EEOS 611
Applied Statistics
Wiser Class No. 1296
Handout 1: Syllabus
1st class: 1/28/08 M
Gallagher home
Revised: 1/28/08
Pdf/html version
Course flyer

SPRING 2008 EEOS611 SYLLABUS
APPLIED STATISTICS

TABLE OF CONTENTS

List of Tables

Key Web Links

Table 1. Web links.		
URL	**Site**	**Description**
http://boston.umassonline.net	WebCT	Students from the Wiser registration system have already been added to the WebCt course.
http://www.lms.umb.edu/studenthandout.pdf	WebCT Student Handout	Information for first-time users of WebCT
http://alpha.es.umb.edu/faculty/edg/files/edgwebp.htm	Gallagher's web page	Contains course-related links
http://www.proaxis.com/%7Epanorama/home.htm	Sleuth web page	Ramsey & Schafer post their datasets and errata pages here.

Course Description

I designed the course for graduate students who use statistics in their research, plan to use statistics, or need to interpret statistical analyses performed by others. The primary audience are graduate students in the environmental sciences, but the course should benefit just about anyone who is in graduate school in the natural sciences. The course is not designed for those who want a simple overview of statistics; we'll learn by analyzing real data. This course or equivalent is required for UMB Biology and EEOS Ph.D. students. It is a recommended course for several of the intercampus graduate school of marine science program options.

I modeled this course after the applied statistics series at the University of Washington. The text for EEOS611, **Ramsey & Schafer's "The Statistical Sleuth,"** was developed for the graduate applied statistics curriculum at Oregon State University. The audience at OSU comprises ecologists, foresters, oceanographers and graduate students in the health sciences. Ramsey & Schafer describe their philosophy of teaching statistics, based strongly on the analysis of case studies, in their J. Stat Ed. article, 'Teaching the Craft of Data Analysis:' **http://www.amstat.org/publications/jse/v11n1/schafer.html**

Knowledge of statistics is essential to doing science. Statistics, like a foreign language or a programming language, can only be learned through doing and practice. The basic steps involved in doing a given statistical analysis with SPSS can be taught in a matter of minutes. However, much of skill involved in doing statistics, and science, is framing an appropriate question to be answered, collecting the relevant data, checking the important assumptions, properly interpreting the results, and presenting them to an audience. Learning how to get the p

values from a computer program is just the beginning of a proper analysis. The more important parts of the course emphasize the hypotheses being tested, selecting the appropriate test statistics, and deciding upon the degree of statistical inference allowed by the data. Finally, all students in the class will present their statistical results to their classmates.

Here are some of the highlights in the course, which follows the textbook closely (with links to the **full outline** below):

- The kinds of statistical inference are possible with a given sampling design (**Chapter 1** in Sleuth and all subsequent chapters)
- Student's t test (**Chapter 2**)
  - What are the assumptions that matter in performing t tests? (**Chapter 3**)
  - What options are available when these assumptions are violated? (**Chapter 4**)
- One-way ANOVA
  - One-way ANOVA and assumptions (**Chapter 5**)
  - *A priori* and *a posteriori* hypothesis tests in ANOVA (**Chapter 6**)
- Regression
  - The history of regression: The method of least squares and Galton's regression to mediocrity, OLS regression model (**Chapter 7**)
  - Assumptions, diagnostics, lack of fit (**Chapter 8**)
  - Multiple regression (**Chapter 9**)
  - Testing hypotheses with the multiple regression model, dummy variables, ANCOVA & Regression diagnostics (**Chapter 10** & **Chapter 11**)
  - Stepwise selection procedures (**Chapter 12**)
- Factorial ANOVA
  - Two-way (**Chapter 13**)
  - Factorial without replication (**Chapter 14**)
  - Factorial designs (**Chapter 24**)
- Time-series analysis: serial correlation and adjustments for serial correlation (**Chapter 15**)
- Repeated measures designs (**Chapter 16)**
- Multivariate analyses, PCA and correspondence analysis (**Chapter 17**)
- Categorical data analysis, odds ratios (**Chapter 18**, **Chapter 19**)
- General Linear Model: Logistic regression (**Chapter 20**)
- Logistic regression for binomial counts (**Chapter 21**)
- Research design & power analyses (**Chapter 23** & **Chapter 24**)

Who should take the course

The course is designed to meet the needs of graduate students from most disciplines in the natural sciences: biologists, chemists, physical oceanographers, and policy students. Examples from the textbook are drawn from a wide array of disciplines including the environmental and health sciences.

I encourage undergraduate and non-matriculated students to take the class, with a warning that this course is a graduate-level course that is also a core-course requirement for doctoral students. It requires more work than most graduate courses. Non-matriculated students should talk to me about taking the course. I need to sign a form required for admission to the class.

Prerequisites & Requirements

AN INTRODUCTORY STATISTICS COURSE REQUIRED

This is **NOT** a first course in statistics. The course and course textbook are designed for students who have had at least one course in statistics, like EEOS601. All of statistics is based on probability theory, so students enrolling in EEOS611 should be familiar with basic concepts in probability. The textbook assumes that you have had a first course in statistics that stresses the probabilistic basis of all statistical tests.

The Spring 2006 course requires at least one prior statistics course or hands-on experience using statistics (or the instructor's permission). Here are some of the things you should know **before** enrolling in EEOS611:

- The meanings of degrees of freedom, p level, confidence intervals, null hypothesis, Type I & II error, and statistical power
- Be able to discuss the need for randomization in experimental and survey design.
- Know how to use the binomial, Chi-square, normal, F, Poisson, and Student's t distributions
- Know how to analyze 2 x 2 contingency tables with chi-square statistics
- Know how to calculate and interpret a least-squares regression

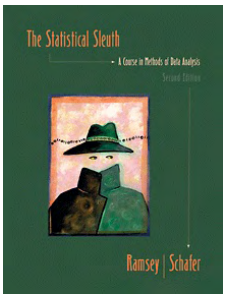
IF YOU DON'T HAVE THE PREREQUISITE COURSE

EEOS601 was offered in Fall 2005, 2006 and will be offered again in Fall 2007. I'd recommend students without the prerequisite take EEOS601 before attempting EEOS611. EEOS261, statistics for geographers is being offered this semester.

Textbooks & Software

REQUIRED

Textbooks and software can be purchased from the UMASS/Boston bookstore in person or online **http://www.efollett.com**. Via the web, you can find cheaper prices, but the UMASS bookstore will buy back your books.



