University of Southern California
Sol Price School of Public Policy
International Public Policy and Management Program
(IPPAM)
Fall 2018, 4 units

PPD 570: Applied Statistics for Planning, Policy, and Management

Thursdays, 6:00 PM – 9:20 PM, VKC 260
Near the IPPAM Office, August 23 through November 29

Computing Lab   Most Fridays, 10:00 – 11:50, RGL 209 or 219
September 7 through November 16

Prof. James Moore [jmoore@usc.edu]
Andrus Center (GER) 203
Los Angeles, CA 90089-0193

Phone: **(**213) 740-0595   Cell: (213) 663-8146

Office Hours: Wednesday 4:30-6:00 PM (prior to ISE 495), Thursday 4:00-5:30 PM (prior to PPD 570), or by appointment. We will pick an additional hour together in class.

Computing Lab Instructor:
Dr. Michael (Cheng-Yi) Lin [chengyil@usc.edu]

Phone: (626) 213-4781

Michael is a Price School PhD and full time staff member with the Milken Institute in Santa Monica. We are fortunate to have his help.

Office Hours: 11:50-12:30 PM (following lab) or by appointment

Graders:
Mr. Qiao Wu [qiaowu@usc.edu, left],
Ms. Phoebe He [whe248@usc.edu].

Mr. Wu has completed PPD 557 and more advanced courses in multivariate statistics and econometrics. Phoebe was the course assistant for PPD 570 just this past summer. Both are highly qualified. We are fortunate to have their assistance.

USC Catalogue Course Description: *Use of statistical reasoning to answer questions related to public policy and management. Students will review and understand selected statistical techniques for analyzing data and for addressing public policy and management questions of interest using applied data analysis.*

sta•tis´•tics *n. Abbr.* **stat. 1**. The mathematics of the collection, organization, and interpretation of numerical data; especially the analysis of population characteristics by inference from sampling. Used with a singular verb. **2**. A collection of numerical data. Used with a plural verb. [German *Statistik,* originally "political science dealing with state affairs," from New Latin *statisticus,* of state affairs, from Latin *status*, manner of standing, position, state. See **sta-**]

sta•tis´•tic *n.* **1**. Any numerical datum. **2**. An estimate of a parameter, as of the population mean, variance, or skew, obtained from a sample. [Back-formation from STATISTICS.]

Usage in **PPD 570**: The most important characteristic of a statistic is that it be a well-behaved random variable. If the statistic is not well enough behaved to have a known probability distribution, then there is usually no reason to be interested in it.

da´•ta *pl. n. Singular* **datum**. **1**. Information, especially information organized for analysis or used as the basis for a decision. **2**. Numerical information in a form suitable for processing by computer. [Latin, plural of DATUM]
   Usage: *Data* is now used both as a plural and as a singular collection: *These data are inconclusive. This data is inconclusive.* The plural construction is the more appropriate in formal usage. The singular is acceptable to 50 percent of the Usage panel.

Usage in **PPD 570**: Find a source of online data.

sto•chas´•tic *adj.* **1**. Of, denoting, or characterized by conjecture; conjectural. **2**. *Statistics.* **a**. Random. **b**. Statistical. [Greek *Stockhastikas,* capable of aiming, conjectural, from *stokhazesthai,* to aim at, guess at, from *stokhos,* target, aim. See **stegh-**]

Usage in **PPD 570**: Realizations of stochastic variables take on random values distributed across some range. Some stochastic variables are vectors. Realizations of deterministic variables are nonrandom, possibly unknown, but by definition nonrandom.

em•pir'´•i•cal *adj.* **1**. Relying upon or derived from observation or experiment: *empirical methods, an empirical conclusion.* **2**. Guided by practical experience and not theory, especially in medicine. **em•pir´•i•cal•ly** *adv.*

Usage in **PPD 570**: In an experimental context, we use systematic empiricism to test existing theories and construct new ones. In a correlation context, we still rely on empiricism, but in a less controlled way.

em•pir•ic (em-pîr'ik, im-) n. **1**. One who believes that practical experience is the sole source of knowledge. **2**. A charlatan: *"we must not … prostitute our past-cure malady / To empirics'"* (Shakespeare). [Latin *empiricus*, from Greek *emperikos*, from *empeira*, experience, from *empeiros*, experienced in **:** *en-*, in + peira, experiment, trial.

