



BIOSTATISTICS

Fall 2008

*"Chance favors
the prepared
mind."*
- Louis Pasteur

Instructor: Dr. Alan Hale (SC134, 610-606-4666 x3510, abhale@cedarcrest.edu)

Course Description [BIO 248]: An introduction to parametric and nonparametric statistical methods commonly used to analyze biological data. Students will learn and apply these methods in order to support their current or future research in ecological, molecular, chemical, or health-related fields.

Course Web Site: In case you are looking at a hard copy of the syllabus, it can also be found at <http://www2.cedarcrest.edu/academic/bio/hale/biostat.html>. Visit the site to replace a lost copy or to check out some interesting web resources that are related to the course material.

Bulletin Board: The bulletin board outside the MacLab (SC132) will be used for this course. At times I will post data sets, answer keys, interesting web sites and news articles; take a break once in a while to check them out.

Goals: The primary goals of this course relate to expanding your horizons in terms of critical thinking and quantitative analyses of data. Working in the sciences without a functional understanding of statistics is like trying to race in a regatta without knowing how to trim the sails. In both cases your chances of success are greatly diminished.

Learning Outcomes/Assessment: Upon successfully completing Biostatistics you should:

- Know how to effectively analyze the data you have collected for your current research, if in fact you are involved in research at this time.
Assessment: 4 examinations, class assignments, optional oral presentation
- Know how to use statistical software to effectively analyze data.
Assessment: 4 examinations, class assignments, homework
- Have shed the intimidation that is so commonly present when the topic of statistics arises.
Assessment: 4 examinations, class assignments, optional oral presentation
- Be able to design a research plan, particularly in terms of choosing statistical tests that effectively address your research questions.
Assessment: 4 examinations, class assignments, optional oral presentation
- Be able to listen to scientific presentations and many forms of media in the real world and know whether the data were analyzed correctly.

Assessment: questions at seminars, 4 examinations, book presentation, oral presentations with associated questions/comments

- Be prepared to understand biostatistics more fully as you become involved in future research programs.

Assessment: 4 examinations, class assignments, homework, optional oral presentation

Content: It is very important to be able to choose the correct statistical test to analyze a particular set of data. In addition, it is very important to make sure the assumptions of a statistical test are met before using a test. Both of these will be addressed in detail throughout the course. Many statistical tests and topics will be covered. These include the characteristics of data in biostatistics, common descriptive statistics, probability distributions, the theory behind hypothesis testing, and an assortment of parametric and nonparametric tests (e.g., t-test, analysis of variance, regression and correlation analyses, goodness of fit tests). Don't forget that a primary goal of the course is to help you develop a good understanding of the basics in statistics; from there you can expand your understanding of data analysis through independent learning or through formal courses.

Textbook: I enrolled in my first statistics class in the early 70's; several more courses followed throughout the same decade and beyond. Although I enjoyed all of the statistics courses, I must say that I never found stat textbooks terribly compelling. However, the primary textbook for BIO 248 has changed this perspective. Not only is it easy to read, fun and captivating, it also leaves the reader with a strong foundation in statistics. Moments ago I completed the chapter on Analysis of Variance; I had just wanted to check on a few things in the chapter but ended up reading the whole chapter from start to finish. Believe it or not, reading this textbook is like reading a good novel. Have fun with: *The Complete Idiot's Guide to Statistics*, 2nd Edition (Robert A. Donnelly Jr. Ph.D. / ISBN 978-1-59257-634-0 paperback) and with *How to Lie with Statistics* (Darrell Huff / ISBN 0-393-31072-8 paperback).

Companion Web Site: The author of your textbook has set up a companion web site to provide a diversity of resources. You'll notice by reading the text that he is quite the family man. If you would like to match described behavior with faces, take a look at the companion web site. Photos of his wife, his children, and himself are provided.

[*Complete Idiot's Guide to Statistics! Companion Web Site*](#)

Supplies: A bound (non-looseleaf) composition notebook will be needed for your homework. The color of the cover and the design on the pages should reflect your personal preferences. A calculator that is programmed to compute simple statistics (e.g., mean, standard deviation) may be useful, though you may choose to use a laptop and Microsoft Excel to analyze data. Both calculators and laptops (Excel only) may be used during examinations.

