Biology/Mathematics 436 Advanced Applied Statistics

"Data analysis is an aid to thinking and not a replacement for." -- Richard Shillington

"It's easy to lie with statistics. But it is easier to lie without them." — Frederick Mosteller

"Statistical thinking will one day be as necessary a qualification for efficient citizenship as the ability to read and write."*--H.G. Wells (this is attributed to Wells but he really didn't say it, see the end of the syllabus for the real quote)*

***Instructor*:** Dr. Sean Walker, Department of Biological Science
***Office*:** McCarthy Hall 389 ***Laboratory*:** McCarthy Hall 342
***Office Hours*:** M 2-5, W 1:30-3:30, and by appointment.
***Phone*: Office** 714-278-3610, **Lab** 714-278-8204, **Cell** 714-686-6821
***Email*:** swalker@fullerton.edu
***Personal Web Page*:** http://biology.fullerton.edu/swalker
***Course Web Page*:** Course Web Site: The course web-site will be on blackboard. You should be able to access it through http://my.fullerton.edu and follow the links to blackboard then to Biol 436 or Math 436.

Required Text
Zar, Jerrold H.1999. Biostatistical Analysis, 4th ed. Prentice Hall, NJ.
ISBN 0-13-081542-X

Other required material is on the CD-ROM I've given you. It contains all the information you could ever want to know about SAS, as well as some data sets that we'll use in class or for homework.

Class Meeting Times
Tuesday and Thursday 2:30 PM - 3:45 PM MH 452
Laboratory Tuesday and Thursday 4:00 PM – 5:15 PM MH 452

Prerequisites: Math 337 or 338; Junior or Senior Standing

Course Description
In this course you will be introduced to many different statistical techniques used to analyze data and will familiarize yourself with computer programs used to do statistical analyses (e.g. JMP, SAS, Minitab, SPSS). In particular, we will focus on statistical techniques used with biological data (these techniques can be applied to basically any data). We will examine one sample and two sample tests, linear models (e.g. ANOVA, including mixed models & regression), basic categorical data analysis including Poisson and logistic regression, and basic re-sampling methods. Also, the class will focus on the proper application of these statistical methods, their assumptions and how to check them and the interpretation of the results obtained. At the end of this course you should be able to carry out and interpret many of the statistical analyses mentioned above as well as design experiments that utilize said techniques.

Computer Programs and Statistics

There are numerous programs that can be used for statistical analysis (e.g. Excel, JMP, SAS, Minitab, SPSS). In class we will utilize SAS. This is a wonderful program that does almost any kind of analysis you can think of (we won't quite do all of this) and, if you use a windows computer, you can obtain a copy of the software. You are welcome to use whatever computer program you would like for completion of your homework and projects. However, I may or may not be able to help you with your analyses if it is a program I am unfamiliar with. If you choose to do work outside of the computer lab, my recommendation to you is to obtain a copy of SAS, JMP, which is another product made by SAS.

Grading Policy: In this course we will utilize the +/- grading scale adopted by CSUF (see pp. 18-19 in the schedule).

Score	Letter Grade	Assigned G.P.A.
>95%	A+	4
≥ 90%	A	4
≥ 87%	A-	3.7
≥ 83%	B+	3.3
≥ 80%	B	3
≥ 77%	B-	2.7
≥ 73%	C+	2.3
≥ 70%	C	2
≥ 67%	C-	1.7
≥ 63%	D+	1.3
≥ 60%	D	1
≥ 57%	D-	0.7
< 57%	F	0

A grade of C or better is required for the course to count towards the Biology or Mathematics major

Assignments

Assignment	Percentage
Email	**1 %**
Homework	**20%**
Take Home Exams (2)	**40%**
Group Project	**19%**
Paper Interpretations (4)	**20%**

Email Assignment (1%) DUE BY FEB 4 2005

It is critical that I know the e-mail address that you actually use. This will facilitate rapid communication between myself and the class. In addition, there are three questions I'd like you to answer when you send me the email.

1) What previous mathematics/biology courses have you taken that are relevant to this course?
2) Is there a particular statistical technique you'd like to see addressed in this course (I'll try but no promises)?
3) What is your favorite math problem or model?