Usage in **PPD 570**: There isn't one. Nobody wants to be called an "empiric."

The point of statistics in applied social science is not to become more theoretical. The objective is to bring theory down to earth, and learn how to reconcile differences between what we believe with what we have taken the time to systematically observe.

Overview

This is a first course in statistics for new graduate students with no substantive prior exposure to the field, or for students who want to review the fundamentals. The course is designed and delivered especially for USC Price School IPPAM students. No prior work in statistics is assumed, but students should be capable of performing mathematics at the level of 11th grade algebra. We will cover the fundamentals of probability and statistics without (much) compromise, and then treat more ambitious topics in a survey fashion.

The lectures for this class will necessarily emphasize basic theory and procedures. There will be numerous applied examples. These applied examples are called "home work."

The lecture meetings will summarize and organize the ideas put forth in the readings. We have only a very limited amount of time to cover this quantity of material, and routine attendance is strongly encouraged. As a matter of courtesy to all parties concerned, please arrive on time, and discuss any premature departures with me prior to the event. If you come late, please be careful not to slam the door as you enter. If you do skip a class, the cost of mitigation is yours, not mine.

I will communicate with you at your USC NetID, which is also your USC email address. You are accountable for the information content of the messages I send to you. I will also post key messages as announcements on the course blackboard website, which you also access with your USCNetID. My email address is above.

Objectives

This class is designed to provide you with:

1. a basic understanding of probabilistic and statistical concepts, with an emphasis on probability;

2. an ability to reason in probabilistic terms;

3. a set of accepted techniques that can be used to analyze, understand, and (hopefully) address many public policy and management problems and related research questions; and the means to acquire new skills in this dimension as needed;

4. an understanding of how to ask statistical questions, and how to treat the information needed to answer these questions or offered in response to these questions;

5. a basic familiarity with statistical computing standards, courtesy of Dr. Lin; and

6. an understanding of why statistical analysis is a key element of your applied social science, graduate education, regardless of whether your objectives are further scholarship or professional practice.

Lectures:

Lectures are offered weekly each Thursday evening during the fall semester. While this is a lecture-based course, questions and informed discussions that result have an important role. Because this is a survey class treating a wide range of techniques, not all topics can be treated in the detail they merit. Questions concerning clarifications, extensions, and applications are welcome and always encouraged, but class members may sometimes find themselves asking reasonable, relevant questions that I will not take the time to answer during lecture. These questions can be pursued on the telephone, in email exchanges, or during office hours.

Evaluation

Objective measures include class participation (a whopping 15%, so make the most of it and speak up with questions and other contributions to class discussion), homework exercises (25%, for which you may work together and receive copious assistance), lab participation and assignments (15%, Dr. Lin grades this component and has the final

word on lab performance), a midterm examination (15%, in class), and a final examination (20%, take home to be worked alone, due on Thursday, December 6th at 9:00 PM). You may turn the final examination in ahead of schedule, but not after the due date.

These weights add up to 90%. An additional 10% will be added to the weight for that objective course component accounting for each student's best performance. There are many ways for a student to turn in a performance that reflects knowledge lower than his or her true state of information, but relatively few ways for a student to deliver a performance reflecting a better state of information than the state he or she actually has achieved. Consequently, I place a premium on the importance of each student's best score because this score includes more information than his or her lower scores. Consider the following example.

Course Component	Weight
Homework: Average of 5	25%
Class Participation	15%
Lab and Lab Assignments	15%
Midterm Exam	15%
Final Examination Dec 6	20%
Floating Increment	10%
Total	100%

The university permits grades of "incomplete (IN)" to be given only if the terms the University defines for issuing a grade of "incomplete" are met. See *SCampus*, https://policy.usc.edu/student/scampus/, the USC Student Handbook, for the conditions under which students might legitimately request a grade of "incomplete," and under which an instructor might legitimately accommodate such a request. Homework assignments will be distributed via blackboard and are due on the schedule indicated. Class members should respect this schedule. Late work will normally be declined.

