relationship between test statistics, p-values, sample sizes, and est. effect sizes. Predict changes in one quantity given changes in others.
24. Explain the relationship between the level of a confidence interval, the sample size, and the width of an interval. Compute required sample sizes necessary to achieve desired margin of errors in applied circumstances.
25. Describe the common practice of using 'dichotomized p-values' and explain the consequences of this practice.
26. When using hypothesis tests for decision-making, identify Type I and Type II errors and their consequences. Compute error rates & power given relevant parameters. Explain the relationship between error rates, power, sample size, and true effect size.

Peer Writing Prompt: Hypothesis testing Wikipedia page.

RStudio assignment: NA

Webwork: HW 7

Assessment: NA

Module 3 (Parametric Inference): Week of October 26

Learning objectives:

27. Conduct hypothesis tests for a difference in independent model proportions $p_1 - p_2$ by hand or in RStudio. Interpret test statistic, p-value, and estimated effect size.
28. Conduct hypothesis tests for a difference in independent model means $\mu_1 - \mu_2$ by hand or in RStudio. Interpret test statistic, p-value, and estimated effect size.
29. Construct confidence intervals for a difference of model proportions or averages. Interpret intervals as providing a range of parameter values that are compatible with collected data. Recognize common misinterpretations of confidence intervals.

Peer Writing Prompt: Responding to your peers.

RStudio assignment: Statistical significance vs. Practical Importance

Webwork: HW 8

Assessment 3: Learning objectives 19 - 26

Module 4 (Principles of study design) Week of November 2

Learning objectives:

30. Recognize instances where two samples contain paired data and apply appropriate paired procedures for μ_d. Explain the consequences of treating paired samples as though they were independent.
31. Given a random process, identify whether causal claims can be investigated and whether observed findings are generalizable to a broader random process. Provide rationale in cases where either is not appropriate that cites issues of biased sampling or confounding variables.

Peer Writing Prompt: The replication crisis.

RStudio assignment: NA

Webwork: NA

Assessment: NA

Module 4 (Principles of study design) Week of November 9

Learning objectives:

32. Express/design a statistical study that includes blocking/randomization diagrammatically.
33. Recognize the problem of pseudoreplication and outline its effect on statistical analyses.
34. Given a context, plan a study that will obtain a desired level of power or margin of error.

Peer Writing Prompt: Responding to your peers.

RStudio assignment: Practice with Inference in RStudio.

Webwork: HW 9

Assessment: NA

Module 5 (Inference for Associations) Week of November 16

Learning objectives:

35. Conduct χ^2 GoF and independence tests by hand or in RStudio. Interpret test statistics, p-values, and estimated effect sizes.
36. Explain factors influencing the results of a χ^2 hypothesis test, including sample size and contingency table dimensions.

Peer Writing Prompt: Research study proposal.

RStudio assignment: Simulating the χ^2 distribution

Webwork: NA

Assessment 4: Learning objectives 27 - 34

Module 5 (Inference for Associations) Week of November 23

Learning objectives:

37. Conduct ANOVA procedures by hand or in RStudio. Interpret test statistics, p-values, and estimated effect sizes (eta-squared).
38. Check assumptions of ANOVA procedure. Explain the effects of violations of these assumptions on the test's performance.
39. After finding sufficient evidence against the null, apply Tukey's HSD procedure in subsequent analyses.

Peer Writing Prompt: Responding to your peers.

RStudio assignment: ANOVA in RStudio.

Webwork: HW 10

Assessment: NA

Module 5 (Inference for Associations) Week of November 30

Learning objectives:

40. Create scatterplots in RStudio and compute correlation between two variables. Interpret the correlation as the approximate average of $z_{xi} * z_{yi}$
41. Compute OLS regression line by hand or in RStudio.
42. Interpret residual diagnostics and justify whether OLS is the optimal model to apply.
43. Evaluate model performance using s/s_y and R^2. Apply variable transformations to improve the performance of a simple linear model.

Peer Writing Prompt: Making predictions.

RStudio assignment: NA

Webwork: NA

Assessment: Learning objectives 35 - 39

Module 5 (Inference for Associations) Week of December 7
Learning objectives: 44. Investigate stability of coefficient estimates as a function of σ_x and address the leverage of outlier values on coefficient estimates. 45. Define extrapolation and explain its influence on predictions. Identify instances where transformations may improve model performance. 46. Conduct simple inferences for β_1. Interpret test statistic, p-value, and confidence interval. Interpret correlation and b_1 as standardized/unstandardized estimated effect sizes.
Peer Writing Prompt: Making predictions (ctd).
RStudio assignment: Outliers & leverage
Webwork: HW 11
Assessment: NA

(EXAM WEEK) Week of December 14
Learning objectives: Final exam week
Peer Writing Prompt: **Responding to your peers.**
RStudio assignment: NA
Webwork: NA
Assessment 6: Learning objectives 40 - 46