Homework Assignments (20%)

For most of the lecture topics you will have assigned homework. These may be from the book or may be data-sets that I have collected over the years for you to analyze. We will generally have time to work on some homework in the laboratory. If the homework requires you to use SAS we will have time in class and there are open laboratory times (TBA).

Take Home Exams (40%)

We will have 2 take-home examinations each worth 20% of your grade. These are to be completed on your own. Each of these will have analyses and problems for you to complete. These will be handed out in class and you will have one-week to complete it.

Paper Interpretations 20% (each one is worth 5%)

The objective of this assignment is to become familiar with interpreting statistical results in the literature. You will present the following information:

1) The hypothesis to be tested.
2) The statistical test or tests used to evaluate the hypothesis.
3) How the statistical test relates to the hypothesis (i.e. how rejection of the null hypothesis is related to the hypothesis being tested).
4) The assumptions of the statistical test being used.
5) How the authors should have evaluated these assumptions (this may or may not be in the paper and **YOU** should, by utilizing the text and other resources, find the assumptions and how to evaluate them).

You can choose any paper you would like but. But you should write the papers on these techniques in this order. For due dates see the schedule.

Technique

1) T-test/One-Way Anova
2) Multi-Factor ANOVA
3) Multiple Regression or
4) Contingency Tables/ Categorical Data Analysis

Group Project (14%)

Groups of 3-4 students (each group must contain students from both Mathematics and Biology) will carry out a data analysis project. Each group will be responsible for analyzing and presenting results from the analysis of at least two, you are welcome to do three, unrelated data sets. These data can come from graduate or undergraduate projects, published papers, or from other sources (e.g. the web, other professors etc.). In addition, you are welcome to use statistical techniques that we may not use in this class. This is your chance to work with your data or to examine a burning question that you might have about some phenomenon.

Each group will be responsible for giving an 10-15 minute oral presentation on their projects and a final written report will be due at the end of the semester. Each student will write their own report on the group project. This can be done in collaboration with the group, HOWEVER, each student will turn in their own report.

GRADUATE STUDENTS will write a report for each analysis project (i.e. they will write at least two) and **UNDERGRADUATES** will write a report for only one of the analyses.

A report for a single data analysis should include the following:

1) A brief, one paragraph, introduction to the hypothesis to be tested.
2) A one paragraph introduction to the statistical test or tests used to evaluate the hypothesis.
3) How the statistical test relates to the hypothesis (i.e. how rejection of the null hypothesis is related to the hypothesis being tested).
4) The assumptions of the statistical test being used and **HOW** you have evaluated these assumptions.
5) A presentation of the results (this should be done graphically if possible).
6) A closing summary paragraph.

The following is a timeline for the project:

1) I must have the name of your group (be creative) and the members of your group by March 1, 2005.
2) The data sets, in excel format, should be turned in to my by March 24, 2005
3) Presentations will occur during the last four weeks of class (starting April 26,2005).
4) Final Group Project Papers will be due on May 26, 2005 and can be handed in with the take home final.

Course Policies

Attendance

Students are expected to attend and participate in lectures, laboratories and mandatory field trips. If you miss class YOU ARE RESPONSIBLE for obtaining the information from classmates NOT from the graduate assistant or instructor.

Exam, Lab and Assignment Make Up Policy

If you cannot take a test at the scheduled time, you should contact Sean (Dr. Walker) as soon as possible with appropriate documentation verifying the circumstances. **PLEASE NOTE** make ups will only be given in the case of documented emergencies or unavoidable conflicts (these must be approved by Sean in advance). Please note, it is **YOUR RESPONSIBILITY** to contact Sean regarding make up assignments, labs, or exams.

Late Assignments

Late work will have 10% of the maximum points for that assignment deducted per day that it is late (weekends count). If there are exceptional circumstances the assignment may be given full credit.

ACADEMIC INTEGRITY: I assume that by remaining enrolled in this class your intentions are HONORABLE, and that you accept responsibility for dutiful attendance, earnest effort toward understanding the subject and pledge that you will not cheat on exams.

