



PM 511a: Data Analysis

Units:	4
Term:	Spring 2021
Lecture:	Friday 1:05 pm – 2:55 pm, Zoom (link in Bb)
Lab:	Friday 3:00 – 4:00 pm , Zoom (links TBA)
Instructor:	Farzana Choudhury, Ph.D. email: fchoudhu@usc.edu https://usc.zoom.us/my/farzana.choudhury
Student Hours:	Tuesday 5.30-6.00 pm
Teaching Assistants:	Anthony Colombo; acolombo@usc.edu Felice Lin; flin@usc.edu

Course Description: This course will introduce rigorous methods of applied statistical data analysis, with a focus on linear regression models for quantitative (continuous) outcomes using SAS.

Prerequisite(s): PM510; open to other students by consent of instructor

Learning Objectives

- Read and interpret computer output from a commonly used statistical package (SAS).
- Employ SAS to perform descriptive and inferential data analysis for continuous outcomes.
- Evaluate the assumptions of linear regressions.
- Interpret results for analyses with continuous outcomes.
- Write the statistical component of a “Methods” and “Results” section for a scientific paper.
- Apply data analysis techniques to real-world data and write and interpret the results as if for a journal article

Required Texts:

- Kleinbaum, D. G., Kupper, L. L., Nizam, A., & Rosenberg, E. S. (2014). *Applied Regression Analysis and Other Multivariable Methods*. (5th ed), Cengage Learning.
- Delwiche & Slaughter (2019) *The little SAS Book: A Primer* (6th ed), SAS Institute. ISBN-13: 978-1642952834
- Ottesen, Delwiche & Slaughter (2020) Exercises and Projects for *The little SAS Book* 6th ed, SAS Institute. ISBN-13: 978-1629596556

Technological Proficiency and Hardware/Software Required

It is mandatory that you have access to a computer and reliable internet. SAS will be used extensively in this course and will be accessed via the internet through a virtual desktop. You must also have a USC activated Zoom account with the Zoom client installed on your computer in order to attend the lecture and lab sessions.

Description and Assessment of Assignments

Homework: There will be 7 homework assignments. Homework must be submitted in electronic form on Blackboard by 11:59pm on the due date. Please allow yourself enough time to complete the homework assignments (~10-20 hours per assignment).

Homework Quizzes: With the exception of homework 1, every homework will also include a homework quiz that consists of questions from the homework, plus additional conceptual questions. The quiz is timed, and you only get one attempt. The quiz is administered through Blackboard. It is due at the same time as the homework. You should complete the homework assignment before attempting the homework quiz, as quiz questions will reference your homework assignment answers.

Lab exercise: Each week, you are expected to watch/read all lab materials and attempt the lab exercise before attending the lab live session. Lab exercises are graded on completeness and submitted in electronic form on Blackboard by 11:59pm on the Thursday after they are assigned.

In-Class Quizzes: There will be two 30 minute "in-class" quizzes. These multiple choice/short answer quizzes are closed book (no calculator) and will begin promptly at the start of class.

Midterm Exam: The midterm exam will be 1.05 pm-2.55 pm (usual lecture time) on week 11. It will be closed book and will cover material through week 9 of the course. There will be no lecture/lab that week.

Final Project: The final project will be a take-home data analysis project. This will be made available during the week 13 and is due at the scheduled final exam period for this class, to be submitted on Blackboard. This project will cover topics from the entire course. This project should be treated like an exam. You may not discuss it with anyone.

Participation: Attendance at all live sessions is expected, though not required. Extra credit (applied as a maximum of 5 points to your Midterm grade) will be given for participation in the live sessions. If you cannot attend the live sessions, they will be recorded. If you would still like to receive extra credit you may write up a thoughtful response/synopsis – something that indicates that you watched the recording of the live session in its entirety. This discussion board post must be made in the same week as the missed session (by the following Wednesday at midnight). In order to encourage active engagement and support our classroom community, students are encouraged to keep their cameras on for the duration of the class session. Virtual backgrounds can be used to remove any distractions that may be present in your physical environment from the view of your peers. If you have any internet connectivity, illness, are in a time zone that makes attendance unreasonable, or other concerns that would impede your ability to have your camera on, please contact the professor directly.

Accessing datasets for lab exercises or HW

The datasets used in your Lab exercises and Homework assignments are available in Blackboard. To find these, navigate to the “Datasets” section of the course, where you will find two folders, one titled “Lab Datasets” and one titled “Homework Datasets”. These folders contain all of the datasets you will need for the entire semester. Please note that, while Lab Datasets includes a single list that applies to all students in the course, Homework Datasets are assigned individually per student. You will be assigned a student number at the beginning of the course, and every homework dataset will be

specific to that student number. You will use this student number on some of your homework quizzes, so please make note of it.

Accessing SAS via the Virtual Desktop

****Please have the virtual desktop set up prior to the first class session.**

In this course you will access SAS by logging in to a virtual computer that has SAS installed on it. Cloudapps, also referred to as a virtual desktop interface (VDI), is a service provided by ITS that allows current USC faculty, staff, and students to access USC-licensed software packages from any computer or mobile device. In order to access this virtual computer, you will need to download VMWare Horizon client. NOTE that there is an option to access the virtual desktop via a web browser, but, we do not allow that in this course. There are step-by-step instructions on how to install this software, including screenshots, at <http://itservices.usc.edu/vdi>. You will log into this software using your USC NetID and password (the same login you use for Blackboard).

If you have any issues with Cloudapps either while installing, or throughout the semester, you can call the ITS Help Desk, 24/7, at 213-740-5555.

