Fall 2003
Biostatistics, BIOL380
Michael Dohm, Ph.D.

Class Location: 105 UCB (new University Classroom Building)
Class Times: Tuesday (T) & Thursday (R) 3:30 - 4:45 PM
MD's Office: PB 14-8
Phone: 933-3155 (with voicemail)
E-mail: dohm@hawaii.edu
Office Hours: M 3:00 - 5:00 PM, plus after class, or by appointment
Teaching Assistant: Jessica Urbic
Main Web site: natsci225.uhh.hawaii.edu
Mirror Site: www2.hawaii.edu/~dohm

Course Description

Biostatistics (BIOL380) is a lecture + hands-on course designed to provide students with skills necessary to design experiments, analyze and report data in the biological sciences.

Major activities in the course will include understanding statistical analyses in biomedical sciences, ecology and evolution, experimental design, concepts of data collection, use of graphics and writing of reports in correct, scientific format. We will emphasize statistical reasoning ("what are the assumptions, what tests are best or most appropriate for this data set and hypotheses...") over memorization of formulas, although the basic understanding of how a particular test works (i.e., how it is calculated and why it is best over other options) will be included.

Biostatistics students will become familiar with Minitab, a computer program, to assist in analyzing data and presenting results of statistical tests. Additional computer skills, including some concepts of data management, manipulation of data in spreadsheets, communication with the instructor via e-mail, and use of file attachments, will also be developed or enhanced.

Our Goals

The ultimate goal of BIOL380 is to have students complete an original research project. All phases of research will be conducted by students (as groups): experimental design, developing hypotheses, gathering and analyzing data. Results and conclusions will be communicated by group presentation to the class and by written report (each member writes their own paper), in standard scientific format suitable for submission to a typical journal. By the end of the semester, student's will be able to demonstrate knowledge of many important statistical terms and concepts, the ability to analyze data and write research reports on their findings.

Students will learn about statistical hypothesis testing using analyses commonly of interest to potential employers, graduate or professional school admission committees, or professional journals. Basic statistical knowledge will include descriptive statistics, assumptions of hypothesis testing, techniques for testing hypotheses involving categorical and quantitative data, T-tests, chi-square tests, analysis of variance, simple and multiple regression. Our emphasis will be on statistical reasoning and communication of our results! We won't "memorize" formulas, nor will we spend a lot of time with hand-calculations, but rather, we will use output from Minitab to make biologically relevant statements about data.

Thus, an important objective of BIOL 380 is course is to provide students with a framework to discuss and critique reports of biomedical research findings and results from Ecology, conservation biology, and genetics.

Statistical reasoning

My definition of "statistical reasoning," it's the ability to understand and use numbers to communicate findings and to support opinions. Statistical reasoning is important to your future, for two reasons. First, the biological sciences are, by their nature, an experimental science. Biotechnology, botany, cell biology, health, environmental, ecology, evolution, genetics, medical, microbiology, molecular biology, physiology, zoology all involve the presentation and analysis of data.

Consider a biomedical example: You work as an AIDS counselor in a major U.S. city. Before you is the result of a male client's HIV test. It is positive, meaning the ELISA and Western Blot tests detected HIV in the blood sample. The client, who insists that he does not fit any known risk behavior group, now sits before you, waiting to hear the results. Your training has given you the following facts: (1) less than 0.01% of men fitting this category (heterosexual, no IV drug use) are infected with HIV in the U.S.; (2) the sensitivity of the tests is very high, 99.9% (sensitivity is defined as the percentage of individuals with a disease who are correctly classified by a test as having the disease); and the specificity of the tests is also high, 99.9% (specificity is defined as the

percentage of individuals without a disease who are correctly classified by a test as not having the disease). So, how do you communicate the client's test results?

Second, statistical reasoning is common to our daily lives. The above example illustrates the problem from the perspective of a counselor, but as citizens, we are faced with a barrage of numbers: our risk of developing a particular disease (breast cancer, 1 in 9 women; prostate cancer, 1 in 9 men, etc.); our risk from dying under general anesthesia (on average, about 0.01% or 1 death every 10,000 U.S. cases); our risk of dying in a car crash versus our risk of dying in a plane crash (see Lecture 1). Data summaries, projections and predictions, and declarations of probability and likelihood are important tools for communicating complex information. The media commonly employ these tools to present information to you: The latest HIV drug tests, putative links between diet and health, effects of environmental toxins on health... the list is long. Often times, the media's portrayal of medical and health news are inadequate precisely because they do not pay enough attention to statistical issues. One result from the media's poor presentation of science results is the impression that just about everything and anything can be shown to adversely affect health... not a very useful conclusion. More importantly, exaggerated claims are made with numbers, and it is difficult some times to sort out fact from interpretation when numbers are used.