- Ramsey, F. L. and D. W. Schafer. 2002. The statistical sleuth: a course in methods of data analysis, Second edition Duxbury Press, Belmont CA, 742 pp & data diskette.
  - New, this textbook is now $127 new, but it is available used in the UMASS/Boston bookstore and online (**http://www.efollett.com**) for $95. Note, if you can pick up an inexpensive copy of the 1st edition, you can use that for the semester. There are very few changes between the two versions.
- Access to SPSS Version 10 to 15 (SPSS versions are backward compatible).
  - This is for an unlimited copy of the full SPSS version. There is a student version of SPSS Version 15 available for about $80 through Prentice-Hall (Do a Google search for vendors). The stripped-down student version does not allow you to do analyses with syntax, and you will be quite frustrated. Rather than just entering a set of ascii commands to perform an analysis, you'll be forced to click away on a number of different windows buttons. I agree with **Raynald Levesque**, the guru of SPSS macros, that syntax is a must for anyone who intends to use SPSS for serious work. I will distribute the solutions to the case studies in SPSS syntax. With the student version, you'll be able to do the analyses but only by using the mouse to click on all of the appropriate icons. I know that money is important to graduate students, so you WILL be able to solve all of the problems with the inexpensive student version. You can go to the library to access the full version if you need it.
  - Each student must have access to a statistical analysis package comparable to SPSS Version 10-15. There are five ways to access SPSS:
    - The bookstore will have copies of the SPSS Version 15 Graduate Pack available for about $200 new: **http://www.journeyed.com/itemDetail.asp?T1=62391573+R**
  - There are copies of SPSS available on dozens of computers in the library. All students in EEOS611 can use the graduate student and faculty computing center on the 5th floor of the library (Matlab and SAS are also available on those computers)
  - There are 6-month and 12-month 'rental' versions of SPSS available through e-academy. The cost is about $79.99 for a 6-mo rental of the the grad pack (which is a poor deal): **http://estore.e-academy.com/index.cfm?loc=spss/main**
  - I am encouraging faculty in Biology and EEOS who will have graduate students in EEOS611 to purchase a University-licensed copy of SPSS for their laboratory

computers. The University owns a site license and faculty can obtain a CD with the program for a small price (Faculty should send an email to **Needcd@umb.ed** requesting a copy of SPSS Version 15 and the latest version of the AMOS software (AMOS7). The cost is $15 for the CD and $15 for each faculty computer on which the program is installed.)

- Any previous version of SPSS (SPSS versions 8-14) would be adequate for the course. SPSS produces Mac versions too, but these would have to be special ordered by the bookstore (takes a few days).

- Students should **not** attempt to take the class using a statistical package other than SPSS

Recommended Books

Ramsey, F. L. and D. W. Schafer. 2002b. The Solutions Manual for The Statistical Sleuth: a course in methods of data analysis. Duxbury Press, Pacific Grove, CA. 40 p. *[This is an important addition to the textbook, but it is onlyat 40 p. and expensive. It is available used for under $20 at the bookstore {online* **http://shop.efollett.com***. }]*

SPSS Base 15.0 Syntax reference guide. This book, available from the **www.spss.com** provides the syntax for unleashing the full power of SPSS. All of the syntax is available through the SPSS help system, but this is the one manual that is well worth the roughly $40 cost.

Other Recommended Probability and Statistics Books

Comments

Before choosing Ramsey & Schafer's "Statistical Sleuth", I purchased and reviewed a number of texts for EEOS611. For a more elementary discussion of some of the topics covered in this course, try **Devore (2000)**. For a different view of much of the same material with a biological focus, **Sokal & Rohlf (1995)** and **Zar (1999)** are excellent. **Underwood (1997)** provides a superb overview of the uses of ANOVA in field ecology, but **Quinn & Keough (2002)** is even better at covering more recent methods, including generalized linear models and multivariate methods. **Winer et al. (1991)**, **Neter et al. (1996)** and **Draper & Smith (1981, 1998)** provide a great deal of additional material on ANOVA, regression, inverse regression, weighted regression and general linear models. These books include far more mathematical detail than is found in Statistical Sleuth, with **Draper & Smith (1998)** expressing the matrix models in sufficient detail that they can be directly written as Matlab m.files. **Campbell & Kenny (1999)** covers regression artifacts, especially regression to the mean and problems with the interpretation of analysis of covariance. **Harrell (2002)** provides an analysis of modern regression modeling procedures. **Manly (1997)** provides a more detailed discussion of Monte Carlo methods for biologists. **Middleton (2000)** provides some nice routines for doing regression and curve fitting with MATLAB™. **Agresti (1996)** provides the best description I've seen of tests of independence, appropriate for EEOS611. Many chemists need to know about propagation of error, a subject not included in Sleuth but covered well in **Bevington & Robinson (1992)** and **Taylor (1997)**.

Agresti, A. 1996. An introduction to categorical data analysis. Wiley, New York. *[This is a superb supplement to the textbook's analysis of r x c contingency tables, and generalized linear models with categorical data]*

Bevington, P. R. and D. K. Robinson. 1992. Data reduction and error analysis for the physical sciences. McGraw-Hill, Boston. 328 pp. *[This book covers the same basics as **Taylor (1997)** but goes into considerably more detail about how to propagate error with complicated functions]*

Campbell, D. T. and D. A. Kenny. 1999. A primer on regression artifacts. The Guilford Press, New York. *[Regression to the mean in all of its different forms.]{?}*

Devore, J. L. 2000. Probability and statistics for engineering and the sciences. Duxbury Press, Pacific Grove CA. 775 p and data CD. *[This text covers some of the same ground as **Ramsey & Schafer (2002)**, but is not as comprehensive or rigorous]*

Draper, N. R. and H. Smith. 1998. Applied Regression Analysis, 3rd Edition. John Wiley & Sons, New York. 706 p, with data diskette.*[This is a major revision of the tremendous 2nd edition, now with a greatly enhanced treatment of regression diagnostics. Includes a superb discussion of inverse regression]*

Fitzmaurice, G. M., N. M. Laird, and J. H. Ware. 2004. Applied longitudinal analysis. John Wiley & Sons, Hoboken NJ. 506 p.

Hayek, L.-A. C. and M. A. Buzas. 1996. Surveying Natural Populations. Columbia University Press. 448 p. *[A superb guide to survey sampling for ecologists. It also has some profound insights into analyzing species evenness.]*

Harrell, F. E. 2002. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Springer-Verlag, New York. 591 pp. *[A superb description of how advanced regression can be used to model data. Spends a fair amount of time with splines, a subject not addressed in sleuth.]*

Hollander, M. and D. A. Wolfe. 1999. Nonparametric Statistical Methods, 2nd edition. John Wiley & Sons, New York. 787 p. *[An extremely valuable reference book that supplements brief discussion of distribution-free statistics in **Ramsey & Schafer 2002**]*

Manly, B. F. J. 1997. Randomization, bootstrap and Monte Carlo methods in biology. Chapman & Hall, London. 399 p. *[This remarkable book shows how with a high speed computer (and a program like MATLAB™), you can generate confidence intervals and test hypotheses for a wide range of biological problems without having to assume a parametric distribution]*

Mayo, D. G. 1996. Error and the growth of experimental knowledge. University of Chicago Press, Chicago. 493 pp. *[Mayo incorporates the hypothesis testing approach of Neyman & Pearson with Popperian falsificationism. Science advances through the falsification and the analysis of error. She also attacks Bayesian inference, but that is of secondary importance.]*

McCullagh and J. A. Nelder. 1989. Generalized linear models, 2nd edition. Chapman & Hall/CRC, Boca Raton Fl. *[A superb guide with a concise description on p 40-41 of how to fit the parameters of Generalized Linear Models. Dave Nichols of SPSS uses this book as one of the major references for the SPSS logistic regression procedures.]*