· **Plagiarism** is the unacknowledged use of another's words or ideas as your own. Use your own words when writing. Use quotation marks and cite the source of any phrase that you "use". Changing one or two words in a sentence is still plagiarism, you must put the information into your own words.

· **Cheating** is the use of another's work as your own. Copying another student's homework, looking at another student's exam, and using information from another student to enhance your performance on a task are all examples of cheating.

Students who violate university standards of academic integrity are subject to disciplinary sanctions, including failure in the course and suspension from the university. University policies are strictly enforced in this course. Please familiarize yourself with the academic integrity guidelines found in the current student handbook.

Withdrawl from courses: CSUF policy regarding withdrawal from classes (UPS 300.016) will be followed. After the first two weeks of the semester, students may be granted withdrawal ONLY by presenting compelling evidence outlining a physical, medical, or emotional condition that prevents completion of the course. **POOR ACADEMIC PERFORMANCE IS NOT EVIDENCE OF A SERIOUS REASON FOR WITHDRAWAL**. Students unable to produce official documentation will be required to take the grade they have earned in the class. Please refer to the course schedule for information on the last day to withdraw with a W grade. Important dates concerning registration or drops are on the inside cover of the CSUF **Spring** 2004 Class Schedule or at:
http://www.fullerton.edu/admissions/policy_and_deadline_information_.htm.

CLASSROOM SAFETY BRIEFING

- In the event of an emergency such as earthquake or fire:
    - Take all your personal belongings and leave the classroom (or lab). Use the stairways located at the east, west, or center of the building.
    - Do not use the elevator. They may not be working once the alarm sounds.
    - Go to the lawn area towards Nutwood Avenue. Stay with class members for further instruction.
    - For additional information on exits, fire alarms and telephones, **Building Evacuation Maps** are located near each elevator.
    - Anyone who may have difficulty evacuating the building, please see me after class.
- Dial 911 on any campus phone, pay phone, or blue emergency phones to connect directly to University Police. Dialing 911 on your cell phone will connect with the Highway Patrol. Tell CHP dispatcher that CSUF Police are the responding agency. Stay on the line until asked to hang up.
- If you want to bring visitors to the classroom, you must obtain permission from the instructor in advance and must sign a volunteer form.
- Visitors to the lab must obtain permission from the Chair and must sign a volunteer form.

- There is no smoking within 20 feet of every campus building. This includes the MH balcony.
- **FOR LAB CLASSES:** Specific hazards or risks in the lab will be discussed prior to each experiment. If you have any questions about the safety of an experiment, please contact me or the lab instructor.
    - If there is a spill of a hazardous chemical, notify your TA immediately.
    - Report all injuries to me or the TA immediately.
    - All students must read and sign the departmental, "Laboratory safety procedures" form at the beginning of each semester.
- **FOR CLASSES WITH FIELD TRIPS:**
    - Make sure you submit an Academic Field Trip Waiver and sign the Participant List for each field trip.
    - Students must comply with all State laws regarding possession, sale and use of alcohol or controlled substances while participating in CSUF related activities.

Tentative Lecture Schedule

Week	Date	Topic	Reading	Assignment Due
1	Feb 1	**Introduction to Class and Preliminary Assessment**		
	Feb 3	Populations, Samples, and Probability	Chapters 1, 2, 3, 4 & 5	
2	Feb 8	The Normal Distribution	Chapter 6	
	Feb 10	One Sample Hypotheses	Chapter 7	
3	Feb 15	One-Sample Hypotheses cont.	Chapter 7	HW Assignment # 1
	Feb 17	Designing Experiments	Hurlbert, Ecological Monographs	
4	Feb 22	Two Sample Hypotheses	Chapter 8	
	Feb 24	Assumptions & Alternatives	Chapter 8	HW Assignment # 2
5	March 1	Paired Sample Hypotheses	Chapter 9	
	March 3	Multi-Sample Hypotheses (ANOVA)	Chapter 10	
6	March 8	Multi-Sample Hypotheses (ANOVA)	Chapter 10	HW Assignment # 3
	March 10	Assumptions of ANOVA & Transformations	Chapter 10 & 13	
7	March 15	Making a model for means	Chapter 17	HW Assignment # 4
	March 17	IN CLASS WORK DAY ON TAKE HOME EXAM		
8	March 22	Multiple Comparison Procedures	Chapter 11 Day & Quinn, Ecological Monographs	Paper Interpretation # 1
	March 24	Non-Parametric Alternatives to ANOVA	Chapter 10	
9	March 29	***SPRING BREAK-NO CLASSES*** ☺		
	March 31			
10	April 5	Multifactorial ANOVA	Chapters 12 & 14	HW Assignment # 5
	April 7	Multifactorial ANOVA	Chapters 12 & 14	

