University of California, Santa Cruz
Department of Applied Mathematics and Statistics
Baskin School of Engineering
Winter 2007

AMS 7 and 7L: Statistical Methods for the Biological and Environmental Sciences

- **Background:** AMS 7 is a 5–unit class with lectures and discussion sections devoted to *statistical methods for the biological and environmental sciences*, and AMS 7L is a 2–unit computing lab connected to AMS 7 in which you'll get a chance to do hands-on statistical data analyses using a package called `JMP` that's popular in biology and environmental studies. The intent is that AMS 7 and 7L should be taken together in the same quarter.

  There is a statistics curriculum requirement for undergraduates majoring in Ecology and Evolutionary Biology (EEB) and Environmental Studies (ES). Until recently students in these majors could choose to take either AMS 5 (Statistics) or AMS 7 (Biostatistics). The new AMS 7/7L represents a complete redesign of the previous AMS 7 (which was focused mainly on medical statistics), with the goal of greater relevance to the biological and environmental sciences.

  Starting in 2006–07, AMS 7/7L is the new statistics curriculum requirement for EEB and ES students. If you're already enrolled in one of these majors, (I think) you can still take AMS 5 if you want (please check with the undergraduate advisors in your Department), but AMS 7/7L (a methods course) will better prepare you for field study and research in the biological and environmental sciences than AMS 5 (an ideas course).

  Students majoring in Molecular, Cell and Developmental Biology (MCDB) or in any of the other majors on campus that have at least a partial biological or environmental focus (such as Health Sciences or Environmental Toxicology) will also potentially find AMS 7/7L useful preparation for field study and research; in particular, as a result of recent curriculum changes, MCDB students can use AMS 7/7L to satisfy one of their laboratory course requirements.

- **Lectures** for AMS 7: TuTh 2–3.45pm, in Baskin Engineering (BE) room 152.

- **Instructor:** David Draper (office BE 135); telephone 459–1295; email

  `draper@ams.ucsc.edu`

  (I'll do my best, but due to the volume of email I receive, I can't guarantee quick response to any message you send me; please put *AMS 7 student* in the subject line of any email message you send me).

- **Web Page:** There is a course web page: its URL is http://www.soe.ucsc.edu/classes/ams007/Winter07/

It will be regularly updated with lecture notes, handouts and announcements.

- **Instructor Office Hours:** will be announced soon (initially my office hours will be held in BE 135; depending on how many people show up, we'll often move to a bigger space nearby called Jack's Lounge — it's the open space on the first floor of BE with white boards, near BE 125).

- **Enrolling:** I'll give permission codes to anybody who wants to take this class (see discussion of informal pre-requisites below).

- **TAs:** There are two TAs for the class:

Name	Email Address	Office
Juan Carlos Laguardia	jlaguard@ams.ucsc.edu	BE 142
Nicole Targowski	nikkit@ams.ucsc.edu	BE 142

Their office hours will also be announced in class soon.

- **Discussion Sections:** These have already been arranged, and you're required to enroll in one of them as part of taking the class.

Section	Day	Time	Place	Enrolled	Max	Status
01A	M	9:30–10:40am	Engineer 2 192	30	30	closed
01B	Tu	8:30–9:40am	Soc Sci 2 179	10	30	open
01C	W	2:00–3:10pm	Engineer 2 194	30	30	closed
01D	Th	noon–1:10pm	Engineer 2 194	29	30	open

- **Lab Sections:** These have also already been arranged, and you're required to enroll in one of them if you're taking AMS 7L concurrently (as is strongly recommended).

Section	Day	Time	Place	Enrolled	Max	Status
01	M	1:30–3:30pm	J Baskin Engr 109	18	30	open
02	Tu	9:30–11:30am	J Baskin Engr 109	25	30	open
03	W	11:30–1:30pm	J Baskin Engr 109	20	30	open
04	Th	7:00–9:00pm	J Baskin Engr 109	29	30	open
05	Tu	7:00–9:00pm	J Baskin Engr 109	6	30	open

If possible we'd like to close lab section 5 since it has so few people in it; we'll talk about this during the first class meeting.