McCulloch, C. E. and S. R. Searle 2001. Generalized, linear and mixed models. Wiley Interscience. *[A great book covering the details of fitting generalized linear models and providing practical advice on which models are appropriate]*

Morrison, D. F. 1976. Multivariate statistical methods. McGraw-Hill, New York. *[This is a classic older reference on multivariate statistics, which covers Hotelling's T^2 and univariate repeated measures in Ch 4 and MANOVA, profile analysis, Greenhouse-Geisser and MANOVA in Chapter 5. Morrison provides justification for Sleuth's use of the T^2 test in Chapter 16 on p. 208: the only pattern required of the covariance matrix is that it is full rank.]*

Neter, J, M. H, Kutner, C. J. Nachtsheim and W. Wasserman. 1996. Applied linear statistical models. Irwin, Chicago. 1408 pp. with data diskette. *[A comprehensive guide to regression and ANOVA designs. This is the book that I go to first when I have questions about the regression models used in **Ramsey & Schafer (2002)**]*

Quinn, G. P. and M. J. Keough. 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press. 520 p. *[This is a tremendous book. If I weren't using Sleuth, I'd be using this as a text.]*

Scheiner, S. M. and J. Gurevitch, ed. 2001. Design and analysis of ecological experiments, 2nd edition. Oxford University Press, Oxford. 415 pp.*[Contains invited review chapters including reviews of power analysis, ANOVA, ANCOVA, repeated measures (Von Ende), and time series. SAS programs to perform the analyses provided]*

Singer, J. D. and J. B. Willett. 2003. Applied longitudinal data analysis: modeling change and event occurrence. Oxford University Press, Oxford. 644 pp. []

Sokal, R. R. and F. J. Rohlf. 1995. Biometry, 3rd Edition. W. H. Freeman & Co., New York. 887 pp. *[A top-notch guide to statistics with many biological examples. This text does a particularly good job with one-way ANOVA. and multiple-comparison tests]*[7]

Taylor, J. R. 1997. An introduction to error analysis: the study of uncertainties in physical measurements, 2nd edition. University Science Books, Sausalito CA 327 pp. *[This is a very good book on the propagation of error, an important topic for any natural scientist]*

Underwood, A. J. 1997. Experiments in ecology. Cambridge University Press, Cambridge. 504 p. *[Very strong on ecological applications of ANOVA, reviews his BACI design: Before-After Experimental interactions.]*

Winer, B. J., D. R. Brown, and K. M. Michels. 1991. Statistical principles in experimental design, Third Edition. McGraw-Hill, New York. 1057 pp. *[This book is EXCELLENT on ANOVA, perhaps the best source on ANOVA in the statistical literature. Hurlbert (pers. Comm.) Strongly recommends this book for ecologists. Chapter 4 on repeated-measures ANOVA compliments Chapter 16 in **Ramsey & Schafer (2002)** and provides the technical basis for the GLM/Repeated Measures statistics in SPSS]*

Zar, J. 1999. Biostatistical analysis, 4th Edition. Prentice Hall, Upper Saddle River, NJ. 663 p plus appendices and index. *[A very good introduction to most of the topics covered in EEOS611. Zar is much stronger in some areas, such as nonparametric tests, than **Ramsey & Schafer (2002)**]*

BACKGROUND TEXTS

For those who want to look at a few texts to refresh their knowledge of statistics and probability, here are several that I can recommend:

Cobb, G. W. 1997. Introduction to design and analysis of experiments. Springer, New York. 795 pp. *[This is a remarkable undergraduate or introductory graduate text that begins with factorial ANOVA in chapter 1 and then fills in the statistical foundation when needed to further understanding of experimental design.]*

Devore, J. L. and N. R. Farnum. 1999. Applied Statistics for Engineers and Scientists. Duxbury Press, Pacific Grove CA. 577 pp. *[This was the textbook used for EEOS601 in Fall 2000]*

Freedman, D, R. Pisani and R. Purves. 1998. Statistics, 3rd edition. Norton, New York. *[This is a wonderful introduction to probability and statistics. It is very elementary though, and the authors' avoidance of any equations really limits the book's usefulness]*

Gonick, L. and W. Smith. 1993. The cartoon guide to statistics. Harper Perennial, New York. 230 pp. *[This is a surprisingly useful little book for reviewing concepts needed for EEOS611]*

Larsen, R. J. and M. L. Marx. 2006. An introduction to mathematical statistics and its applications, 4th edition. Prentice Hall, Upper Saddle River, NJ . *[This is the latest edition of the text I used for EEOS601 in 2001 & 2003. Meng Zhou used the text in Fall 2006. I'd highly recommend it for the sections of probability. It isn't strong on inferential statistics nor does it have much discussion of random sampling and survey designs.]*

For Matlab®

Hanselman, D and B. Littlefield. 2001 Mastering MATLAB®: a comprehensive tutorial and reference. Prentice-Hall, Upper Saddle River NJ. *[This is a superb supplement to the extensive Matlab manuals. There are new editions every couple of years]*

Middleton, G. V. 2000. Data analysis in the earth sciences using MATLAB™. Prentice-Hall, Upper Saddle River, NJ. 260 pp. and diskette. *[An innovative text presenting MATLAB™ programs for regression, PCA, cluster analysis, ternary diagrams and other topics]*

Some well-written histories of statistics & statisticians

Bell, E. T. 1937. Men of mathematics. Simon & Schuster. *[Includes biographies of Legendre, Poisson & Gauss]*

Salsburg, D. 2001. The lady tasting tea: how statistics revolutionized science in the twentieth century. W. H. Freeman & Co., New York. 340 pp. *[This is a wonderful little book, with superb New Yorker style articles on Fisher, Wilcoxon, Deming and other giants of 20th century statistics]*

Stigler, S. M. 1986. The history of statistics: the measurement of uncertainty before 1900. Belknap Press, Cambridge.

Stigler, S. M. 1997. Statistics on the table. Belknap Press, Cambridge.

Class Logistics

TIME & LOCATION

The class will meet from 10 to 11:15 am on Monday & Wednesday in Presentation Room 3, Lower Level library.

The class format (**described below**) includes students presenting their statistical analyses and interpretations to their classmates at the start of each class.

OFFICE HOURS AND E-MAIL

Eugene D. Gallagher

My in-person office hours will be Monday & Wednesday immediately after class (11:30 am to 12:30 am in my lab W-3-043). Make an appointment by email in advance or check with me at the end of lecture if you want to meet with me.

I can also meet with students any time that I'm in my office. My office is in the Science Building, 1st Floor, Room 55. If possible, make an appointment in advance after class, by E-mail (**Eugene.Gallagher@umb.edu**) or by calling ([617] 287-7453). The easiest way to get a quick answer is by email. My fax is (617) 287-7474.

