SOC 301-01 - Social Statistics Syllabus

Pacific University - Fall 2016

Dr. Chester Ismay

- **Course meeting time**: Mondays and Wednesdays 2:15 PM - 3:50 PM in Walter Annex, Room 101
- **Course webpage**: http://ismayc.github.io/teaching/soc301-f2016
- **Office Hours**
  - Mondays 1:00 PM - 2:00 PM second floor of the Library near the stairs
  - Wednesdays 11:30 AM - 12:30 PM second floor of the Library near the stairs
  - By appointment (Send email to chester@pacificu.edu or talk to me after class.)
- **Office Phone Number**: Not available (Happy to meet in the Library or another place on campus as needed though.)
- Email: chester@pacificu.edu

Overview

Welcome to the Age of Data, where information is all around us, helping us live happier, healthier lives. Or does it? Do we know yet if cell phones cause cancer? Have we come to a decision on whether we should be eating lots of meat or none at all to stay healthy? Despite all of this information, it can be challenging to turn it into the knowledge from which we can make sound decisions. Statistics is the field that aims to bridge this gap between information and knowledge and this course is an application-oriented introduction to modern statistical modeling and inference in the social sciences.

The main focus of this course will be on data visualization as the bridge to statistical and data science topics from a social perspective. We will avoid the traditional, formula-based methodologies of "plug-and-chug" statistics courses and instead use the statistical computing language R to help build intuitions and solve problems in a reproducible way. We will strive to identify the correct problem types and focus on the interpretation of results and plots.

Course Goals

Upon leaving this course you'll be able to answer the following questions:

- What is a p-value and how is it used in social science literature? How does it relate to a confidence interval and why are both of them important topics for the social sciences?
- What makes an effective plot for a given set of variables?
- What goes into making a good statistical analysis? What assumptions are made?
- How do we start a statistical analysis? What questions should we ask?
- How can we create an effective story with data?

We will be using the RStudio Server in nearly every class period and you can expect weekly (and frequently daily) assignments to be due using it. The expectation is that by the end of the course you are able to conduct statistical analyses and produce appropriate plots and do so in a reproducible environment. The RStudio Server is the best way to accomplish these goals.

Student Learning Objectives

Upon completion of this course, students who have been full participants in the course should be able to:

- Create tidy data sets by carefully identifying the observational units and the types of variables that make up the data set.
- Make and interpret the Five Named Graphs: histograms, boxplots, barplots, scatter-plots, and line-graphs.
- Use multivariate thinking by producing and discussing appropriate plots showing the relationships between three (or more) variables.
- Identify the appropriate plot and type of analysis needed to answer a given social research question.
- Describe the process of simulation to build inferential statistics concepts.
- Explain the results of hypothesis testing and confidence intervals in layman's terms.
- Check that assumptions are met in order to perform traditional statistical tests (t-tests, ANOVA, regression, etc.) and be able to provide an alternative randomization-or-bootstrap-type analysis if assumptions are not met.
- Use RStudio and R Markdown to conduct reproducible, well-documented, statistical analyses.
- Discuss how social science research depends on carefully made statistical graphics and analyses.

Course Description from the Catalog

Introductory statistics course for students in the social sciences. The emphasis of the course will be on understanding how social scientists use numerical data to understand social phenomena, and how to use and interpret statistical measures and techniques commonly reported in the social sciences literature. Prerequisite: MATH 122. Does not meet Social Sciences core requirement (2010 catalog). 4 credits.

Textbook

A MODERN DIVE into Data with R (2016) by Chester Ismay and Albert Y. Kim. The online HTML version is preferred but a PDF version is also available. This textbook is free and open-source and will be a key component of the course. You'll frequently be asked to read through the textbook and complete the Learning Checks and Review Questions for homework to check for comprehension.

The textbook is more general in focus than social statistics specifically, but it is intended to be as user-friendly and engaging as possible. The labs and other assignments throughout the course will be more applied to social statistics in particular.

Additional Resources

Getting used to R, RStudio, and R Markdown (2016) by Chester Ismay. This is also a free online resource I put together to help completely new users of R not be quite as intimidated in learning what might initially be scary. We'll refer to this document frequently throughout the course, especially the chapter entitled "Deciphering Common R Errors", which will be updated as the semester goes along.

OpenIntro: Introductory Statistics with Randomization and Simulation (2014), by David M. Diez, Christopher D. Barr, and Mine Çetinkaya-Rundel, available in three formats: pdf, tablet-friendly pdf, and paperback edition. You'll notice a pattern here: the textbook is also free and open-source.

Class components

This course has three components: Problem Sets, Labs/Projects, and Exams/Quizzes.