11	April 12	Nested ANOVA	Chapter 15	
	April 14	Simple Linear Regression & Correlation	Chapter 17 & 19	Paper Interpretation # 2

12	April 19	Analysis of Covariance (comparing regressions)	Chapter 18	HW Assignment #6
	April 21	Multiple Regression	Chapter 20	
13	April 26	Multiple Regression cont.	Chapter 20	PRESENTATIONS START
	April 28	Goodness of Fit	Chapter 22	HW Assignments #7
14	May 3	Contingency Tables	Chapter 23	
	May 5	Contingency Tables	Chapter 23	Paper Interpretation # 3
15	May 10	Contingency Tables Cont. & Log-linear models	Chapter 23	HW Assignment # 8
	May 12	Logistic Regression	Chapter 15 from Moore and McCabe, Trexler & Travis, Ecology	
16	May 17	A brief introduction to Generalized Linear Models	GENMOD_SUGI.pdf	HW Assignment # 9
	May 19	Resampling Methods	Chapter 14 from Design and Analysis of Ecological Experiments	Paper Interpretation # 4
17	May 26	Take Home and Final & Group Project Paper DUE by 5 pm		

Tentative Laboratory Schedule

Week	Date	Topic	SAS INFORMATION RELATED TO TECHNIQUE FOR THE DAY
1	Feb 1	**Introduction to Class and Preliminary Assessment**	
	Feb 3	Populations, Samples, and Probability	Introduction to SAS & the Analyst Ch 1, 2 SAS_ANALYSTAPP_V9_1.pdf
2	Feb 8	The Normal Distribution	Descriptive Statistics 2-Proc UNIVARIATE (in Volume 3, Chapter 4) & MEANS BASE_SASPROC_V9_1.pdf
	Feb 10	One Sample Hypotheses	Home-work Day
3	Feb 15	One-Sample Hypotheses cont.	Chapter 8 Hypothesis Tests SAS_ANALYSTAPP_V9_1.pdf pp 212-218, Constructing Tests in EXCEL
	Feb 17	Designing Experiments	More PROC UNIVARIATE
4	Feb 22	Two Sample Hypotheses	Home-Work Day
	Feb 24	Assumptions & Alternatives	Chapter 8 (SAS_ANALYSTAPP_V9_1.pdf) and PROC TTEST (SAS_STAT_UG_V9_1.pdf)
5	March 1	Paired Sample Hypotheses	PROC NPAR1WAY (SAS_STAT_UG_V9_1.pdf)
	March 3	Multi-Sample Hypotheses (ANOVA)	Home-Work Day
6	March 8	Multi-Sample Hypotheses (ANOVA)	Chapter 10 (SAS_ANALYSTAPP_V9_1.pdf) & PROC ANOVA (SAS_STAT_UG_V9_1.pdf)
	March 10	Assumptions of ANOVA & Transformations	Home-Work Day
7	March 15	Making a model for means	PROC ANOVA (SAS_STAT_UG_V9_1.pdf)
	March 17	IN CLASS WORK DAY ON TAKE HOME EXAM	
8	March 22	Multiple Comparison Procedures	Chapter 10 (SAS_ANALYSTAPP_V9_1.pdf) & PROC GLM continued
	March 24	Non-Parametric Alternatives to ANOVA	PROC NPAR1WAY (SAS_STAT_UG_V9_1.pdf)
9	March 29	***SPRING BREAK-NO CLASSES***	

