Advanced Statistical Inference

PSYC UN1660

- **Course Location:** Schermerhorn 200B
- **Course Time:** Tuesday/Thursday 1:10 PM - 2:25 PM

- **Instructor:** Greg Jensen
- **Office:** Schermerhorn 500
- **Office hours:** 3:00pm – 4:30pm Wed., 11:00am – 12:30pm Thu.
- **Email:** ggj2102@columbia.edu

- **Teaching Assistant:** Raymond Crookes
- **Office:** Schermerhorn 419
- **Office hours:** 10:00am – 12:00pm Tue.
- **Email:** raymond.crookes@columbia.edu

Prerequisites

Instructor's permission **and** PSYC UN1610 (or the equivalent). Students should be comfortable with algebra and the basic concepts in classical statistical methods.

Notes on Prerequisites: At a minimum, students should be comfortable with algebra (including the connection between logarithms and exponents) and the basic concepts in classical statistical methods (e.g. elementary probability, μ, σ, z, t, linear regression, and ANOVA). Additionally, familiarity with such topics as probability theory, information theory, Bayes' rule, and calculus will each facilitate understanding of the material. Note that a high school AP Statistics course, as presently defined by The College Board, does *not* cover enough material to qualify as a prerequisite. Past experience with computer programming is not required, but will make it much easier to get started.

Course Description

This course outlines elements of statistical inference. Students will receive training in the use of software to evaluate both experimental data and psychological theory. In doing so, students will construct models that can both describe scientific results and also predict future outcomes.

Course Overview

This course introduces students to Bayesian inference and model construction. Building on a foundation of basic probability theory, we will begin with *maximum a posteriori* parameter estimation and proceed to briefly overview multivariate linear models. The course will then transition to more advanced topics: Model selection using information criteria, Markov chain Monte Carlo techniques, and nonlinear regression. We will conclude with an introduction to multi-level models.

These techniques, presently at the cutting edge of empirical psychological research, equip aspiring scientists with the practical skills they will need to grapple with experimental data. To this end, coursework will

primarily involve hands-on manipulation of real data from published studies, in order to relate abstract statistical concepts with concrete experimental scenarios. Rather than working with equations, students will undertake analyses using computer programming. Please read this syllabus carefully to make sure you haven't missed anything important.

Course Role in the Academic Curriculum: Statistical inference is a major activity of research scientists, in psychology and otherwise. The objective of the course is both to deepen students' understanding of how data can inform theory and vice versa, as well as giving them practical skills that are in high demand. Psychology majors, as well as students studying psychology in the Post-Baccalaureate Certificate Program, may count this course toward the statistics requirement or as an elective. Psychology concentrators may count this course as an elective. Neuroscience & Behavior majors may use this course to satisfy the P3 research methods/statistics requirement.

Course Requirements

Textbook: *Statistical Rethinking* (Required) by Richard McElreath (2015). Each lecture is matched to a corresponding chapter of the textbook. When a chapter is listed more than once, you are expected to *reread* the chapter each time. This material is very challenging, and you should not expect to understand all of it on your first read-through.

***Using R* (Provided on Courseworks)** by J. H. Maindonald. This electronic guide explains a wide variety of R's basic operations. Only chapter 2 ("An Overview of R") is required to get started.

Software: *R* (Required), available from The R Project for Statistical Computing. We will also use *RStudio* to make R a little easier to work with. You **must** use R to complete homework assignments and projects. Both R and RStudio are free, open source, and available for a wide variety of computing platforms.

Attendance: Attendance and class participation are expected. Because we will be grappling both with abstruse concepts and with messy data, an understanding of the material can only be achieved by asking questions when things are unclear.

Homework: There will be seven homework assignments, each worth 4% of your final grade. These will consist of a mix of self-directed tutorials, short-answer questions, and practice problems. Please note the following:

- **Think of these as "homework labs."** Lectures are not sufficient or satisfactory for learning to *do* statistics. As such, the homework assigments are not just practice problems, but are instead a core part of your statistical training. Do not put off completing them or you will very likely fall behind in the course.
- **Collaboration on homework assignments is permitted and encouraged.** Students may work together on homework assignments, but each student must turn in a *separate* assignment, written in their own words. Copying and/or cosmetic editing of another student's assignment will be considered plagiarism, and thus incur disciplinary action.
- **Turn in homework as text files.** Homework is provided in the form of "R Markdown" scripts (with a *.Rmd file extension), which can be opened in RStudio and modified to produce a final output. Turning in homework assignments as plaintext (*.txt) or Markdown (*.md) files is OK, but it should be straightforward for the grader to run your scripts without modification to get the answers you report. Assignments turned in using other formats (e.g. as Word documents) will be penalized.
- **Show all work for every problem.** It is assumed that you are using RStudio to get your answers. Consequently, every answer should include the *complete* script used to generate both the results of your analysis and the corresponding figures.
- **Homework must be uploaded to Courseworks by 5:30 on the day it is due.** Late assignments will be penalized by 0.4% (out of 4%) for each day late.

Projects: There will be six "projects," each worth between 10% and 15% of your final grade. These are more involved than homework assignments, and will require applying the tools presented in the course to analyze provided data and interpret the results.

- **Projects are *not* collaborative: Do not work with other students.** Although closer in spirit to a writing research paper, your work on each project should be treated as though it were an exam. Sharing your work on a project with another student is grounds for immediate referral to the Dean's Discipline process and possible subsequent sanction. *However,* questions directed to the instructor or TA are OK.
- **Take document preparation seriously.** As with homework, turn projects in as text files and take advantage of R's scripting to show all your work.
- **Projects must be uploaded to Courseworks by 5:30 on the day it is due.** Late projects will be penalized by 2% a day for each day late.