Problem Sets

Problem sets reinforce the concepts from the reading as well as the material that we address in class. You're encouraged to work on the problem sets with fellow students, though you'll each need to turn in your own work. Problem sets are due at 10 AM the morning of nearly all class meetings and will usually be turned in online via Google Forms or other methods. Late problems sets will not be accepted, so you are encouraged to work on them as soon as possible instead of waiting until the last minute on the day of class.

We will also have frequent checking for understanding using Plickers and other low stakes quizzing strategies in class. This is done to ensure the coverage of topics is at the right level while allowing you the chance to check your own understanding as much as possible in a low stress environment.

Labs/Projects

Almost every week you'll be working with data and organizing your work in an R Markdown document on the RStudio Server and these will be called our labs. Whenever possible I will make the lab instructions available to you on Friday for completion in the following week. An initial completed draft of your work on the lab will be due on **Tuesdays** at 3:00 PM. This will give me time to look over your work and provide you with feedback. In order to receive feedback and a score on the lab, you'll need to have attempted **ALL** problems on the lab. I encourage you to reach out to me on Monday or earlier if you are having troubles with the lab.

You should complete final work on your lab (based on updating according to my suggested revisions) by **Thursdays** at 3:00 PM. As noted in the last paragraph, failure to turn in the draft by 3:00 PM on Tuesdays will lead to a zero score for your lab in that given week. I'll only put the score for your final revised lab into the gradebook, so this is to further encourage you to seek out and respond to my feedback as you learn R. These labs can be worked on in groups of two to three people and you will likely have time in class on both Mondays and Wednesdays to work on them and ask questions.

Remember that you may be asked on Individual Quizzes and the final exam to explain R code or output. You may want to just split a lab up, but this will be to your detriment in my experience. Carefully work through the problems with someone else or at the very least meet to discuss after you have tried ALL of the problems yourself.

You will also have the opportunity to use the RStudio Server to complete a group project with your peers near the end of the semester. This group project will include your writing of a data journalism piece in the style of http://FiveThirtyEight.com. It will include carefully created graphics and clean descriptions to a layperson and will serve as a culminating experience in the course. You'll be given feedback on this project in multiple stages over the course of the semester.

Exams/Quizzes

We'll have several examinations and quizzes throughout the semester in order to challenge your understanding and provide me and you with a sense of where you're at. Each of these assessments will build on previous knowledge and you can think of them as being **cumulative** in much the same way the final exam will be. This is to your benefit. I want to help you learn how to study cumulatively and practice so that final exams aren't as terrifying. I'll provide plenty of tips and we'll practice a lot in class to help you with this. My goal is for you to see how the course all fits together whenever possible.

Some will be more traditional pen and paper but the majority will be done with the computer using the RStudio Server. These will come in the form of both take-home and in-class varieties and will appear in both individual and pair forms.

A cumulative final exam will be given on December 13th from 8:30 to 11 AM.

Grading Criteria

This course will provide you with many opportunities to demonstrate your learning of statistical concepts and your ability to apply them to a variety of social data sets and contexts. You'll frequently have the opportunity to revise your work (especially on labs and problem sets) based on feedback and in doing so I hope to remove as much stress as possible from the assessment aspect of the course for you. I also reserve the right to allow for revisions on exams and quizzes, but you shouldn't expect that with each assessment.

That being said, I do EXPECT hard work and focus from you as students. I want you to aspire to learn and grow as students. The content from a course that is too easy is quickly forgotten at the end of the semester and a course that is too difficult can cause anxiety that interferes with genuine learning opportunities. I strive to find the right balance of challenge and fairness for you.

Your course grade will be determined by the following basis with relative weights given:

- Attendance/Participation/Problem Sets – 10%
- Cumulative Quizzes and Exams – 40%
- Labs/Group Project – 30%
- Final Exam - 20%

More details on the specifics of these weights and a further breakdown will be given as the semester goes along.

I won't assign or give extra credit in this course. There are plenty of opportunities to show me (and yourself) that you are improving and learning the material. Getting behind will get you in trouble quickly though as I've found getting out of a rut academically is far harder than you may think.

Grading System

I strongly encourage you to focus on your own learning of the materials and how they align with the goals for the course I have laid out. You should identify for yourself how well you are understanding the material and be able to assess this frequently. This will allow you to keep the knowledge that you obtain during the course much after the semester is over. I am required to assign you letter grades though and they will be assigned as follows:

A: 90-100%; B: 80-89%; C: 70-79%; D: 60-69%; F: less than 60%

These cut-offs may be lowered at the end of the semester and I reserve the right to give +/- grades as I see fit.

Course Schedule

The references to **Chapters**/**Sections** here correspond to chapters in the **MODERN DIVE into Data with R** book. The **RStudio & R Markdown** references correspond to the **Getting used to R, RStudio, and R Markdown** book.