Statistical Software: "A bird in the hand is worth two in the bush." It's late at night, you're sitting in your dorm room, and you need to run some statistical tests. Which would you rather do: hike through the snow to get to a computer lab that may or may not be open, or curl up on your bed with loads of blankets

and your laptop? Right! Consequently, we will be making good use of Microsoft Excel and the guidance provided by the author of your textbook. Another reason for learning how to run a great diversity of statistical tests with Excel is because wherever you go in the future, Excel will follow, whereas some of the more specialized packages (e.g., *Statistica*) may not. Nonetheless, know that *Statistica* is installed on all of the PCs in our W.M. Keck Bioinformatics Laboratory (MB22). It integrates data management, statistical analysis, graphics, and presentation. Its data sets, charts and graphics can easily be transferred to Microsoft Word and PowerPoint documents, as is the case with Excel.

Other (purely optional) Resources: Before I list the following resources I want to make it clear that one can do quite well in BIO 248 with out these. I include them on the syllabus because some students want to push Excel to its limits. For those of you who find this attractive, you might consider picking up one or more of the following books to use as handy resources during times of stress. Keep in mind that you are not likely to find these in the college bookstore; each is approximately \$25 through online book distributors. So here they are: (1) *Statistical Analysis with Excel for Dummies* (Joseph Schmuller); (2) *Excel Data Analysis for Dummies* (Stephen L. Nelson); (3) *Excel Charts for Dummies* (Ken Bluttman). There is one more resource that you might want to consider in the future if you happen to take a liking to statistics and Excel. Excel as a component of Microsoft Office comes with a good set of statistical tests. Nevertheless, you may get really hooked on Excel and want an expanded arsenal of statistical tests, including nonparametric tests.. I found myself in this situation with my own research and while helping others with their statistics. I tested a number of statistics packages available on the web and have been quite happy with the one I finally purchased (\$245). XLSTAT Pro can be purchased on the web [<http://www.xlstat.com/>], but be sure you get **student pricing** if you decide to get a copy. Student pricing is approximately \$50.00, but in the case of this significant discount, the software license is for only one year. I recommend that you master what is free on Excel before you jump to XLSTAT Pro. If you're looking for one particular test that you need for your research, you might just google it to see if there is any freeware available. And don't forget, *Statistica* in the Bioinformatics Lab is loaded with statistical tests.

Grading: Final course grades will be based on the following examinations and assignments, and will be assigned on a percentage basis. Grades will be posted on the web throughout the semester so you will have the option to check how you did on the recent exam from the comforts of your dorm room. You will be given a password, which you should keep to yourself, to access your grades at the following site:

[Online Grade Reports](#)

By all means, if you see a mistake on your grade summary, please bring it to my attention.

<u>Activity</u>	<u>Points</u>	<u>Session/Info</u>
-----------------	---------------	---------------------

Examination #1 [Key]	100	Session 7 75 minute written examination; calculator, textbook, and laptop (Excel only) allowed
Examination #2 [Key]	100	Session 14 75 minute written examination; calculator, textbook, and laptop (Excel only) allowed
Examination #3 [Key]	100	Session 21 75 minute written examination; calculator, textbook, and laptop (Excel only) allowed
Homework Notebook	30	Session 25
Assignments [<u>group or individual</u> , your choice]	10 per assignment	Assigned sporadically during semester. Deadline is typically in one week.. Expect approximately five assignments.
<i>How to Lie with Statistics</i> Presentation (Group Project)	20	Group presentations are given during Session 26.
(Optional) Research Statistics: Oral Presentation; Limited Number - First Come/First Served <i>Abstract due 1 week before presentation.</i>	maximum of 20 bonus points, including 5 for the abstract	Session 28 Sign-up sheet posted outside MacLab - 10 minute presentation plus 4 minutes for questions and comments
Final Examination	100	Final Exam Period 3 hour written examination; calculator, textbook, and laptop (Excel and course website) allowed
Attendance	50	Think of your understanding of statistics as a tree. If the tree is missing a section of its trunk, it's not difficult to be toppled. (-2 pts. per absence)

Examinations 1, 2 and 3: These will be in-class, written examinations. The amount of time available to complete the examinations will not exceed 75 minutes. The better you understand the material prior to your arrival, the less time you'll need to complete it. Do not plan to learn material during the exam because you will run out of time; your textbook is there for quick reference, not for extensive guidance. You may use a calculator, laptop, and the primary textbook (Donnelly) to complete each examination. If you do not have a laptop computer, you will have access to either the classroom computer or computers in a computer lab (e.g., MB20 and 22) if you choose to use one. Use of a laptop (or desktop computer) must be restricted to the statistical tools available through Excel, though for the final exam you will have the option to access the course website for useful links (e.g., Scheffe Test, important tables).. The internet (including email), text files, and non-Excel programs (except for what was noted above) must not be used during examinations; an exam grade of zero will be given to individuals who violate this rule. Although the focus of an exam will be on material covered since the previous exam, keep in mind that in statistics new material builds upon concepts learned previously, and therefore consider each exam as being comprehensive.