- **Structure:** The content of the combined course AMS 7/7L will be presented in four weekly meetings: the TuTh lectures, a 70-minute discussion section, and a 2–hour computer lab. It's your responsibility to attend one of the discussion sections (quizzes that are a part of the grade for AMS 7 will be given in discussion sections every week). To keep the class sizes roughly uniform I ask you to regularly go to the section you're enrolled in, but from time to time you can go to another one if you need to.

  Discussion sections and labs will start on Mon 8 Jan 2007; please go to your chosen discussion section and lab next week.

  Holidays to note this quarter: no classes of any kind (discussion sections or labs) will be held on campus on **Mon Jan 15** and **Mon Feb 19**; if your regular discussion sections and/or labs take place on Mon, you'll need to go to one of the other discussion sections and/or labs during those two weeks of this quarter.

- **Individual tutoring:** Some hours of individual tutoring will be available for those who most need it. You should get the great majority of your help in this course by coming to class, discussions sections, labs, and the office hours that the TA(s) and I will give; it's best to regard the modest availability of individual tutoring as a last resort after these other resources prove insufficient. If you feel you would benefit from individual tutoring, please see me to request this. Tutoring is also available for qualified students through the Multicultural Engineering Program (MEP); see `mep.soe.ucsc.edu` for details.

General Content

Statistics is the **study of uncertainty**: how to measure it, and what to do about it. **Uncertainty** is a state of incomplete or imperfect information about something of interest to you, for example (a) the percentage p of the deer who live on the UCSC campus as of December 2006 who have chronic wasting disease or (b) the pollution status of Monterey Bay in 2012 if a law regulating the dumping of refuse from ships into the Bay comes into effect in 2008. Statistics comes up mainly in two kinds of things people do:

- *Science* (knowledge for its own sake), and
- *Decision-making* (putting that knowledge to work to make a choice among different possible actions).

Science is mostly about *facts* (for example, the percentage p mentioned in (a) above is 0.8%) and *relationships* (for instance, how the wing length of newborn sage sparrows relates to their age). Statistics is helpful with both: coming up with *estimates* and give-or-takes (measures of uncertainty) about facts (for example, on the basis of some data I have I might estimate p to be 0.8%, give or take 0.2%), and identifying which relationships are *causal* ("Smoking causes lung cancer and heart disease in humans") and which are just *associations* ("Drinking soda pop causes polio," or so they thought for awhile back in the 1930s; it does turn out that soft drink consumption and polio incidence were associated with each other, but as it happens neither was causing the other). Along the way we'll learn some of the most important basic rules of **probability**, which is the part of mathematics devoted to quantifying uncertainty.

Decision-making is mostly about **predicting** the future under different sets of conditions and choosing your favorite future; for example, policy-makers might need to choose between enacting or not enacting the law regulating the dumping of refuse from ships into Monterey Bay mentioned in (b) above, and until they gathered some data and figured out how to analyze it they would be uncertain about the two possible futures {amount of pollution in the Bay in 2012 if the law were not enacted} and {amount of pollution if the law were enacted}. Statistics has a lot to say both about how to predict things and how to figure out how accurate your predictions are likely to be.

Statistics is good both for telling you how much (or little) you know about something and for figuring out how to **design** *experiments* or *sample surveys* to get new information (*data*) to reduce your uncertainty. There's a lot of emphasis on good *graphics*: drawing pictures of your data that provide insight not readily found just by looking at the numbers (for example, a *scatterplot* of polio deaths against soft drink consumption). Statistics includes both **descriptive** methods to summarize *factuals* ("The death rate within 30 days of admission for patients aged 65 and over with a principal diagnosis of heart attack at these 10 hospitals from January through July 2005 was 17%") and methods to draw **inference** about *counterfactuals* ("I'm pretty sure that I would have gotten there faster if I had taken Soquel instead of the freeway"). Along the way we'll talk about *power* and *sample size calculations* (methods for figuring out how much data you should gather in any given situation: surprisingly, it's possible to have *too much data*), and methods for quantifying the strength of the relationship between two variables (*correlation, regression*, the *analysis of variance*, and the *analysis of categorical data*).