Weekday	Date	Content	Material Due
Monday	August 29	Introduction (Chapters 1 and 2)	-
Tuesday	August 30	-	Pre-test
Wednesday	August 31	RStudio & R Markdown	PS1
Monday	September 5	Labor Day (No classes)	-
Wednesday	September 7	RStudio & R Markdown	-
Monday	September 12	Tidy data (Sections 3.1 and 3.2)	PS2, Quiz 1
Tuesday	September 13	"	Lab 1 draft due
Wednesday	September 14	Tidy data II (Sections 3.3 to 3.5)	PS3

Weekday	Date	Content	Material Due
Thursday	September 15	"	Lab 1 final due
Monday	September 19	Histograms and Boxplots (Sections 4.1 to 4.3)	PS4
Tuesday	September 20	"	Lab 2 draft due
Wednesday	September 21	Barplots (Section 4.4)	PS5
Thursday	September 22	"	Lab 2 final due
Monday	September 26	Scatter-plots and Line-graphs (Sections 4.5 and 4.6)	PS6
Tuesday	September 27	"	Lab 3 draft due
Wednesday	September 28	Grammar of Graphics (Sections 4.7 and 4.8)	PS7
Thursday	September 29	"	Lab 3 final due
Monday	October 3	Review for Exam 1	PS8, Quiz 2
Tuesday	October 4	"	-
Wednesday	October 5	Take Home Exam 1 (Chapters 1 - 4)	-
Thursday	October 6	"	-
Monday	October 10	Work on Lab 4 in class	Take Home Exam 1 Due at 2
Wednesday	October 12	Five Main Verbs & The Pipe (Sections 5.1 and 5.2)	Lab 4 draft due at 4 PM
Friday	October 14	"	Lab 4 final due at 3 PM
Monday	October 17	Five Main Verbs & The Pipe (Sections 5.1 and 5.2) (Part 2)	PS9
Wednesday	October 19	The Pipe and Joins (Sections 5.2 and 5.3)	PS10
Monday	October 24	Random sampling and Simulation (Sections 6.1 and 6.2)	Quiz 3, PS11
Tuesday	October 25	"	Lab 5 draft due
Wednesday	October 26	Bootstrapping (Section 6.3)	-
Thursday	October 27	"	Lab 5 final due
Monday	October 31	Review for Exam 2 (Chapters 1-6)	-
Tuesday	November 1	"	-
Wednesday	November 2	Exam 2 in class (Chapters 1-6)	-
Thursday	November 3	"	Midterm course eval
Monday	November 7	Hypothesis Testing (Chapter 7)	PS12
Tuesday	November 8	"	Lab 6 draft due
Wednesday	November 9	Hypothesis Testing II (Chapter 7)	PS13
Thursday	November 10	"	Lab 6 final due
Monday	November 14	Confidence Intervals (Chapter 8)	PS14
Tuesday	November 15	"	Lab 7 draft due
Wednesday	November 16	Correlation/Simple Regression (Sections 9.1 and 9.2)	Quiz 4, PS15
Thursday	November 17	"	Lab 7 final due
Monday	November 21	Multiple Regression (Section 9.3)	PS16
Tuesday	November 22	"	-
Wednesday	November 23	Thanksgiving Holiday (No Classes)	-
Thursday	November 24	Thanksgiving Holiday (No Classes)	-
Monday	November 28	Multiple Regression II (Section 9.3)	-
Tuesday	November 29	"	Draft of Final Project Due
Wednesday	November 30	Multiple Regression III (Sections 9.4 to 9.6)	Quiz 5, PS 17
Thursday	December 1	Work on Final Project	-
Monday	December 5	Course Review (Chapter 10)	Final Project, PS 18
Tuesday	December 6	-	Post-test
Wednesday	December 7	READING DAY	-
Tuesday	December 13	Final Exam	8:30 AM to 11:00 AM

What to Expect in Class

In nearly all classes, I will devote the first five to ten minutes of class and the last five minutes of class to activities designed to help you retain the material you have just encountered or to reflect on previous content. It is imperative that you value these opportunities and your attendance and participation during them will help you succeed in the course.

We'll be working through problems together in class and discussing content from your reading assignments. This class may be different from other classes you take or your experiences in classrooms from the past. I prefer to think of myself as your guide on the side instead of the sage on the stage. I'm here to assist you as needed, but I don't want to be the passer of all knowledge to you. In other words, I'm hoping to create an environment where you develop understanding based on each of your own needs rather than just dictating how and what to think to you.