Angeliki Evgenidou, Teaching Assistant

The teaching assistants for this course is Angeliki Evgenidou, a doctoral candidate in EEOS. Angeliki's email is **Angeliki.Evgenido001@students.umb.edu**. Her fax number is the EEOS departmental fax: (617)287-7474. Her office is located in Dr. Gallagher's lab Wheatley-3rd Floor- Rooms 41-43, but Angeliki will be doing most of the grading during the early part of semester from her home.

WEBCT

WebCT/Vista 4 is the online software for the UMASS/Online system. It can be accessed here: **http://boston.umassonline.net/**

After you have registered for the course, you should be automatically registered for the WebCT course. WebCT will contain handouts with the problems assigned for each class, datasets, the slides shown in class, and SPSS syntax to solve the case studies.

Reading Assignments posted on WebCT/Vista

We'll be covering most of the content in **Ramsey & Schafer (2002)**. The reading assignments are provided in the **full outline** below. In the folders on WebCT are the handouts and assignments for each class. These will include:

- A pdf handout of the class slides, posted on WebCT at 9 am on class days
  - Answers to the homework problems to be discussed that day
  - The "Best of the Day" homework solutions.
  - Background information on the new topic
    - the case studies and their solution in SPSS
    - the new problems assigned for each chapter
- Electronic versions of the new data & SPSS syntax for the 2 new case studies to be introduced in class
- A shockwave movie. I've recorded 10 min to 20 min movies showing how each of the case studies is solved in SPSS. These are posted on WebCT.

WebCT discussion boards

There are discussion boards on WebCT for posting questions about the course and problems. The TAs and I will check these boards daily and post responses. During class, you will be graded on class participation, and part of class participation is checking and posting on the discussion boards. The discussion on the boards is informal. It is a spot for informal conversation about a topic, including statistics in the news.

SPSS files on WebCT

There is a CD that ships with Statistical Sleuth with the data files for all of the case studies and problems. The files are also available on the Sleuth web page (**see below**). I'll have updated versions of the files for the case studies and the assigned problems posted in files for each chapter on the WebCT server. Each one of the case studies will have a data file and syntax file posted on WebCT. The data files for most of the assigned problems will be posted on WebCT. These will always be available on the WebCT collaborative files section, and each file will be linked to the appropriate class session in the outline section.

Camtasia movies on how to solve the Case studies with SPSS

In previous offerings of EEOS611, I demonstrated how to solve the case studies in SPSS using the graphical user interface in SPSS and using syntax. This was an inefficient use of class time. For the 2006 class, I've made movies, using the Camtasia screen capture software, of the steps necessary to solve the case studies. The syntax for each of the case studies will be posted on WebCT. I'll post the movies in shockwave format. These 5-15 minute movies take about 1-2 minutes to download with a T1 internet connection.

"STATISTICAL SLEUTH" WEB PAGE

Ramsey maintains a web page at Oregon State for "The Statistical Sleuth." This page at **http://www.statisticalsleuth.com** contains the data from the book, available as ASCII, EXCEL, MINITAB, STATAQUEST, SAS (.sd2), SPSS, SAS transport (.xpt) (for JMP IN) files. There is also an Errata page, containing a few corrections to the book, and sample midterm and final exam questions. For input to SPSS, I've found the Excel files to be the best to use.

CLASS FORMAT

The class format that I use for this class is a modified form of a format used by Dr. Pete Jumars in his graduate course in non-parametric statistics at the University of Washington. I use this method because:

- I believe statistics is best learned by solving problems. During the semester, you'll solve about 2 computation problems per class.
- Students need experience not only choosing and using the appropriate statistical tests, but also presenting their results to others
- Students should also be able to answer questions about the statistical tests that they performed
- Students will be able to see the work submitted by other students. Peer pressure often greatly improves the quality of the analyses.
- As an instructor, I get immediate feedback about whether students have grasped the concepts involved.

During the third class, after the first homework assignment is graded, each 75-minute class will use the following format:

I. **Discussion of problems**: During the first half of class period **T**, the homework handed in at the beginning of class **T-1** will be discussed. I will have emailed your homework back to you by the morning of class, so check your email and bring a copy of your solution to class. The topics of these homework problems would have been introduced during the lecture in class period **T-2** (that is, one week earlier).
I will usually assign three computational problems for each class, and I'll pick at least three different students to present their solutions to the class. I will often choose two different students to present the same problem, if the approach they took was different. I'll notify students by 5 pm the day before class whether they are presenting the next morning. I'll pick student work based on these criteria: **1)** Exceptional work, with the best solution noted as **"Best of the Day"**, **2)** Innovative solutions, **3)** Campus location (at least one presentation will be from a distant location – Lowell, Dartmouth, or Amherst – for each class), and **4)** Student presentation frequency (with 18 students, each student should expect to present at least every 2-3 weeks). The students will be notified the day before class that they will be presenting. So, check your email or come to class prepared to present. I will make electronic versions of the graphics and bulleted lists for each student presentation. If a student wishes to make their own presentation that would be far preferable. Each student must present their solution in 5-10 minutes.

II. **New material:** During the 2nd half of each class, I'll introduce a new statistical topic. The textbook is set up so that there are two case studies per chapter. I'll introduce these case studies and show how they can be solved with SPSS. You should have read this material in advance so that you can ask questions about the topic. If you come to class unprepared to ask questions, then you will have missed your chance to ask me questions in person before you have to solve homework problems on the topic. There will be discussion sections set up in WebCT (**see below**), so you can post questions about the homeworks in WebCT that can be answered by me, the TA's, or your fellow students. You'll be assigned 3 problems to solve based on this new material.

The following table shows what students should be do before each class.

TABLE 2. CHECKLIST TO PREPARE FOR CLASS	
ITEM	**BEFORE CLASS**
Readings from Sleuth	Read the new material for class **T**. If possible, try to work out the two case studies in SPSS. The syntax for these problems will be posted on **WebCT** and there will be a movie posted showing how the case studies can be solved in SPSS.
Watch the Movie	There will be 10-20 minute movies posted on WebCT showing how to solve the case studies with SPSS.
Do Homework	Do the assigned homework for Class **T**, based on material presented during the last half of class **T-1**. Th e data are on the CD provided with the text, or you can obtain the data in SPSS format on **WebCT**. Homeworks must be posted on WebCT by Monday and Thursday at noon.
Prepare homework problem presentation	Check your email the afternoon before class. You may have been picked to present **your** solution to a homework problem handed in during class **T-1**. If you've been chosen, you'll be notified with a copy of your corrected homework. I'll make electronic versions of your graphics. If you want to make your own presentation slides, please do so, but you'll need to send an electronic form of your presentation by 9 pm the night before class.

GRADING

Table 3 shows how the allocation of points in the class.

Table 3. Grading. The grades will be continually updated on WebCT.

Area	Points in WebCT Grade Book	% of grade
Homework	225	45%
Class Presentations & Discussion	77	10%
Midterm exam	100	20%
Final Exam	100	25%
Total	502	100%

Homework

You will be responsible for submitting homework for nearly every class. The homework based on Monday's lecture will be due on Thursday at 10 am. The homework based on Wednesday's lecture will be due the next Monday at 9:50 am (before class). The homeworks should be posted electronically in the WebCT grade book. **Late homework will not be accepted.**

You will not be able to pass the class if you fail to turn in homework on time. The TA and I need to be able to grade problems on the day that they are handed in. I provide answer keys at the start of the next day's class, so there really is no way to give partial credit for homework that is more than 1 day late.