A sample grade calculation follows below. *This is probably an "A-," or at worst a "B+":

Student X:	Score (out of 100)	Weight	Contribution
Homework: Average of 5	90.	25%	22.50
Class Participation	88.	15%	13.20
Lab & Lab Assignments	85.	15%	12.75
Midterm Examination	82.	15%	12.30
Final Examination	93.	20%+10%=30%	27.90
Course Total			**88.65**∗

Homework assignments are for instruction as well as evaluation. I am willing to discuss the homework assignments during office hours, and in class for that matter. Anyone who wants a perfect score on any homework assignment can probably get one by asking for assistance.

Cooperation is almost always the least expensive means of overcoming difficulty, so I urge students to work homework assignments in teams. The midterm and final examinations must be completed alone. You are strongly encouraged but not required to execute homework assignments in teams of two to four members. Team members will receive identical grades on group assignments. If you have not contributed to the completion of a homework assignment, please do not pester your colleagues to fraudulently append your name to their work. This would be a violation of University Conduct Code § 11.15, 11.17, 11.21, and 11.31. See the information below on academic integrity.

All homework should be submitted electronically as an email attachment to me at jmoore@usc.edu. Since you are allowed to work in groups, I think we can skip the blackboard assignment manager, unless the graders ask me to proceed otherwise. I will acknowledge your email submissions. Please do not submit homework assignments as multiple documents or files. Integrate you work for each assignment into as single document. <u>Keep an electronic copy of your submissions for your records, in any event</u>. Let me repeat that. <u>Keep an electronic copy of your submissions for your records, in any event.</u>

Website

Most class handouts will be distributed through the course website https://blackboard.usc.edu/. There is additional information available at the website, more than we will use in our course. It is there for reference. You are not responsible for every document posted there. I will direct you to what you need to know.

You should use your USC NetID to log into this site. The site is password protected with the same password as your USC NetID. You may also view a record of your scores on the website if you access the site.

Before you can log into the course website, or use USC email, you must execute the USC Information Technology Services (ITS) first login procedure. You have probably already done so, but if not, go to

https://netid.usc.edu/account_services/activate_account

You will see a screen like the one below.

Tip: If you need help from USC Information Technology Services, call them. I doubt that an email to *consult@usc.edu* has ever drawn a response in the history of the institution.

Additional Points the Provost Wants Me to Mention, Which I Do as a Courtesy[1]

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

---

[1] I offer this content in spite of the fact that it presents USC to you as a kind nanny pseudo-state, and seems to suggest that your natural role in the organization is as some flavor of victim. It is not. This is most of the content Provost wants me to share with you, but I have not included everything, because I disagree with some of the most subjective elements.

forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, http://policy.usc.edu/scientific-misconduct.

The Sol Price School of Public Policy adheres to the University of Southern California's policies and procedures governing academic integrity as described in *SCampus*. Students are expected to be aware of and to observe the academic integrity standards described there, and should expect those standards to be enforced in PPD 570, because they will be.

Support Systems

Student Counseling Services (SCS) – (213) 740-7711 – 24/7 on call
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention, engemannshc.usc.edu/counseling.

National Suicide Prevention Lifeline – 1 (800) 273-8255
Provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week, www.suicidepreventionlifeline.org.

Relationship and Sexual Violence Prevention Services (RSVP) – (213) 740-4900 – 24/7 on call
Free and confidential therapy services, workshops, and training for situations related to gender-based harm, engemannshc.usc.edu/rsvp.

Sexual Assault Resource Center
For more information about how to get help or help a survivor, rights, reporting options, and additional resources, sarc.usc.edu.

The Office of Disability Services and Programs
Provides certification for students with disabilities and helps arrange relevant accommodations, dsp.usc.edu.