Homework: We live in a society where people expect things to be easy: get rich quick strategies; the ultimate diet pill; learn a language faster than ever before; perfect abs in nothing flat. Maybe someday there will be options similar to those in Matrix, but currently if you want to find success, the name of the game is hard work. Developing an intuitive grasp of statistics requires a good bit of your time and thought. Some of this time and thought should be devoted to explaining concepts and solving statistical problems, two important components of this homework assignment. At the end of each chapter in the Donnelly text there are "Your Turn" questions/problems (Q/P). Each student should complete within a bound composition notebook every Q/P in the textbook. If Excel or another program that performs statistical analyses is used to do some of the analyses, the important parts of the output should be taped neatly into the respective Q/P section of your notebook. Do not tape folded paper into your notebook, instead cut from the output that which is important and which fits on the page of your notebook. Tape all edges. After completing the Q/P of a given chapter, you should sign and date the last page used for the chapter's Q/P answers. It is recommended that students complete each Q/P before looking at the correct answer at the back of the textbook. Corrections to the original answers should be noted, this is a sign of learning. Answers in the composition book should be in one's own words, not as written in the textbook.

Assignments: Class time is typically used for lecture, however at times we use it for data collection and analyses. To carry the latter to completion, group or individual assignments are commonly due within one week. Each assignment is submitted via email as a single Word file with charts and data incorporated into the same file. Student names should be placed at the top of each assignment; do not use headers or footers for this purpose because Dr. Hale will be combining all assignments into one file, adding comments, and emailing the composite file to all students enrolled in the course. Students and groups are encouraged to follow the prescribed format of the assignment.

***How to Lie with Statistics* Presentation:** The book, *How to Lie with Statistics*, was written in 1954. Although cell phones, laptops and iPods did not exist back then, it will become obvious as you read this book that human behavior has changed very little since then. People continue to stretch the truth to convey their intended message. Begin to read this book at the beginning of the semester and watch for

contemporary examples that illustrate how you can lie with statistics. At some point, groups of approximately four students will be formed and each group will select two chapters to focus on. First come, first served. Near the end of the semester each group will present (approx. 15 minutes) their two chapters to the class. Presentations should include an overview of the primary message within each chapter and several examples found in the media these days. The latter will involve watching a lot of TV, listening to the radio, surfing the web, catching the latest show on your iPod, or reading your favorite newspaper or magazine. Be sure your examples closely match the chapter content. A word for the wise: procrastination does not mesh well with this project. Begin early in the semester.

Final Examination: The final exam will be given during the final exam period and will cover all aspects of the biostatistics course, including readings, lectures, and oral presentations. As with Exams 1, 2 and 3, a calculator, laptop, and primary textbook are allowed. Unlike Exams 1, 2 and 3, you will have access to the course website during the final exam; you will need important tables and online programs.

Optional Presentation of Research Statistics (Extra Credit): Students will be giving presentations so the class can see several examples of how statistical analyses can significantly enhance the quality of one's scientific research. If you are currently conducting lab or field research, or if you know where you are headed in the future with your research, or if you participated in an exciting summer research program, I encourage you to share your discoveries with the class. An important goal of this presentation opportunity is to encourage you to work out the statistics you want to use for your research well before you have collected all of your data through the statistical design associated with your research. Presentations should be ten minutes long with an additional four minutes for questions and comments. The focus of your talk should be on the research question(s), hypotheses, data collected, statistical analyses and conclusions. It is your responsibility to convince the class that the chosen data set and statistical analyses adequately address the research question(s). Please keep in mind that members of the class will do their best to identify errors in your statistical design and conclusions, this is not because they are malicious individuals, but rather because they realize that it is a good learning experience for all. Recall that this is an optional presentation. A maximum of 20 bonus points can be earned by giving a presentation to the class. The number of points you receive will reflect the clarity of your presentation, the quality of your statistical design and analyses, and your ability to respond to questions and comments. One lecture period has been set aside for presentations, therefore a total of 5 slots will be available. A sign-up sheet will be posted on the bulletin board outside SC132. It will be first come, first served, however, do not sign up for a slot until you are sure you are willing to give a presentation. If you sign up but then withdraw your name, ten points will be deducted from your grade. Why, you ask? It is possible that more than five individuals will want to present their research; everyone participating is entitled to the same amount of prep time. Please note: An electronic copy of an abstract (~ 1 page) is due one week before the presentation; in addition to an overview of the research project and its findings, the abstract should include at least the sample sizes and statistical tests used in the study. Dr. Hale will distribute the abstract to all members of the class prior to the presentation.