Collaboration on homework*

I encourage discussion among students about the phrasing of homework problems and possible methods of solution. However, the calculation of the solutions must be an individual effort. You can discuss the problems but the solution of the problems should be done individually. The written portions and interpretation of the results must be done individually.

Homework format

Homework solutions should be concise, usually 2 pages or less. Present the problem, the solution with presentation-quality graphics, the SPSS syntax if appropriate, and a strong conclusion. In

years past, some students have handed in 15-page data or mere printouts of 20 pages of SPSS output. Data dumps of this sort are inappropriate and won't be accepted.

Homework must be submitted in the WebCT gradebook in electronic form. I need to be able to make copies of your solutions for class discussion (see **class format** & **Student Presentations**), so electronic submission of your homework is essential. You can submit your homework as a Wordperfect document, a Word document, an rtf file, a pdf, or as an html file. I would strongly prefer receiving your homework as WordPerfect or Word documents. I can readily cut and paste from these formats. Post your electronic submission in the gradebook section of WebCT.

At the top of each homework, please list the number of hours that you spent on the homework. I do this for two reasons: 1) I want to keep the number of hours on each homework under 3 hours (6 hours per week for the class), and 2) if some students are spending an inordinate amount of time on homeworks, I may call them in to help them pick up some tricks on getting data into and out of SPSS. I'll try to produce a boxplot for each homework assignment to see whether the goal of under 6 hours is being met.

Homework grading

The TA and I will grade each homework on a 10-pt scale. Each solution should be a stand-alone document. You should clearly state the problem being analyzed, the statistical test, the results of that test, and your interpretation of the result. Again, this can almost always be done in 1-2 pages. You must submit your homework electronically in the WebCT gradebook. The grades and comments will be returned to you electronically.

Student presentations & a final project

At the start of every class, we'll have two to three student presentations of solutions to homework problems submitted in the previous class. Angeliki Evgenidou, the TA will pick the best solutions by student who haven't yet presented. There will be 40-45 student presentations during the semester, so if there are 15-20 students, you can expect to do 2 or at most 3 presentations. Angeliki & I will notify students on Friday at noon if they are to present on Monday. We'll notify students by Tuesday noon if a student is presenting in Wednesday's class.

If you are chosen to present, please post an electronic version of your presentation on WebCT by 9 am on class day, and link your file to that day's class's outline. In previous years, I made up the slides for the students, but that wasn't as effective as students making their own slides. Scientific meetings now insist on Powerpoint for presentations, so that would be the preferred format for Presentations. I will also accept Corel Presentations slides or pdf's.

All students will be expected to deliver a 10-minute talk on a research design of their choice during the **final three classes**. If there are more than 15 students in the class, we'll alter the lecture schedule to allow 3 or more classes for student presentations and discussions. Please read Chapter 23 for ideas on your experimental design.

During each class section, there will be opportunities for you to ask questions of your fellow students and the instructor. There will also be opportunities to post questions on the web and participate in discussions on WebCT.

Midterm & final Examinations

I will base 20% of your grade on the midterm examination, which will be a 75-min in-class examination on **Class 15, Weds, March 26, 2008** covering Chapters 1-8. For this exam, you may use your copy of Statistical Sleuth and class notes. Links to copies of previous midterms and finals will be posted on **WebCT**, but you should note that some previous midterms were take-home exams.

Statistics on the web

Table 4. Statistics on the web.

URL	Description	Site
http://www.bmj.com/collections/statsbk/index.shtml	Statistics at Square One Ninth Edition T D V Swinscow Revised by M J Campbell, University of Southampton. *[An overview of some statistical methods]*	British Medical Journal collections
http://members.aol.com/jeff570/mathword.html	Earliest Known Uses of Some of the Words of Mathematics	
http://members.aol.com/jeff570/stat.html	Earliest Uses of Symbols in Probability and Statistics	
http://name.math.univ-rennes1.fr/bernard.delyon/textbook/stathome.html or **http://www.statsoft.com/textbook/stathome.html**	Electronic Statistics Textbook (online)	
http://www.itl.nist.gov/div898/handbook/index.htm	**Engineering Statistics Handbook** - Includes a nice section on exploratory data analysis including the Anscombe quartet.	National Institute of Standards & Technology

Table 4. Statistics on the web.

URL	Description	Site
http://physics.nist.gov/Pubs/guidelines/contents.html or the pdf **http://physics.nist.gov/Pubs/guidelines/TN1297/tn1297s.pdf**	Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results by Barry N. Taylor and Chris E. Kuyatt	National Institute of Standards & Technology
http://davidmlane.com/hyperstat/	Hyperstat online statistics journal	
http://www.stat.uiowa.edu/~rlenth/Power/	Java applets for power & sample size	Russ Lenth University of Iowa
http://www.amstat.org/publications/jse/	Journal of Statistics Education (free on web)	
http://mathworld.wolfram.com/topics/ProbabilityandStatistics.html	Mathworld's Probability & Statistics	
http://www.pitt.edu/~wpilib/statfaq2.html	Rich Ulrich's FAQ. A collection of his posts from the statistical educator's user's group.	
http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/index.htm	SAS/STAT user's guide, including many references on methods	
http://www.ats.ucla.edu/stat/SPSS/	SPSS Resources from UCLA	
http://www.ats.ucla.edu/stat/spss/examples/default.htm	SPSS Textbook examples & data from UCLA	
http://www.ats.ucla.edu/stat/spss/webbooks/reg/default.htm	SPSS Web Books: Regression with SPSS (from UCLA)	
http://support.spss.com/default.htm	SPSS Support page, incuding white papers (log on as guest/pw guest if you don't want to register)	
http://ourworld.compuserve.com/homepages/jsuebersax/agree.htm#basics	Statistical methods for rater agreement	

Table 4. Statistics on the web.