	March 31	☺	
10	April 5	Multifactorial ANOVA	Chapter 10 (SAS_ANALYSTAPP_V9_1.pdf) & PROC GLM continued
	April 7	Multifactorial ANOVA	Chapter 10 (SAS_ANALYSTAPP_V9_1.pdf) & PROC GLM continued
11	April 12	Nested ANOVA	Chapter 10 (SAS_ANALYSTAPP_V9_1.pdf) & PROC GLM continued
	April 14	Simple Linear Regression & Correlation	Home-Work Day
12	April 19	Analysis of Covariance (comparing regressions)	PROC CORR (BASE_SASPROC_V9_1.pdf) REG, & GLM (SAS_STAT_UG_V9_1.pdf) Chapter 11 (SAS_ANALYSTAPP_V9_1.pdf) &
	April 21	Multiple Regression	Chapter 11 (SAS_ANALYSTAPP_V9_1.pdf) & PROC REG & GLM
13	April 26	Multiple Regression cont.	Home-Work Day
	April 28	Goodness of Fit	EXCEL & PROC FREQ (SAS_STAT_UG_V9_1.pdf)
14	May 3	Contingency Tables	PROC FREQ & CATMOD (SAS_STAT_UG_V9_1.pdf)
	May 5	Contingency Tables	Home-Work Day
15	May 10	Contingency Tables Cont. & Log-linear models	PROC LOGISTIC & GENMOD (SAS_STAT_UG_V9_1.pdf)
	May 12	Logistic Regression	PROC GENMOD & CATMOD (SAS_STAT_UG_V9_1.pdf) /Home-Work Day
16	May 17	A brief introduction to Generalized Linear Models	PROC GENMOD(SAS_STAT_UG_V9_1.pdf)
	May 19	Resampling Methods	Programming examples
17	May 26	Take Home and Final & Group Project Paper DUE by 5 pm	

Homework Assignments for BIOL/MATH 436

Homework assignment #1

Chapter 3: 3.1, 3.2, 3.3, 3.2
Chapter 4: 4.1, 4.2
Chapter 5: 5.1, 5.3, 5.8, 5.14

Homework Assignment #2

Chapter 6: 6.1 a, b; 6.5

Also, utilizing the data in 6.5 use PROC UNIVARIATE to obtain a histogram of the data, a normal probability plot, and the Shapiro-Wilk Test for normality. Make sure you do the statistical analysis and interpret it for full credit.

Chapter 7: 7.1, 7,2, 7.3

Homework Assignment #3

Chapter 8: 8.1, 8.2, 8.5
Chapter 9: 9.1 & 9.2

Homework Assignment #4

Chapter 10: 10.1 Also, a graphical presentation of the data.
ANOVA data from CD (ANOVA_1) details are in the folder

Homework Assignment # 5

Chapter 10: 10.5 (compare this result to the result you obtained in 10.1)
Chapter 11: 11.2 (also use Tukey & Student-Newman-Keuls) & 11.4
ANOVA data from CD (ANOVA_1.xls) details are in the folder

Homework Assignment #6

Chapter 12: 12.1, 12.2
Chapter 14: 14.1,
Chapter 15: 15.1

Homework Assignment #7

Perform a simple linear regression analysis of the data from problem 19.1 p. 410
Regression Data #1 (REGR_1.xls) details in folder

Homework Assignment #8

ANCOVA Data #1 (ANCOV_1.xls) details are in the folder
Chapter 20: 20.1
Chapter 22: 22.2

Homework Assignment #9

Chapter 23: 23.1, 23.3 (for 23.3 use a log-linear model as well)

The Story on the H.G.Wells Quote..

The actual quote is "The great body of physical science, a great deal of the essential fact of financial science, and endless social and political problems are only accessible and only thinkable to those who have had a sound training in mathematical analysis, and the time may not be very remote when it will be understood that for complete initiation as an efficient citizen of one of the new great complex world-wide states that are now developing, it is as necessary to be able to compute, to think in averages and maxima and minima, as it is now to be able to read and write." and occurs in the sixth chapter of *Mankind in the Making*. The original quote was modified in a book written in 1929 called *Studies in the History of Statistical Method*. Anyway, the story, as I understand it can be found at: http://www.dartmouth.edu/~chance/chance_news/recent_news/chance_news_11.03.html#item11.