Statistics uses math, mainly probability, but common sense and good judgment are at least as important as math in most good statistical work. A long time ago (in the late 1700s) a really good mathematician, Pierre Simon de Laplace, put it best:

Statistics is common sense reduced to calculation.

If any of this sounds relevant to your interests, maybe this course is for you.

General Style

The course will be based on a series of **case studies** drawn from my own consulting work and that of people whose work I'm familiar with (including a variety of examples from the main text for the course and other biological/environmental statistics textbooks, and possibly also from journal articles in the biological and environmental sciences). These case studies will mainly come from the natural and social sciences and medicine, but there will also (for example) be decision-theory examples from business and other fields. The case studies typically have four components:

(1) In the first step we fully examine the *real-world problem* and make the central question(s) clear.

(2) Then we "invent" one or more *methods* to solve the problem in step (1).

(3) Next we apply the methods from step (2) to completely *solve* the problem and understand the real-world implications of the solution.

(4) Finally, we stand back and examine the *general properties* of the methods "invented" in step (2): what other kinds of problems can they help to solve? Under what conditions do they work best, and what does it take to make them fail?

I like to help people learn in an *interactive* fashion, with questions and answers going back and forth between you and me on a regular basis during the "lectures." In this manner we'll trace the discovery process that led to the original development of the methods we study. The idea is for some real learning to occur in class, not just note-taking.

Text

There is only one text book for the course, and it's required:

- Triola MM, Triola MF (2006). *Biostatistics For the Biological and Health Sciences.* Boston MA: Pearson-Addison Wesley (ISBN 0–321–19436–5).

You can get this book (new or used) at the Bay Tree bookstore; you can also find it new or used on the web at places like `half.com` or `amazon.com`. I'll also draw some examples and case studies from

- Zar JH (1999). *Biostatistical Analysis* (fourth edition). Upper Saddle River, NJ: Prentice Hall (ISBN 0–13-081542–X);

this book is not required (and in fact is not even recommended; it does have some good examples, but it's far too dry and difficult to read).

Course Prerequisites and General Education Codes

The formal prerequisites for the class are as follows:

> Score of 31 or higher on mathematics placement exam, or AMS 3, or AMS 11A, or MATH 3, or MATH 11A, or MATH 19A, or by permission of instructor.

I'm happy to give a permission code to anybody who wants one; basically you should be comfortable with high school mathematics at roughly the level of college algebra; in particular, no calculus will be used in this class (there will be liberal use of formulas involving summation notation, which I'll review soon). If you have any questions about whether you satisfy these prerequisites, please see me.

This course satisfies the following General Education Codes: **IN** and **Q**.

Course Requirements and Grades

My basic approach to grades is to try to get everybody to work hard to absorb as much of the material as they can in one quarter and then give the best grades I can, more or less consistent with past grading standards for the course. (The grade distribution is usually approximately 25–35% A, 35–45% B, 20–30% C, 0–10% D/F; anyone who sincerely tries in this class — by turning in every assignment and taking every quiz and exam, and demonstrating a basic level of understanding of the material — will pass the course.) **You'll probably notice that the material in the course is cumulative, and that its difficulty level rises slowly each week.** The final grade for AMS 7 will have four components: homework, midterm, discussion sections, and final exam.

- **Homework** (about 25%) will be assigned about 5 times during the quarter and due at the beginning of class 1–$1\frac{1}{2}$ weeks later. In order that solutions be available promptly, and because of the procedural problems inherent in the grading for a large class, **LATE HOMEWORK WILL NOT BE ACCEPTED FOR ANY REASON**. To compensate for emergencies or bad luck, your lowest homework score will be dropped from the grade computation (each homework will have about the same weight). Note that none of the homework assignments is optional.

  One possible strategy in view of the dropping of the lowest score is of course to neglect to turn in an assignment, but people who have done this in the past in courses like this one have noticed that they are unprepared on the corresponding material at exam time.