URL	Description	Site
http://www.bmj.com/collections/statsbk/index.shtml	Statistics at Square One Ninth Edition T D V Swinscow Revised by M J Campbell, University of Southampton. *[An overview of some statistical methods]*	British Medical Journal collections
http://www2.chass.ncsu.edu/garson/PA765/statnote.htm	Statnotes: Topics in Multivariate Analysis, by G. David Garson	
http://www.statsoft.com/textbook/stathome.html	Statsoft electronic statistics textbook	
http://www.cmh.edu/stats/index.asp	Steve's attempt to teach statistics, including many definitions	
http://davidakenny.net/cm/causalm.htm	Structural equation modeling	
http://www.amstat.org/sections/SRMS/whatsurvey.html	Surveys (from the ASA)	
http://www.stat.ucla.edu/history/	UCLA's History of statistics	UCLA
http://www.ats.ucla.edu/stat/spss/library/spssmixed/mixed.htm	UCLA Mixed models (an SPSS html slideshow, available from SPSS, posted at UCLA)	UCLA
http://www.stat.ucla.edu/history/people/	UCLA Portraits of statisticians	UCLA
http://trochim.human.cornell.edu/kb/	William Trochim's Research Methods database	

Outline of the Course

I. Introduction to the course **(Class 1)**
   - A. Who should take the class
   - B. The textbook and class logistics
   - C. Topics to be covered during the semester
   - D. Introduction to WebCT & CENTRA
   - E. Brief introduction to SPSS on the PC

II. **Chapter 1** Drawing statistical conclusions (**1**) **(Class 2)**
- A. Case studies
  1. Motivation and Creativity
  2. Sex discrimination in employment
- B. Statistical inference and study design
- C. Measuring uncertainty in randomized experiments
- D. Measuring uncertainty in observational studies
- E. Related issues
- F. Summary

III. **Chapter 2** Inference using t-distributions (**28**) **(Class 3-4)**
- A. Case studies
  1. Bumpus's Data on Natural Selection — An observational study
  2. Anatomical abnormalities associated with schizophrenia — An observational study
- B. One-sample t-tools and the paired t-test
  1. The sampling distribution of a sample average
  2. The standard error of an average
  3. The t-ratio based on a sample average
  4. Unraveling the t ratio
- C. A t-ratio for two sample inference
  1. Sampling distribution of the difference between tow independent sample averages
  2. Standard error for the difference of two averages
- D. Inferences in a two-treatment randomized experiment
- E. Related issues
- F. Summary
- G. Exercises

IV. **Chapter 3** A closer look at assumptions (**56**) **(Class 5)**
- A. Case studies
  1. Cloud seeding to increase rainfall – A randomized experiment
  2. Effects of Agent Orange on troops in Viet Nam – An observational study
- B. Robustness of two sample t-tools
- C. Resistance of the two sample t-tools
- D. Practical strategies for the two sample problem
- E. Transformations of the data
  1. The Logarithmic Transformation
  2. Interpretation after a Log Transformation
  3. Other Transformations for Positive Measurements
- F. Related issues
- G. Summary
- H. Exercises

V. **Chapter 4** Alternatives to the t-tools (**85**) (**Class 6**)
- A. Case studies
  1. Space shuttle O-Ring Failures – An observational study
  2. Cognitive Load Theory in Teaching – A Randomized Experiment
- B. The rank-sum test
  1. The Rank Transformation
  2. The Rank-sum statistic
  3. Finding a p-value by normal approximation
  4. A confidence interval based on the rank-sum test
- C. Other alternatives for two independent samples
  1. Permutation tests
  2. The Welch t-test for comparing two normal populations with unequal spreads
- D. Alternatives for paired data
  1. The Sign test
  2. The Wilcoxon Signed Rank Test
- E. Related issues
  1. Practical and Statistical Significance
  2. The Presentation of Statistical Findings
  3. Levene's test for Equality of two variances
  4. Survey sampling
- F. Summary
- G. Exercises

VI. **Chapter 5** Comparisons among several samples (**113**) **(Class 7-9)**
- A. Case studies
  1. Diet restriction and longevity
  2. Benjamin Spock conspiracy trial
- B. Comparing any two of the several means
  1. An ideal model
  2. The pooled standard deviation
  3. t-tests and confidence limits for differences of means
- C. The one-way analysis of variance F-test
  1. The extra sum of squares principle
  2. the ANOVA table
  3. More applications of the extra sum of squares principle
- D. Robustness and model checking
  1. Robustness to assumptions
  2. Diagnostics using residuals
- E. Related issues
  1. Further illustrations of different sources of variability
  2. Kruskal-Wallis nonparametric ANOVA
  3. Random effects
  4. Separate confidence intervals and significant differences.
- F. Summary
- G. Exercises

1. Ex 5.23 *T. rex* temperature

VII. **Chapter 6** Linear combinations and multiple comparisons of means (**149**) **(Class 9 & 10)**
- A. Case studies
  1. Discrimination against the handicapped
  2. Sexual selection in swordtails
- B. Inferences about linear combinations of group means
- C. Simultaneous inferences
- D. Some multiple comparison procedures
- E. Related issues
- F. Summary
- G. Exercises

VIII. **Chapter 7** Simple linear regression: a model for the Mean **(174)** **(Class 11)**
- A. Case studies
  1. The Big Bang
  2. Meat Processing and pH
- B. The simple linear regression model
- C. Least squares regression estimation
- D. Inferential tools
  1. Tests and confidence limits for slope and intercept
  2. Describing the distribution of the response at some value of explanatory variable
  3. Prediction of a future response
  4. Calibration: Estimating the X that results in $Y=Y_o$ {See also **Draper & Smith 1998, Chapter 3**}
- E. Related issues
- F. Summary
- G. Exercises

IX. **Chapter 8** A Closer look at assumptions for simple linear regression **(206)** **(Class 12-14)**
- A. Case studies
  1. Island area and number of species – an observational study
  2. Breakdown times for insulating fluid under different voltages – a controlled experiment
- B. Robustness of least-squares inferences
- C. Graphical tools for model assessment
- D. Interpretation after log transformations
- E. Assessment of fit using the analysis of variance
- F. Related issues
- G. Summary
- H. Exercises

X. **Chapter 9** Multiple Regression **(235)** (**Class 14**)
- A. Case studies
  1. Effect of light on meadowfoam flowering – a randomized experiment
  2. Why do some mammals have large brains for their size – an observational study
- B. Regression coefficients
  1. The multiple linear regression model
  2. Interpretation of regression coefficients
- C. Specially constructed explanatory variables
  1. A squared term for curvature
  2. An indicator variable to distinguish between two groups
  3. Sets of indicator variables for categorical explanatory variables with more than two categories
  4. A product term for interaction
  5. A shorthand notation for model description
- D. A strategy for data analysis
- E. Graphical methods for data exploration and presentation
- F. Related issues
- G. Summary
- H. Exercises

XI. Midterm Exam (3/22/06 W) **(Class 15)**

XII. **Chapter 10** Inferential tools for multiple regression **(267)** (**Class 16**)
- A. Case studies
  1. Galileo's data on the motion of falling bodies – a controlled experiment
  2. The Energy costs of echolocation by bats – an observational study.
- B. Inferences about regression coefficients
  1. Least squares estimates and standard errors
  2. Tests and confidence intervals for single coefficients
  3. Tests for confidence limits for linear combinations of coefficients
  4. Prediction
- C. Extra-sum-of-squares *F* tests
- D. Related issues
- E. Summary
- F. Exercises

XIII. **Chapter 11** Model checking and refinement **(304)** **(Class 17)**
- A. Case studies
  1. Alcohol metabolism in men and women – an observational study
  2. The blood-brain barrier – a controlled experiment
- B. Residual plots
- C. A strategy for dealing with influential observations
  1. Assessment of whether observations are influential
  2. What to do with influential observations
- D. Case-influenced statistics
  1. Leverages for flagging cases with unusual explanatory variable values
  2. Studentized residuals for flagging outliers