Student Support and Advocacy – (213) 821-4710
Assists students and families in resolving complex issues adversely affecting their success as a student EX: personal, financial, and academic, studentaffairs.usc.edu/ssa.

USC Emergency Information
Provides safety and other updates, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible, emergency.usc.edu.

USC Department of Public Safety – UPC: (213) 740-4321 24-hour emergency or to report a crime.
Provides overall safety to USC community, **dps.usc.edu.**

Teaching Assistant

Technically, PPD 570 does not have a teaching assistant. In the interests of better serving you, we have arranged instead for two highly qualified graders for PPD 570 rather than a single teaching assistant. Our graders are Qiao Wu (qiaowu@usc.edu), and Phoebe He (whe248@usc.edu). Both are highly qualified. Qiao took the course in the summer of 2017 and completed subsequent coursework in multivariate statistics and econometrics. Phoebe was the course assistant for PPD 570 just this past summer.

The Lab Instructor, Dr. Lin, is responsible for the statistical computing lab. Questions about how to proceed with respect to the lecture material and home works should be directed to me. Questions about how to proceed with respect to the computing lab and lab assignments should be directed to Dr. Lin. Qiao and Phoebe are available to answer questions about grading decisions, but they do not have the range of responsibilities associated with a Teaching Assistant, so other questions about the material should normally come to me.

Computing and the Computer Lab

Modern statistical procedures were revolutionized by the availability of low cost computers, just as computing has revolutionized just about everything else. Our assignments are small and fundamental enough to be executed by hand with a calculator, but you are encouraged to organize your work with Excel if so inclined.

The course includes a computing lab experience to provide you hands on experience with statistical software, specifically the R programming language. The labs are organized and led by our very experienced Lab Instructor, Dr. Michael Lin [chengyil@usc.edu].

Previous editions of this course relied on the Strategic Analysis System (SAS, formerly the Statistical Analysis System). We have shifted to R because of its pervasiveness in data analytics and informatics applications. PPD 558, Multivariate Statistical Analysis, relies on the Stata statistical package, and students who elect to take PPD 558 subsequently will have the opportunity to participate in a Stata lab first.

There will be computer lab sections during most weeks in the semester, scheduled so as not to conflict with other IPPAM courses. The computing labs will meet on the Thursdays from 10:00 AM to Noon. R instruction begins starting September 7, and runs until November 16, in King Hall Computer Center, RGL 209 or 219. IPPAM may have some other orientation activities scheduled for you during the first two Fridays.

Week	Thursday	Time	Location	Topic
1	August 24	10:00-Noon	RGL 209*	Usually an IPPAM activity
2	August 31	10:00-Noon	RGL 209*	Usually an IPPAM activity
3	September 7	10:00-Noon	RGL 209*	Introduction
3	September 14	10:00-Noon	RGL 209*	SAS Data Sets I & Descriptive Statistics
4	September 21	10:00-Noon	RGL 209*	SAS Data Sets II & Variable Mgmt I
5	September 28	10:00-Noon	RGL 209*	Variable Mgmt II & SAS Data Sets III
7	October 12	10:00-Noon	RGL 209*	SAS Data Sets IV & Data Sources
8	October 19	10:00-Noon	RGL 209*	SAS Data Sets IV & Data Sources
9	October 26	10:00-Noon	RGL 209*	Correlation & Data Presentation
10	November 2	10:00-Noon	RGL 209*	Regression Analysis I
11	November 9	10:00-Noon	RGL 209*	Regression Analysis II
12	November 16	10:00-Noon	RGL 209*	Lab Final Exam
13	Thanksgiving			
14	November 30			
15	Reading period			

* or RGL 219
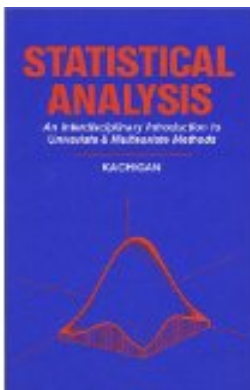
Required Texts For the Statistics Lecture

Kachigan, Sam Kash, _Statistical Analysis: An Interdisciplinary Approach to Univariate and Multivariate Methods_, Radius Press, 1986. This is available in USC Gift and Convenience Store (formerly the USC Book Store).