Textbook(s)

Required text: We will use a new textbook, "Introduction to Biostatistics," by Tom Glover and Kevin Mitchell. This is a first edition book (published in 2002 by McGraw-Hill).

My lecture notes: I write out my lecture notes and make them available to you before the start of class. They are extensive and cover more information than your required textbook. I've been doing this since I first taught the class at UHH, so there are copies floating around (see Online textbook). Note however, that the current notes take precedence.

Online textbook. To help you prepare for class, I have provided last semester's notes (Fall 2002) as an ADOBE PDF file for you to download (click here to get the book). Feel free to download and distribute the book. Much of what we will cover this semester is contained within those pages, and you may find the additional source of information helpful.

Other texts (recommended but not required). Last year (2001-2002), we used two different books, listed below. I still recommend them, and you may find used copies around.

1 Zar, J. 1999. Biostatistical Analysis, 4th Ed., Prentice Hall Press.
2 Dytham, C. 1999. Choosing and Using Statistics: A Biologist's Guide. Blackwell Scientific Press.
3 Sokal, R. R., and F. J. Rohlf. 1996. Biometry, 3rd edition. I also highly recommend this textbook for those of you who plan to continue in biology.

A note about these books. Glover and Mitchell's text book was written for courses like ours: a one semester introduction to biostatistics for undergraduates. Therefore, there's not a lot of detail, just the facts and some nice homework problems to help you understand the mechanics of the formulas. My lectures supplement the text and together provide what I believe is a solid introduction to biostatistics. You are responsible for all material discussed or presented in class plus whatever assigned readings from Glover & Mitchell.

Zar's book (and Sokal & Rohlf's book), however, is a reference book. It is not a text book in the sense you may be more familiar with. At times, the writing is quite dense and may be a bit intimidating. Zar's book is a pretty typical introduction to statistics for biology students, generally first-year graduate students. Zar's discussions of the formulas and what goes into a particular statistical test are top-notch, and our new book references Zar extensively. The single best feature of the Zar book is that it is comprehensive -- it will serve you very well for future work you may do. In fact, if you plan to continue on (graduate school in the biosciences), I highly recommend that you pick up a copy of Zar: you will need a reference book by your side as you pursue your research interests to help you justify the kinds of statistical analyses you use and work out assumptions of your tests.

Dytham's little green book is very readable and covers much of the same material, but in a much more superficial way -- a true introduction to the material. Dytham's book also provides some tips on using Minitab. Together, the books provide a sound foundation to statistics.

I have placed a number of statistical textbooks on reserve for your use at the library.

Assignments

Homework: There will be five short assignments that will help you understand concepts that we cover in class. Assignments involve analyzing data relevant to lecture topics and writing short reports that include presenting experimental designs, methods and statistical results using tables and graphs. You may work together on homework, *but each student must turn in their own homework*. Your textbook also has many nice problems to work on: I will recommend many, but will not grade work from the text book. Of course, the more you do, the better your understanding will be!

Small Group Project: Groups of 2-3 students will design a project of interest in the biological sciences. The project must include the following elements: experimental design, data collection, data analysis, and conclusions. Graded elements of the project include the following:

1. Project design proposal (group written): Student groups will turn in a one-page proposal with the following elements: (1) a descriptive project title; (2) 1-2 sentences to introduce what the project will address; (3) 1-2 sentences to justify why the project is interesting; (4) 2-4 sentences to address the primary hypotheses that will be tested; (5) 2-3 sentences to describe what type of experiment will be used to gather data that addresses the hypotheses; and (6) 2-3 sentences to describe how data will be analyzed (e.g., what statistical procedures might you use to test the hypotheses).

2. Oral Report (group presentation): **Student groups will present the report** to the class at the end of the semester (last week of class). As a group, students will take 10-15 minutes of class time to present the results of their project.

3. Final Written Report (individual): **Each student will also submit a final**, written manuscript that follows standard scientific journal format. The report is done in collaboration with the student's group, but each student must turn in their own paper, written only by the student.

