



University of Southern California
Sol Price School of Public Policy
International Public Policy and Management Program (IPPAM)
Fall 2015, 4 units

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

Prof. James Moore [jmoore@usc.edu]
 Tuesdays, 6:00 PM – 9:20 PM, RGL 209
 Some Tuesdays, 10:00 – 11:50, KOH 208

3650 McClintock Avenue, Suite 200
 Los Angeles, CA 90089-1450

Office: OHE 200M (weekdays, some weekends & evenings)
 Phone: (213) 740-0595

Office Hours: Tuesday, 2:30 – 3:45, prior to **PPD 570**
 Thursday, 10:00 – 11:15

734 West Adams Blvd., Room 204
 Los Angeles, CA 90089-7725

Research Office: KER 204 (some weekends & evenings)
 (213) 743-1920

Computing Lab / **TBD [TBD@usc.edu]** Phone: (213) ***-****
 Course Assistant:

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 *Stokhastikas*, 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 observe.

Overview

This is a first course in statistics for new graduate students with no substantive prior exposure to the field. The course is designed and delivered especially for USC Price School IPPAM students. No prior work in statistics is assumed, but students must 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. Attendance is important, and thus strongly encourage. As a matter of courtesy to all parties concerned, please arrive on time, and discuss any premature departures with me prior to the event.



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 website. 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; 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 Tuesday evening during the fall semester. We have only a very limited amount of time to cover this material, and routine attendance is strongly encouraged. If you do skip a class, the cost is yours, not mine.

This is a lecture-based course, but questions and informed discussions 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 observations), homework exercises (25%, for which you may work together and receive copious assistance), lab participation and assignments (15%, the lab assistant grades this component), a midterm examination (15%, in class), and a final examination (20%, take home to be worked alone, due on Tuesday, December 15 at 9:00 PM). You may turn the final examination in ahead of schedule, but not after the due date.



Course Component	Weight
Homework: Average of 5	25%
Class Participation	15%
Lab and Lab Assignments	15%
Midterm Exam	15%
Final Examination (Dec 11)	20%
Floating Increment	10%
Total	100%

Incomplete grades can only be given if the terms the University defines for issuing a grade of “incomplete” are met. See *SCampus*, the USC Student Handbook, for the conditions under which students might legitimately request a grade of “incomplete,” and 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.

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.

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 less bias than his or her lower scores. Consider the following example.

Student X:	Score (out of 100)	Weight	Contribution
Homework: Average of 5	90.	25%	22.50
Class Participation	88.	15%	13.20
Lab and Lab Assignments	85.	15%	12.75
Midterm Examination	82.	15%	12.30
Final Examination	93.	20%+10%=35%	27.90
Course Total			88.65*

* This is probably an “A-,” or at worst a “B+”

Cooperation is usually the least expensive means of overcoming difficulty, so I normally urge students to work homework assignments in teams. The final examination 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. The class is small enough that, since you are allowed to work in groups, I think we can skip the blackboard assignment manager. I will acknowledge your email submissions. Keep an electronic copy of your submissions for your records, in any event. Let me repeat that. Keep an electronic copy of your submissions for your records, in any event.

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.

To log into the course website, you must execute the USC Information Technology Services (ITS) first login procedure. This is web based. Go to

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

You will see a screen like the one below.

Activate your USC NetID

Your USC NetID gives you access to **online resources and tools**. Follow the steps below to get started.

- 1 Find your USC NetID
- 2 Set secondary email address
- 3 Create new password

10-digit USC ID number (from your USCard) [Help me find it](#)

Date of birth (MM/DD/YYYY) [Why we're asking](#)

Find your USCnet ID

Call us at 213-740-5555
consult@usc.edu



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.

A Couple of Additional Points the Provost Would Like Me to Mention

Academic Integrity

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 Section 11, *Behavior Violating University Standards* <https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions>. Other 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

A number of USC’s schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the *American Language Institute* (<http://dornsife.usc.edu/ali>), which sponsors courses and workshops specifically for international graduate students.

The Office of Disability Services and Programs provides certification for students with disabilities (http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html) and helps arrange the relevant accommodations.

If an officially declared emergency makes travel to campus infeasible, *USC Emergency Information* <http://emergency.usc.edu> will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.

Computing

Modern statistical procedures have been revolutionized by the availability of low cost computers. 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 Strategic Analysis System (SAS, formerly the Statistical Analysis System, http://www.sas.com/en_us/company-information.html). The labs are organized and led by our Course Assistants, **TBD [TBD@usc.edu]** and **TBD [TBD@usc.edu]**.