Supplemental Readings and Lecture Notes for PPD 570: Applied Statistics for Planning, Policy, and Management, Fall, 2018. This is available on the course website.

Assigned readings are important and will contribute significantly to your understanding of the lecture material. Fortunately, both the required text and the lecture notes are reasonably clear. The first-best strategy is to skim the assigned material before class, attend lecture, listen carefully, and then read the text assignments with discrimination.

<u>**Tentative Course Outline**</u>

<u>**Topics**</u> <u>**Assigned Readings**</u>   <u>**Week of**</u>

  I. **Fundamental Concepts** Aug 23 **K** pp 1-28
II. **Data Reduction**

    A. Frequency Distributions **K** pp 29-42

    B. Central Tendency Aug 30 **K** pp 43-53
                                    HW 0 due

III. **Probability**

    A. Basic Probability
      …Sample spaces and set logic
      …Simple discrete probability distributions **K** pp 73-87
      …Expected value of a random variable   Sept 6 **K** pp 97-98
      …Conditional probability **K** pp 87-97

    B. Advanced Probability Topics
      …Bayes formula and decision trees Sept 13 **K** pp 476-488
      …Permutations and combinations **HW 1 due** **K** pp 462-468

IV. **Inference**

    A. Variation Sept 20 **K** pp 54-72

    B. Sampling Distributions: **K** pp 102-131
      …Samples
      …Normal distribution
      …Statistics as random variables
      …Central limit theorem Sept 27, **HW 2 due**
      …Proportions as means: Binomial distribution
      …Hypergeometric distribution **K** pp 471-473
      …Poisson distribution **K** pp 473-476

    C. Parameter Estimation: Oct 4 **K** pp 134-159
      …Point and interval estimates
      …t-statistics

USCPrice
Sol Price School of Public Policy

D. Hypothesis Testing: **K** pp 160-185, 189
…Types of hypothesis sets
…Tests of a mean
…Type I and Type II errors

MIDTERM EXAMINATION   Oct 11   **In Class**

V. Association (Fundamentals of Multivariate Analysis)

A. Simple and Serial Correlation Analysis   Oct 18   **K** pp 195-226, 233-
234

                                                            B pp 125-149 (lecture notes)

B. Nonparametric Statistics   **K** pp 452-461

C. Analysis of Category Data   Oct 25   **K** pp 342-354
                                            HW 3 due
D. Simple Regression Analysis: **K** pp 238-259
…Basic linear model and assumptions
…Significance tests and confidence bands
…Proportion of variance explained

E. Multiple Regression Analysis   Nov 1   **K** pp 259-271
                                                            H pp 61-72 (lecture notes)

F. Multiple and Partial Correlation Analysis
(Qualitative): Relationship to Multiple Regression   **K** pp 226-232

G. Analysis of Variance (ANOVA)   Nov 8
…One-Way ANOVA and Multiple Regression   **K** pp 272-282
…Factorial Designs   **HW 4 due**   **K** pp 282-297

H. Multiple Regression Using Matrices
…Matrix Arithmetic   Nov 15   **R** pp 1-10 (lecture notes);

THANKSGIVING RECESS   Nov 22
…Matrix Generalization of the General Linear Model

FINAL EXAMINATION   Distributed: Nov 29
        Due: Dec 6, 2018

This schedule is optimistic, but it has been met before. If we cannot, I know what material to triage. Hopefully, you will find these techniques useful in many ways. I certainly do. In any event, this schedule is a roadmap, not a contract, and will probably be subject to minor revisions subject to our joint capacity for progress.

Andrus Center, Room 203, 3715 McClintock Avenue (to the left)

The Andrus Center first floor lobby is under renovation though October, so if you meet me in my office there you should enter from the west (Vermont) side of the building, via the elevator, which has exterior access. I will arrange for Thursday office hours in IPPAM space in VKC.