To summarize, a group may consist of from one to three students; members of a group work together to identify a project, write a proposal, design the appropriate experiment, gather data, analyze data, then present their results to the class in the form of a 10-15 minute talk (typically a PowerPoint presentation). However, every student is responsible for turning in a written report, using scientific format, in which the student presents the elements of the project complete with his or her own interpretations of what the project was about and what conclusions, if any can be drawn from the project.

Exams: Three exams, two will be given in class with a take-home final (due 1 week:

1. Exam I will cover material from weeks 1-6.
2. Exam II will cover material from weeks 7 - 12.
3. The final exam will be cumulative (40%), plus cover material from weeks 13-15 (60%).

The exams will focus on your understanding of important concepts, your ability to evaluate experimental designs, and your ability to recognize appropriate statistics to be employed given sets of data or particular experimental designs.

Cheat sheets: You will be allowed to bring a calculator plus one 8 ½ x 11 in. paper (2-sided) with notes to the two midterm exams. The final is a take-home exam, so you will be allowed to use your notes, lecture notes, plus your textbook and other reference material, but without consultation among friends, of course!

Class Participation: You are expected to come to class prepared, to participate fully in class by asking questions, bringing attention to media announcements of relevance to biostatistics, by helping in data collection needed for homework projects. I also encourage all of you to assist fellow students with computer and Minitab related questions. While I will not take attendance on a regular basis, I do note who attends class. I use the "Socratic method" on occasions: You can expect to be called upon to answer questions at least a few times during the semester, and how you respond will be reflected in class participation. Lastly, attendance at both sessions of oral project reports (the last two days of class) is mandatory: absence from one (or both) of the sessions will be reflected in class participation scores.

Scoring for class participation is as follows: everyone starts with a score of 5/10. During the semester, your participation score will go up, or down from 5 for a maximum of 0 points possible, the equivalent of one-half of a homework assignment.

Grading

Assignment	Points
Homework #1	**20**
Homework #2	**20**
Homework #3	**20**
Homework #4	**20**
Homework #5	**20**
Group Projects	
Design (group)	**10**
Presentation (group)	**20**
Written (individual)	**40**
Exam #1	**60**
Exam #2	**60**
Final Exam (cumul.)	**100**
Participation	**10**

Letter Grade	%	Point Range
A	92.0 -100	368-400
A-	89.0 >	356-367
B+	86.0 >	344-355
B	80.0 >	320-343
B-	78.0 >	312-319
C+	75.0 >	300-311
C	67.0 >	268-299
D	50.0 >	200-267
F	< 50.0	0-199

Bonus points -- there will be opportunities to earn a few bonus points from time to time, at my discretion. No more than 10 points will be possible, however!!

Course Policies

Deadlines -- we have to live with deadlines. Exams, homework, class projects, we need enough time to complete assignments, but at some point, we, or the work we produce, needs to be evaluated.

Late? Simply put, I will not and cannot accept any excuse for missing a test, or turning a homework or other assignment to me late, without substantial penalty, unless you make arrangements with me BEFORE the deadline date. If a paper is late without approval, I will grade it, but for every day past the deadline, 10% will be taken. If the paper is more than 4 days late, the grade will be zero. Again, you must get prior approval for turning in material late. You can contact me by e-mail, by phone, but if you do not hear back from me, then you cannot assume that the excuse has been accepted.

Class attendance -- Yes, I do take attendance into account (see participation). The material in the class comes fast and furious. Every day you miss means more frustrations for you in terms of understanding the material. You must keep in mind that, although I post my lecture notes, I also make statements in class that will not appear in the web notes -- you are responsible for all material presented in class. I've found that class attendance by students is directly proportional to class performance: those that come to class do much better than those that attend sporadically. See notes under class participation for additional information. I will also conduct review sessions prior to all exams: these will include an in-class review plus generally a Saturday session. We'll work out the dates and times as we go.

Office hours. I have reserved 2 hours per week for office hours (Monday 2-4 PM). I also welcome you to make an appointment with me so that we can address any questions or concerns you may have about biostatistics. However, please refrain from attempting to drop-in on me without having first made an appointment (e-mail or by phone). I am employed full-time as a researcher and my schedule is quite full during the day.

Computer assistance. There will be an additional 2 hours per week scheduled in LS-1 so that you may get help with Minitab and related computer software during the semester. In general, the teaching assistant for the class will be available at that time to answer questions. Exact times will be announced in class. Attendance at these sessions is optional but encouraged.

Additional Assistance? Any student with a documented disability who would like to request accommodations should contact the University Disability Services Office - Campus Center Rm. 311, 933-0816 (V), 933-3334 (TTY) - as early in the semester as possible.