3. Cook's distances for flagging influential cases
4. A strategy for using case influence statistics

E. Refining the model
1. Testing terms
2. Partial residual plots

F. Related Issues
1. Weighted regression for certain types of non-constant variance
2. Measurement errors in explanatory variables

G. Summary

H. Exercises

XIV. **Chapter 12** Strategies for variable selection **(338)** **(Class 18-19)**

A. Case Studies
1. State average SAT scores – an observational study
2. Sex discrimination in employment – an observational study

B. Specific issues relating to many explanatory variables
1. Objectives
2. Loss of precision
3. A strategy for dealing with many explanatory variables

C. Sequential variable selection techniques
1. Forward selection
2. Backward elimination
3. Stepwise regression
4. Sequential variable selection with the SAT data
5. Compounded uncertainty in stepwise procedures

D. Model selection among all subsets

E. Analysis of the Sex discrimination data

F. Related issues

G. Summary

H. Exercises

XV. **Chapter 13** The Analysis of Variance for Two-way classifications **(374)** (**Class 20-21**)

A. Case studies
1. Intertidal seaweed grazers – A randomized experiment
2. The Pygmalion effect in training programs – A randomized experiment

B. Additive and nonadditive models for two-way tables
1. The Additive Model
A Regression Parameterization for the additive two-way model
2. The Saturated, nonadditive model
3. A strategy for analyzing two-way tables with several observations per cell.
4. The analysis of variance F-test for additivity

C. Analysis of the seaweed grazer data
1. Initial assessment of additivity, outliers and the need for transformation
2. The analysis of variance table from the fit to the saturated model
3. The analysis of variance table for the fit to the additive model
4. Answers to specific questions of interest using contrasts

5. Answers to specific questions of interest using multiple regression with indicator variables

D. Analysis of the Pygmalion data
   1. Initial Exploration and check on the additive model
   2. Answering the question of interest with regression
   3. A closer look at the regression estimate of treatment effect
   4. The p-value in the randomization distribution

E. Related Issues
   1. Additivity and nonadditivities
   2. Orthogonal contrasts
   3. Randomized blocks and paired-*t* analyses
   4. Should insignificant block effects be eliminated from the model?
   5. Multiple comparisons
   6. An alternate parameterization for the additive model

F. Summary

G. Exercises

XVI. **Chapter 14** Multifactor studies without replication (**409**) **(Class 22)**

A. Case Studies
   1. Chimpanzees Learning Sign language – a controlled experiment Fouts (1973)
   2. Effects of ozone in conjunction with sulfur dioxide and water stress on soybean yield – a randomization experiment

B. Strategies for analyzing tables with one observation per cell

C. Analysis of the Chimpanzee learning times study

D. Analysis of the soybean data

E. Related issues
   1. Nested ANOVA

F. Summary

G. Exercises

XVII. **Chapter 15** Adjustment for serial correlation (**436**) (Class 23)

A. Case Studies
   1. Logging practices and water quality – an observational study
   2. Measuring global warming – an observational study

B. Comparing the means of two time series
   1. Serial correlation and its effect on the average of a time series
   2. The standard error of an average in a serially correlated time series
   3. The first serial correlation coefficient
   4. Pooling estimates and comparing means of two independent time series with the same first serial correlation

C. Regression after Transformation in the AR(1) model
   1. The serial correlation coefficient based on regression residuals
   2. Regression with filtered variables

D. Determining if serial correlation is present
   1. An easy large-sample test for serial correlation
   2. The nonparametric runs test
   3. The Durbin-Watson test statistic
E. Diagnostic procedures for judging the adequacy of the AR(1) model
   1. When is a transformation of a time series indicated
   2. The partial autocorrelation function (PACF)
   3. Bayesian information criterion
F. Related Issues
G. Summary
H. Exercises

XVIII. **Chapter 16** Repeated Measures (**462**) (**Class 24**)
A. Case Studies
   1. Sites of short- and long-term memory — A controlled experiment
   2. Oat Bran and cholesterol — A randomized crossover experiment
B. Tools and strategies for analyzing repeated measures
   1. Types of repeated measures studies
   2. Profile plots for graphical exploration
   3. Strategies for analyzing repeated measures
C. Comparing the means for bivariate responses in two groups
   1. Summary statistics for bivariate responses
   2. Pooled variability estimates
   3. Hotelling's T^2 statistic.
   4. Checking on assumptions
   5. Confidence ellipses
D. Related Issues
E. Summary
F. Exercises

XIX. **Chapter 17** Exploratory tools for summarizing multivariate responses (**497**) (Class 25)
A. Case studies
   1. Magnetic force on rods in printers
   2. Love and Marriage — an observational study
B. Linear combinations of variables
C. Principal components analysis
   1. The PCA train
   2. Principal components
   3. Variables suggested by PCA
   4. Scatterplots in PCA space
   5. The factor analysis model and PCA
   6. PCA usage
D. Canonical correlation analysis
E. Introduction to other multivariate tools
   1. Discriminant function analysis
   2. Multidimensional scaling
   3. Correspondence analysis

4. PCA and Empirical Orthogonal Functions (EOFs)

F. Summary

G. Exercises

XX. **Chapter 18** Comparisons of proportions or odds (**529**)

A. Case Studies
   1. Obesity and heart disease in American Samoa
   2. Vitamin C and the common cold

B. Inferences for the difference of two proportions

C. Inference about the ratio of two odds

D. Inference from retrospective studies

E. Summary

F. Exercises

XXI. **Chapter 19** More tools for tables of counts (**552**)

A. Case studies
   1. Sex role stereotypes and personnel decisions – a randomized experiment
   2. Death penalty and race of murder victim – an observational study

B. Population models for 2 x2 tables of counts
   1. Hypotheses of homogeneity and independence
   2. Sampling schemes leading to 2 x 2 tables
   3. Testable hypotheses and estimable parameters

C. The χ-squared test
   1. The Pearson χ-squared test for Goodness of Fit
   2. X-squared test of independence in a 2 x 2 table
   3. Equivalence of several tests for 2 x 2 tables

D. Fisher's exact test: the randomization (permutation) test for 2 x 2 tables
   1. The randomization distribution of the difference in sample proportions
   2. The hypergeometric formula for one-sided P-values
   3. Fisher's exact test for observational studies
   4. Fisher's exact test versus other tests

E. Combining results from several tables with equal odds ratios
   1. The Mantel-Haenszel Excess
   2. The Mantel-Haenszel test for equal odds in several 2 x 2 tables
   3. Estimate of the common odds ratio

F. Related issues
   1. r x c tables of counts
   2. Higher dimensional tables of counts
   3. Analysis of SUV fatalities & the Ford Explorer (new problem)

G. Summary

H. Exercises

XXII. **Chapter 20** Logistic regression for binary response variables (**579**)

A. Case studies
   1. Survival in the Donner Party – An observational study
   2. Birdkeeping and lung cancer – A retrospective observational study

B. The logistic regression model

C. Estimation of the logistic regression coefficients

D. the Drop-in-deviance test
E. Strategies for data analysis using logistic regression
F. Analysis of case studies
G. Related issues
H. Summary
I. Exercises