There will be seven or eight computer lab sections during the course of the semester, scheduled so as not to conflict with other IPPAM courses. The computing labs will meet on the following Tuesdays from 10:00 AM to Noon, starting September 8, and running until November 24, in King Hall Computer Center, KOH 208:

Week	Date	Time	Location	Topic
1				
2				
3	Tuesday, September 8	10:00-Noon	KOH 208	Introduction
4	Tuesday, September 15	10:00-Noon	KOH 208	SAS Data Sets I & Descriptive Statistics
5	Tuesday, September 22	10:00-Noon	KOH 208	SAS Data Sets II & Variable Mgmt I
6				
7	Tuesday, October 6	10:00-Noon	KOH 208	Variable Mgmt II & SAS Data Sets III
8				
9	Tuesday, October 20	10:00-Noon	KOH 208	SAS Data Sets IV & Data Sources
10	Tuesday, October 27	10:00-Noon	KOH 208	Correlation & Data Presentation
9				
12	Tuesday, November 10	10:00-Noon	KOH 208	Regression Analysis I
13	Tuesday, November 17	10:00-Noon	KOH 208	Regression Analysis II
14	Tuesday, November 24	10:00-Noon	KOH 208	Lab Final Exam
15				

There are 12 weeks in this interval, but only nine lab sessions. The computing lab does not meet every week.



Considerable statistical consulting assistance is also available from the ITS consultants at (213) 740-5555 or send email to stats@usc.edu. See also <http://itservices.usc.edu/stats/sas/> and <http://itservices.usc.edu/stats/sas/elearning/>.

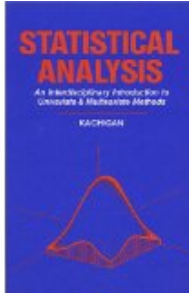
Teaching Assistant

We don't have one. It is just us. The Course Assistant 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 **TBD**, who is very qualified. They will be collaborating to determine your lab grades.



Required Texts For the Statistics Lecture

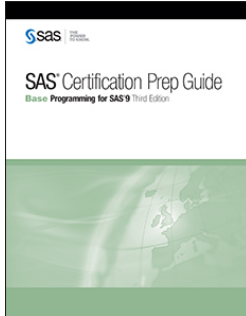


Kachigan, Sam Kash, *Statistical Analysis: An Interdisciplinary Approach to Univariate and Multivariate Methods*, Radius Press, 1986. This is available in USC Gift and Convenience Store (formerly the USC Book Store).

Supplemental Readings and Lecture Notes for PPD 570: Applied Statistics for Planning, Policy, and Management, Fall, 20015. 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 and listen carefully, and then read the text assignments with discrimination.

Recommended Texts For the Computing Lab



Additional reading material you need for the computing lab will be provided by Michael and Luna. Some will come from SAS, *SAS Certification Prep Guide: Base Programming for SAS 9 (Third Edition)* SAS Institute Inc., 2011.

<https://support.sas.com/pubscat/bookdetails.jsp?pc=63049>. Don't purchase this book. It is expensive. We will distribute what you need.



Tentative Course Outline

<u>Topics Assigned Readings</u>	<u>Week of</u>	
I. Fundamental Concepts	Aug 25	K pp 1-28
II. Data Reduction		
A. Frequency Distributions		K pp 29-42
B. Central Tendency	Sept 1 HW 0 due	K pp 43-53
III. Probability		
A. Basic Probability		
...Sample spaces and set logic		K pp 73-87
...Simple discrete probability distributions		K pp 97-98
...Expected value of a random variable	Sept 8	K pp 87-97
...Conditional probability		
B. Advanced Probability Topics		
...Bayes formula and decision trees	Sept 15	K pp 476-488
...Permutations and combinations	HW 1 due	K pp 462-468
IV. Inference		
A. Variation	Sept 22	K pp 54-72
B. Sampling Distributions:		K pp 102-131
...Samples		
...Normal distribution		
...Statistics as random variables		
...Central limit theorem	Sept 29, HW 2 due	
...Proportions as means: Binomial distribution		
...Hypergeometric distribution		K pp 471-473
...Poisson distribution		K pp 473-476
C. Parameter Estimation:	Oct 6	K pp 134-159
...Point and interval estimates		
...t-statistics		
D. Hypothesis Testing:		K pp 160-185, 189
...Types of hypothesis sets		
...Tests of a mean		
...Type I and Type II errors		