Assignment Submission Policy

No late assignments or final projects will be accepted. For the written portion of the homework, students may discuss problems and strategies with one another, but must turn in individual write ups. The homework quiz must be completed individually (no discussion with anyone else). For the final project, all work must be your own (no discussing problems or strategies with anyone else, including TAs or tutors). This course has a zero-tolerance policy for cheating or copying of any kind. *Any evidence of shared written work, programming code, or other form of cheating on the homework or final project will result in zero credit for all students involved, submission of the incident to the Academic Integrity Coordinator for the Keck School of Medicine for adjudication and, likely, an F in the course for all students involved.*

Grading Breakdown

Assignment	% of Grade
Homework (7)	20%
Labs (9)	15%
Midterm Exam	25%
In-Class Quizzes (2)	5%
Final Project	35%
TOTAL	100%

Grading Scale

Letter Grade	Overall Percent
A	93.0 – 100%
A-	90.0 – 92.9%
B+	87.0 – 89.9%
B	83.0 – 86.9%
B-	80.0 – 82.9%
C+	77.0 – 79.9%
C	73.0 – 76.9%
C-	70.0 – 72.9%
D+	67.0 – 69.9%
D	63.0 – 66.9%
D-	60.0 – 62.9%
F	< 60%

Approximate Course Schedule: A Weekly Breakdown*

	Lecture Topics	Lab Topics	Due Dates
Week 1 January 15	Introduction to SAS statistical Software	No Lab	HW 1 due January 21
Week 2 January 22	Using SAS to Manipulate and Analyze Your Data		
Week 3 January 29	Review of Basic Statistics	Lab 1: Manipulating datasets in SAS (set/merge)	HW 2 & Lab 1 due February 4
Week 4 February 5	Linear Regression Analysis, Part I, Assumptions, Fitting, Interpretation, Inference	Lab 2: SAS Graphics	Lab 2 Due February 11
Week 5 February 12	Linear regression Analysis, Part II, Prediction, Assessing Assumptions, Transformations	Lab 3: Linear Regression	HW 3 & Lab 3 due February 18
Week 6 February 19	In-Class Quiz 1 Regression with categorical predictor variables, splines	Lab 4: Sample size and power	Lab 4 Due February 25
Week 7 February 26	Multiple linear regression, Part I	Lab 5: SAS macros, Quiz review	HW 4 & Lab 5 due March 4
Week 8 March 5	Multiple linear regression, Part II, Scaling, Multiple/Partial Correlations	Lab 6: Multiple regression, Correlation	Lab 6 due March 11
03/05/2021: Last Day to drop without "W"			
Week 9 March 12	Wellness day (No class)		
Week 10 March 19	Confounding and interaction	Lab 7: Assessing confounding and interaction in SAS, Midterm review	HW 5 & Lab 7 due March 25
Week 11 March 26	Midterm	No Lab	
Week 12 April 2	Multiple regression residual diagnostics, collinearity	Lab 8: Review midterm, model selection	Lab 8 Due April 8
Week 13 April 9	Model selection	No Lab	HW 6 Due April 15
04/09/2021: Last day to drop with "W"			
Week 14 April 16	Analysis of variance (ANOVA)	Lab 9: ANOVA	HW 7 & Lab 9 due April 22
Week 15 April 23	In-Class Quiz 2 Final Project Assigned	No Lab	
Week 16 April 30	Wellness day (No class)		Final Project Due May 6, 11:59pm

*This might change slightly, if necessary

Statement on Academic Conduct and Support Systems

Academic Conduct:

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, “Behavior Violating University Standards” policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, policy.usc.edu/scientific-misconduct.

Support Systems:

Student Health Counseling Services - (213) 740-7711 – 24/7 on call

engemannshc.usc.edu/counseling

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call

suicidepreventionlifeline.org

Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-4900 – 24/7 on call

engemannshc.usc.edu/rsvp

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED) | Title IX - (213) 740-5086

equity.usc.edu, titleix.usc.edu

Information about how to get help or help a survivor of harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants. The university prohibits discrimination or harassment based on the following protected characteristics: race, color, national origin, ancestry, religion, sex, gender, gender identity, gender expression, sexual orientation, age, physical disability, medical condition, mental disability, marital status, pregnancy, veteran status, genetic information, and any other characteristic which may be specified in applicable laws and governmental regulations.

Bias Assessment Response and Support - (213) 740-2421

studentaffairs.usc.edu/bias-assessment-response-support

Avenue to report incidents of bias, hate crimes, and microaggressions for appropriate investigation and response.

USC Support and Advocacy - (213) 821-4710

studentaffairs.usc.edu/ssa

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

The Office of Disability Services and Programs - (213) 740-0776

dsp.usc.edu

Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

Diversity at USC - (213) 740-2101

diversity.usc.edu

Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call

dps.usc.edu, emergency.usc.edu

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call

dps.usc.edu

Non-emergency assistance or information.

MPH Competencies

Students who complete this course will meet the following MPH competencies for the Biostatistics/Epidemiology Track:

- Advise fellow investigators on the design, conduct and data analysis for studies in the health sciences.
- Serve as data coordinators, managers and analysts for epidemiologic studies or public health-related projects, *e.g.*, multi-center clinical trials, government surveys, outcome databases, etc.
- Determine the appropriate study design to analyze a community health problem.
- Locate and access existing sources of data including vital statistics records, disease registries, and clinic and hospital records that might bring light to the problem.
- Analyze basic relationships between risk factors and outcome data using data management and statistical software.
- Seek additional biomedical, statistical, and computing assistance when appropriate.
- Interpret the results of research reports that focus on public health and/or policy implications.
- Identify potential bias in research reports and evaluate the likelihood that these potential biases actually explain the findings.
- Write the results of data analysis in a coherent report that can be used to guide public policy.
- Identify behavioral/social/cultural and epidemiological factors in a particular setting/problem and analyze how these factors affect disease.