Research Seminars: Students are encouraged to attend research seminars within the department, or beyond. During the talks, focus on how the speaker has used statistics to form conclusions. Do not hesitate to ask the speaker questions about his/her statistical analyses.

Grading Scale

Students with percent averages in the categories shown below will receive the respective letter grades.

<u>Final Grade</u>	A	A-	B+	B	B-
<u>Total Percent</u>	92.50-100%	89.5-92.49%	86.5-89.49%	82.5-86.49%	79.5-82.49%
C+	C	C-	D+	D	F
76.5-79.49%	72.5-76.49%	69.5-72.49%	66.5-69.49%	59.5-66.49%	less than 59.5%

Attendance: Statistics can be challenging, so take advantage of the opportunity to have someone help you understand the subject.

Computer Use Policy and Honor Code: The Biology Department supports the CCC Computer Use Policy and the Honor Code as described in the Customs Book.

Lecture Schedule - Fall 2008



Session	Date	Topic	Chapter(s)
1	Aug 26	<u>Introduction to Course</u> <u>Discussion of Assignments and Examinations</u> <u>Role and Samples of Statistics in Science</u> <u>Microsoft Excel</u>	Syllabus
2	Aug 28	<u>Let's Get Started*</u>	1
3	Sept 2	<u>Data, Data Everywhere and Not a Drop to Drink</u>	2
4	Sept 4	<u>Displaying Descriptive Statistics</u>	3
5	Sept 9	<u>Calculating Descriptive Statistics: Measures of Central Tendency</u>	4

6	Sept 11	<u><i>Calculating Descriptive Statistics: Measures of Dispersion</i></u>	5
7	Sept 16	Examination #1 <u><i>Sample Exam</i></u>	1-5
8	Sept 18	<u><i>Introduction to Probability</i></u>	6
9	Sept 23	<u><i>More Probability Stuff</i></u>	7
10	Sept 25	<u><i>Counting Principles and Probability Distributions</i></u>	8
11	Sept 30	<u><i>The Binomial Probability Distribution</i></u>	9
12	Oct 2	<u><i>The Poisson Probability Distribution</i></u>	10
13	Oct 7	<u><i>The Normal Probability Distribution</i></u>	11
14	Oct 9	Examination #2 <u><i>Sample Exam</i></u>	6-11
	Oct 14	<i>Fall Break</i>	<i>Enjoy!</i>
15	Oct 16	<u><i>Sampling</i></u>	12
16	Oct 21	<u><i>Sampling Distributions</i></u>	13
17	Oct 23	<u><i>Confidence Intervals</i></u>	14
18	Oct 28	<u><i>Introduction to Hypothesis Testing</i></u>	15
19	Oct 30	<u><i>Hypothesis Testing with One Sample</i></u>	16
20	Nov 4	<u><i>Hypothesis Testing with Two Samples</i></u>	17
21	Nov 6	Examination #3 <u><i>Sample Exam</i></u>	12-17

22	Nov 11	<u><i>The Chi-Square Probability Distribution</i></u>	18
23	Nov 13	<u><i>Analysis of Variance</i></u>	19
24	Nov 18	<u><i>Correlation and Simple Regression</i></u>	20
25	Nov 20	<u>Testing for Normality and Homoscedasticity</u> <u>Nonparametric Statistics</u> <i>Notebooks Due!</i>	Course Web Site
26	Nov 25	<u>Group Presentations:</u> <u><i>How to Lie with Statistics</i></u>	Book
	Nov 27	<i>Thanksgiving Break</i>	
27	Dec 2	<u>Scenario Session</u> <u>Putting It All Together</u>	Sessions 1-26
28	Dec 4	<u>Oral Presentations:</u> <u>Research Statistics</u>	-
Final	Dec 11-15	<u>Final Examination</u> <u>Sample Question</u>	Sessions 1-27 <i>How to Lie with Statistics</i> Oral presentations

* textbook chapter titles

Alan B. Hale (abhale@cedarcrest.edu)

Last updated: 21 July '08