You'll be peer reviewing other students' work; you'll be solving problems on the board or using RStudio; you'll be actively working to learn about social statistics inside the classroom. I will lecture from time to time, but this will not be a class where you can passively sit back and copy everything from the board. You'll need to play along in RStudio, to answer and ask questions during class, and to work well with your classmates to solve problems.

Most importantly, I want the classes to be fun and I want you to value what we cover. It's my job to make it interesting, it's your job to put in the effort and have a growth mindset about statistics. You'll be amazed at the end of the semester just how much knowledge you've gained and how your hard work has paid off.

Course attendance policy

Attendance will be collected and active participation in in-class and out-of-class activities is expected. Three absences of any sort are allowed. That being said I understand all the demands that are on you as students. Keep me posted on what is going on if you must miss class. These three absences can serve any purpose you like.

If you have a medical emergency or another valid institutional reason for missing class, we can set up a plan together. You'll find that missing classes hurts your understanding of the material and map of how the whole course fits together. Each additional class missed will result in a 10% grade reduction. I take TIMELY attendance very seriously.

Makeup quizzes

Makeup quizzes and exams are rarely given as I usually drop the lowest score at the end of the semester to help account for this. It isn't fair to the other students in the class to take an assessment at a later date and it hurts your flow in the class because you'll be playing catch-up with other material.

What to Expect Outside of Class

This course **requires** you to read the textbook and to check for understanding as you learn new material and review past material. You'll be working on Learning Checks, Review Questions, and doing statistical analysis and making graphics in RStudio. This will be time-consuming, but it is my goal to separate these assignments into smaller pieces. This will help ensure that you aren't procrastinating until the last minute on a big assignment and also that you are using the new knowledge gained as much as possible just when you learn it.

Tips for success:

1. Spend 15 minutes every day (even on the weekend if possible) quizzing yourself on concepts from the course. I recommend you create flash cards as the course progresses and ask yourself to provide the definition when given a term or to explain what particular R code produces. The process of forcing yourself to think about the concept even if you can't get the answer exactly right will help you learn.

2. Don't just read over your notes or the book multiple times. One of the most common things I hear from students that didn't perform well on a quiz or assignment is that they studied the material a lot by reading over the book many times. This doesn't help you learn the content. It creates a false sense of understanding by creating an association with the words. YOU HAVE TO QUIZ YOURSELF if you want to learn and succeed.

3. Create your own quizzes. One of the best strategies you can do if you have had trouble with exams and quizzes in the past is to identify the key ideas from the class and create your own problems to solve in an environment similar to the one in which you take quizzes and exams.

4. Make mind maps that relate all of the content throughout the course together. By creating a diagram that connects concepts together you'll see how everything is related and that will help you create the story of the course instead of seeing concepts as disjoint pieces.

5. Study with someone else especially with the R material. But DON'T divide the lab in half and not talk carefully about the questions together. You'll need to practice with R nearly every day if you want to learn how to use it. It's going to be like learning a foreign language at the beginning. You'll need to learn the syntax and practice "speaking it" as much as possible.

6. Ask questions of me and of your classmates. I'm always open to answering questions and you should never feel like your question isn't appropriate. You should also look for opportunities to help others and teach them the content. This will help you as a learner as well.

7. Enjoy the learning process! Failure to not understand everything all at once is to be expected. Our brains need to struggle to learn. You can't build muscles by sitting on the couch and you can't learn statistics by not struggling a little bit. Devote some time every day to thinking about a more difficult problem than what is given to you. How would you solve it? What's more difficult about it? Evaluate your learning and find out what works best for you.

Additional Information

Academic Misconduct Policy

Pacific University has no tolerance for academic misconduct/cheating. It is university policy that all acts of misconduct and dishonesty be reported to the Associate Dean for Student Academic Affairs. Sanctions that may be imposed for such misconduct range from an "F" for the assignment, an "F" for the course, and suspension or dismissal from the university. Forms of academic misconduct include but are not limited to plagiarism, fabrication, cheating, tampering with grades, forging signatures, and using electronic information resources in violation of acceptable use policies.

Learning Support Services for Students with Disabilities

If you have documented challenges that will impede your learning in any way, please contact our LSS office in Clark Hall (ext.2717; lss@pacificu.edu). LSS staff will meet with students, review the documentation of their disabilities, and discuss the services that Pacific offers and any appropriate ADA accommodations for specific courses.

Tutoring and Learning Center (TLC)

The TLC is located in Scott Hall, 1st-floor. The center focuses on delivering one-on-one and group tutoring services for foreign languages, math and science courses and writing skills in all subjects. Students should consult with the center's director for information on tutoring available for other subjects. Day and evening hours; walk-ins welcome!

Unauthorized Recordings

Students are prohibited from making audio and/or visual recordings of lectures or presentations without prior consent of the instructor or presenter.