Extra Credit: Many concepts and procedures in this course are hard, even when taught at an introductory level. Consequently, students who devote additional time to practicing with the material deserve to be recognized for their efforts.

- **Students may complete problem sets from the textbook for credit.** Chapters 2 through 13 each have three sets of practice questions: Easy, Medium, and Hard. If a student completes the set of Easy practice questions in a chapter, they may earn up to 0.25% toward their final grade, applied to their final grade. Each set of Medium practice questions are worth up to 0.5%. Thus, across 12 chapters, the maximum possible extra credit available is 9% ($12 \times 0.25\% + 12 \times 0.5\%$). The Hard practice questions are not eligible for extra credit.
- **Extra credit is due in three blocks.** If a batch of extra credit questions is turned in later than the deadline listed on the schedule, its contribution to the final grade is halved. Students are welcome to submit any problem set at any time; the half-off penalty is intended to focus attention on the material that the class is currently dealing with, and also to help spread out the grading load.
- **Treat problem sets like homework when preparing them.** Show all work and submit the resulting scripts as text files. Each problem set should be turned in as a separate file, with your name, the chapter number, and the question difficulty clearly labeled.
- **All extra credit must be turned in by no later than December 8th.** As a courtesy, if you anticipate that you will take advantage of this option, please try to complete sets of problems for credit throughout the semester, rather than trying to power through them all at the last minute.
- **The maximum your final grade can reach with extra credit is 90%.** If your total points on homework and projects reach or exceed 90%, then you are not eligible for extra credit.

Students With Disabilities

Students with disabilities registered for this course and who require classroom accommodations should get in touch with me as soon as possible. Additionally, stop by the Office of Disability Services (ODS), located in Wien Hall, Suite 108A, to register for support services, if you have not already done so.

Grading: Final grades will depend on completed assignments, according to the following breakdown.

- **Homework:** 28% (4% each)
- **Project 1:** 10%
- **Project 2:** 10%
- **Project 3:** 12%
- **Project 4:** 12%
- **Project 5:** 13%
- **Project 6:** 15%

Letter grades are subsequently assigned according to the following criteria:

Bound		Grade		Bound
	$>$	A+	$\geq$	97
97	$>$	A	$\geq$	93
93	$>$	A-	$\geq$	90
90	$>$	B+	$\geq$	87
87	$>$	B	$\geq$	83
83	$>$	B-	$\geq$	80
80	$>$	C+	$\geq$	77
77	$>$	C	$\geq$	73
73	$>$	C-	$\geq$	70
70	$>$	D	$\geq$	60
60	$>$	F		

Academic Integrity

Students are expected to do their own work. Getting someone else to do your work for you, or passing someone else's work off as your own, is a breach of academic integrity and is a serious offense in academic circles. Whether you're getting help from another student, or copying text from the Internet, anyone presenting the work as their own without attribution will be referred to Dean's Discipline, with appropriate action to follow.

Details regarding Columbia's definition of academic dishonesty (including examples), as well as the consequences, are available in **Columbia's Guide to Academic Integrity**. Inform yourself about what the university's policies on cheating and plagiarism are, because "I didn't know!" is not an acceptable defense.

If you are falling behind on the material, or otherwise feeling mounting pressure from the class, it is *always* better to contact me or your TA, rather than resorting to academic dishonesty. We're here to help, and we're enthusiastic about doing so.

Schedule of Classes

Date	#	Topic	Reading	Homework Due
09/06/2016	1	What Are Statistics For?	Ch. 1	
09/08/2016	2	Statistics As Programming	Maindonald Ch. 2	
09/13/2016	3	Probability As Uncertainty	Ch. 2	Homework 1
09/15/2016	4	Bayesian Updating	Ch. 2	
09/20/2016	5	Maximum A Posteriori	Ch. 3	Homework 2
09/22/2016	6	Parameter Space	Ch. 3	
09/27/2016	7	Linear Models	Ch. 4	Project 1
09/29/2016	8	Simulation	Ch. 4	
10/04/2016	9	Multivariate Models	Ch. 5	Homework 3
10/06/2016	10	Masking Effects	Ch. 5	
10/11/2016	11	Categorical Predictors	Ch. 5	Project 2
10/13/2016	12	Information Theory	Ch. 6	Extra Credit (Ch. 2 to 5)
10/18/2016	13	Overfitting & Regularization	Ch. 6	Homework 4
10/20/2016	14	Information Criteria	Ch. 6	
10/25/2016	15	Interactions	Ch. 7	Project 3
10/27/2016	16	Markov Chain Monte Carlo	Ch. 8	
11/01/2016	17	MCMC Unchained	Ch. 8	Homework 5
11/03/2016	18	Solving Impossible Problems	Ch. 8	
11/08/2016		**NO CLASS**		
11/10/2016	19	Generalized Linear Models	Ch. 9	Project 4
11/15/2016	20	Logistic Regression	Ch. 10	Extra Credit (Ch. 6 to 9)
11/17/2016	21	Models of Counts	Ch. 10	
11/22/2016	22	Models for Likert Scales	Ch. 11	Homework 6
11/24/2016		**NO CLASS**		
11/29/2016	23	Multi-Level Models	Ch. 12	Project 5
12/01/2016	24	Shrinkage	Ch. 12	
12/06/2016	25	Multi-Level Regression	Ch. 13	Homework 7
12/08/2016	26	Unbounded Horizons	Ch. 15	Extra Credit (Ch. 10 to 13)
12/20/2016				Project 6

(Syllabus subject to revision)