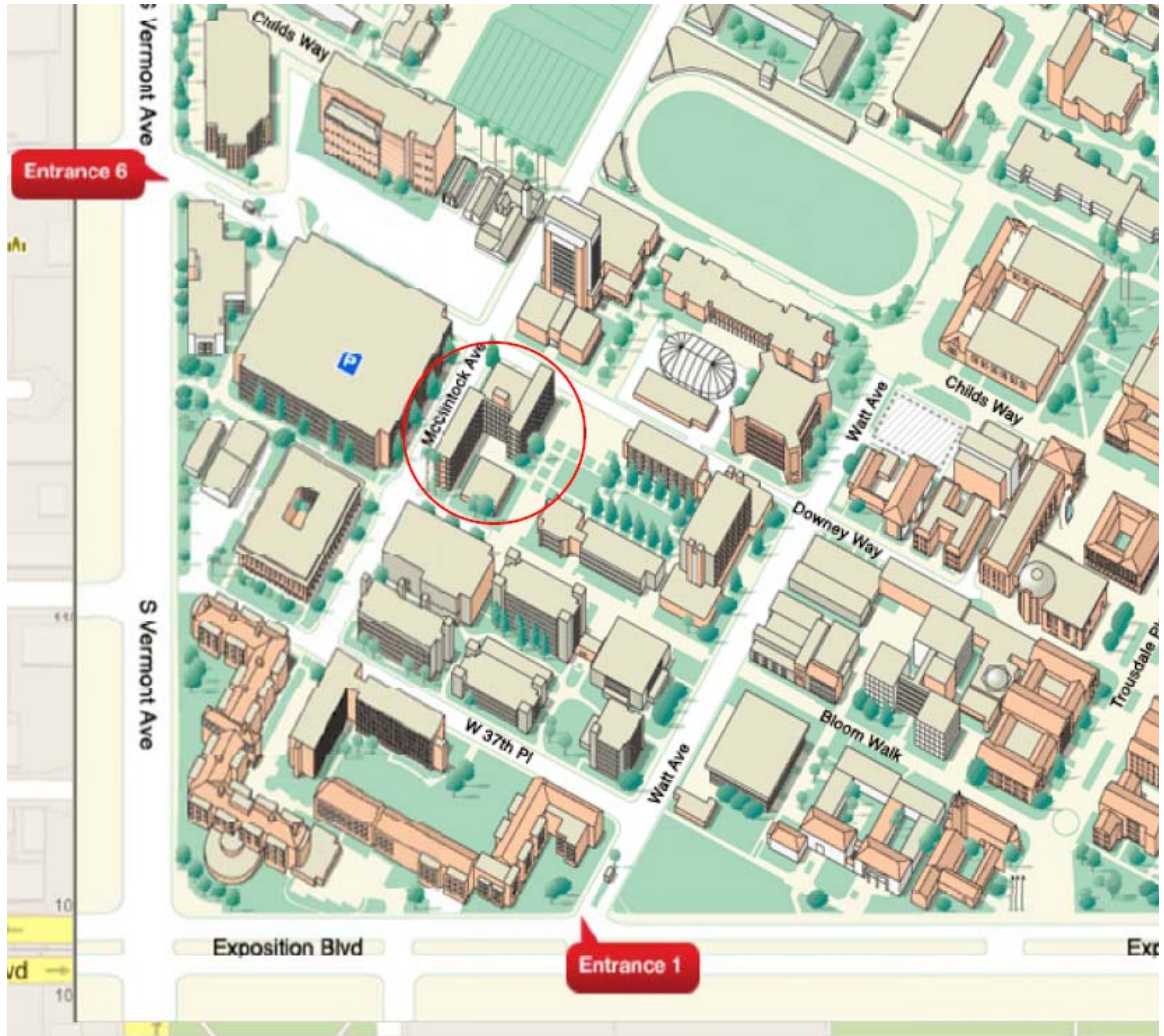


MIDTERM EXAMINATION	Oct 13	In Class
V. Association (Fundamentals of Multivariate Analysis)		
A. Simple and Serial Correlation Analysis	Oct 20	K pp 195-226, 233-234 B pp 125-149 (lecture notes)
B. Nonparametric Statistics		K pp 452-461
C. Analysis of Category Data	Oct 27 (BOO!) HW 3 due	K pp 342-354
D. Simple Regression Analysis: ...Basic linear model and assumptions ...Significance tests and confidence bands ...Proportion of variance explained		K pp 238-259
E. Multiple Regression Analysis	Nov 3	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) ...One-Way ANOVA and Multiple Regression ...Factorial Designs	Nov 10 HW 4 due	K pp 272-282 K pp 282-297
H. Multiple Regression Using Matrices ...Matrix Arithmetic	Nov 17	R pp 1-10 (lecture notes);
...Matrix Generalization of the General Linear Model	Nov 24 HW 5 due	W&W pp 331-340 (lecture notes)
I. Review/Spare Date	Dec 1	
FINAL EXAMINATION		Distributed: Dec 1 Due: Dec 15, 2015

This schedule is optimistic, but it has been met before. Hopefully, you will find these techniques useful in many ways. In any event, this schedule is not a contract, and will probably be subject to minor revisions subject to our joint capacity for progress.



Olin Hall of Engineering, Suite 200, 3650 McClintock Avenue



This exercise is ungraded, but I will record submissions. Response is required.

1. Consider the following expression, which gives y as a function of p ,

$$y = f(p) = \ln [p/(1 - p)] = \ln [p \cdot (1 - p)^{-1}].$$

I want you to invert this function and solve for p as a function of y ,

$$p = g(y) = f^{-1}(y).$$

Record the time it takes you to complete the problem. Some of you might not be able to solve the problem. Some of you will consider it trivial. If you cannot attack the problem, the time required to complete it is ∞ .

2. Send me email (jmoore@usc.edu) identifying the members of your homework group. Copy the Course Assistants (**TBD@usc.edu** and **TBD@usc.edu**).
3. Welcome to the University of Southern California. This won't hurt a bit.