  The purpose of the homework is to develop facility in statistical thinking through regular practice, and to provide early and regular feedback on your performance in the course. **Solutions to the homework will be posted shortly after the due date in a glass case on the wall near BE 125.** You should consult these solutions and compare them to your own, and talk with the TA(s) and/or with me to resolve any confusions that remain. There is an enormous volume of homework that the graders must examine in a short time, and it's impossible for them to make detailed comments on each paper and still return them quickly enough to be useful to you. For this reason you should examine the posted solutions carefully, even if you receive a high score on your paper, since there may be some ideas you've not fully understood.

- **Midterm** (about 25%). This will be a take-home open-book open-notes exam given out around the end of the fifth week and due a week later. This will not come early enough for you to use it in any decision you might need to make about dropping the course, but you should have enough feedback from the homework and quizzes by then to make that decision.

- **Discussion sections** (about 20%). Statistics is something you learn by doing, so it's important to work a lot of problems, both by yourself and by talking with other people. You've already enrolled into a discussion section; attendance at these sections is required. The idea is to have sessions in which the TAs lead the discussion on how to solve some problems, chosen to illustrate in practice the topics being considered in lecture at that time. There will typically be one problem like the ones solved in the discussion section or like what's going on in class at the time; you'll be asked to solve this problem (open-book, open-notes) and turn your solution in for credit as a kind of small quiz.

- **Final exam** (about 30%). The final will be an in-class, open-book, open-notes exam. It will be cumulative, but with emphasis on the material after the midterm.

In AMS 7L there will be a series of 4–8 computer labs for you to complete (in a take-home fashion, on your own time) and turn in for credit; they will be equally weighted in the final grade.

Two final notes about grades:

- Incompletes will be given only in clear cases of emergency.
- Anybody who is a senior and who needs to pass this course by the end of this year to graduate should start working today—waiting til nearly the last minute to take the course does not guarantee a passing grade.

Collaboration, Plagiarism, and Cheating

You're encouraged to form study groups for the purpose of discussing the homework problems, but **all of the written work you turn in for this class must be your own efforts**. Even though the volume of homework the graders will be evaluating is large, it's surprisingly easy to spot instances where someone has simply copied someone else's solution, and this will be even easier to identify with the take-home midterm (unlike the homework, **you're not allowed to discuss the take-home midterm problems with anybody else**). In fairness to the many people who do not cheat, instances of plagiarism and other forms of cheating will be dealt with as harshly as possible in the UCSC system.

Calculators

Everybody should have available a **calculator** (with charged batteries or solar power) for use during exams and discussion sections. It's important that this machine have a square root key in addition to the usual arithmetic operations, and it's helpful to have at least one memory; for some calculations we'll do toward the end of the class, logarithm and exponential keys are also helpful. You can find calculators like this for $5–10 (or even less) these days.

Lectures, Discussion Sections, and Readings

You're responsible for everything that goes on in class, discussion sections, and labs, and for obtaining any written material that's distributed. **The TA(s) and I will often refer back in lectures, discussion sections, and labs to handouts originally covered in previous classes, so I recommend that you gather together all of the handouts**

for the course in a notebook or ring binder and bring it to all lectures, discussion sections, and labs.

You should do the assigned readings *before* coming to class, discussion section, or lab. Ordinarily, the lecture will discuss aspects of the readings in detail or will present additional material not contained in the texts. Neither the lectures nor the readings can be substituted for one another. The discussion sections will sometimes introduce new material and will involve turning in some written work at their conclusion, so regular nonattendance will clearly hurt your chances of performing well. **There is a strong causal relationship in this class between {taking all of the homeworks and quizzes seriously} and {getting a good grade}**.

Preparing Homework

Here are some guidelines for getting your homework ready to turn in; please follow them. The graders have an amazingly small amount of time to look at your paper and pass judgment on it—anything you can do to improve its form, by making it relatively neat and easy to follow, will maximize your chance of a good grade on the homework.

- **Submit** homework on 8.5 by 11 or 8.5 by 14 paper only, and make sure that your **name** is **clearly printed** on all pages of anything you turn in.
- Use **staples** or paper clips to hold together submissions of more than one page.
- Write **legibly** and **coherently**. Manuscripts that are unintelligible in either content or handwriting are not likely to be looked on favorably.