XXIII. **Chapter 21** Logistic regression for binomial counts (**609**)
A. Case studies
1. Island size and bird extinctions – an observational study
2. Moth coloration and natural selection – A randomized experiment
B. Logistic regression for binomial responses
C. Model assessment
D. Inferences about logistic regression coefficients
E. Extra-binomial variation
F. Analysis of moth predation data
G. Related issues
H. Summary
I. Exercises

XXIV. **Chapter 23** Elements of Research design (**669**)
A. Case study
Biological control of a noxious weed – a randomized experiment
B. Considerations for forming research objectives
C. Research design tool kit
1. Controls and placebos
2. Blinding
3. Blocking
4. Stratification
5. Covariates
6. Randomization
7. Random sampling
8. Replication
9. Balance
D. Design choices that affect accuracy and prediction
1. Attaching desired precision to practical significance
2. How to improve a confidence interval
E. Choosing a sample size
1. Studies with a numerical response
2. Studies comparing two proportions.
3. Sample size for estimating a regression coefficient
F. Steps in designing a study
1. Stating the objective
2. Determining the scope of inference
a. What experimental units will be used?
b. What are the populations of interest
3. Understanding the system

4. Deciding how to measure a response
5. Listing factors that can affect the response
6. Planning the conduct of the experiment
7. Outlining the statistical analysis
8. Determining the sample size

G. Related issues – a factor of four
H. Summary
I. Exercises

XXV. **Chapter 24** Factorial treatment arrangements and blocking designs

A. Case study
   1. Amphibian crisis linked to ultraviolet – a randomized experiment
B. Treatments
   1. Choosing treatment levels
   2. The rationale for several factors
C. Factorial arrangement of treatment levels
   1. Definition and terminology for a factorial arrangement
   2. The 2^2 factorial structure
   3. The 2^3 factorial structure
   4. The 3^2 factorial structure
   5. Higher order factorial arrangements
D. Blocking
   1. Randomized blocks
   2. Latin square blocking
   3. Split Plot designs
E. Summary
F. Exercises

XXVI. **Chapter 22** Log-linear regression for Poisson counts

A. Case studies
   1. Age and elephant mating success
   2. Treatment for epileptic seizures
B. Log-linear regression for Poisson responses
C. Model assessment
D. Inferences about log-linear regression coefficients
E. Extra-Poisson variation and the log-linear model
F. Further issues
G. Summary
H. Exercises

Lecture schedule & reading assignments

The pdf's for each slideshow will be posted on WebCT by 9 am on class days. The reading assignments are shown in Table 5 (page 30). The homework problems will be added soon.

TABLE 5. LECTURE SCHEDULE & READING ASSIGNMENTS

NO	DATE	TOPIC	REQUIRED READINGS	HOMEWORK DUE
1/23/05 M		**SEMESTER START**		
1	1/28/08 M	Introduction to the course; HW1 Assigned	**Syllabus (html)** **Syllabus (pdf)**	Read Syllabus & Chapter 1
2	1/30/08 W	Chapter 1 Drawing Statistical; Conclusions; ***HW2 Assigned***	Chapter 1	***HW1*** due 1/31/08 W 9:50 a) Log on to WebCT b) Send me an email with contact information: mailing address, home phone, type of statistical analysis you are particularly interested in c) Load Case 1.1 & 1.2 data on SPSS d) Watch Case 1.1 & 1.2 Camtasia movies
3	2/4/08 M	Chapter 2: Inference using t-distributions	Chapter 2 & **Sterne & Smith (2001)**	***HW2*** on Chapter 1 due 9:50 am 2/4/08 M **Ex. 1.16** in Sleuth Planet Distances and Order from the Sun Parts a through e **Ex 1.21** (p. 25 in Sleuth) Do the analysis of 5 papers as described - brief Pick the best of the five for a 2-3 paragraph description suitable for class presentation.

TABLE 5. LECTURE SCHEDULE & READING ASSIGNMENTS

NO	DATE	TOPIC	REQUIRED READINGS	HOMEWORK DUE
4	2/6/08 W	Chapter 2 (continued)	Chapter 2 (continued)	
5	2/11/08 M	Chapter 3: A closer look at assumptions	Chapter 3	
6	2/13/08 W	Chapter 4: Alternatives to the t-tools	Chapter 4	
	2/18/08 M	President's Day University Holiday		
7	2/20/08 W	**Chapter 5** Comparisons among several samples (1 of 2)	Chapter 5 (1 of 2)	Read Chapter 5: ANOVA **Conceptual exercises, Chapter 4**
8	2/25/08 M	**Chapter 5** Comparisons among several samples (2 of 2)	Chapter 5 (2 of 2)	Read Chapter 5: ANOVA **Conceptual exercises, Chapter 4** Post 1 comment & 1 reply on Ch 4 conceptual problems
9	2/27/08 W	**Chapter 5 (cont) Chapter 6 (start)** Linear combinations and multiple comparisons of means	Chapter 6	HW 7 due Read Chapter 6: Chapter 6 Linear combinations and multiple comparisons of means
10	3/3/08 M	**Chapter 6** continued	Chapter 6	HW8 due Thus
11	3/5/08 W	**Chapter 7** Simple Linear regression: a model for the mean	Chapter 7	HW 8 on Chapter 6 due

TABLE 5. LECTURE SCHEDULE & READING ASSIGNMENTS

NO	DATE	TOPIC	REQUIRED READINGS	HOMEWORK DUE
12	3/10/08 M	**Chapter 7** Simple Linear regression: a model for the mean **Chapter 8** A Closer look at assumptions for simple linear regression, including lack-of-fit	**Chapter 7 & 8**	HW9 on chapter 7 due Monday 3/12/07 9:50 am Extension to Thus 3/15/07 Noon 7.30 Brain activity in violin players. Also, estimate the expected neuronal activity and 95% confidence interval for a violin player with 15-years experience.
13	3/12/07 W	**Chapter 8** A Closer look at assumptions for simple linear regression, including lack-of-fit	**Chapter 8**	Readings for Monday I.Read Chapter 9
	3/16-3/23	Spring Break		
14	3/24/08 M		Chapter 8	
15	3/26/08 W	In Class midterm exam, Chapters 1-7	Covers Chapters 1-7	No homework due
16	3/31/08 M	**Chapter 9 Multiple Regression**	**Chapter 9**	No homework due
17	4/2/08 W	**Chapter 10** Inferential tools for multiple regression		HW11 On Chapter 9
18	4/7/08 M			

TABLE 5. LECTURE SCHEDULE & READING ASSIGNMENTS

NO	DATE	TOPIC	REQUIRED READINGS	HOMEWORK DUE
19	4/9/08 W			
20	4/14/08 M	Chapters 12 & 13		
21	4/16/08 W			
4/18/08 Th		Pass/Fail Deadline/Course Withdraw Deadline		
4/21/08 M		**Patriots Day, University Closed**		
22	4/23/08 W			
23	4/28/08 M			
24	4/30/08 W			
25	5/5/08 M			
26	5/7/08 W			
27	5/12/08 M			
28	5/14/08 W Last class			
5/19/08-5/23/08		**Final Exam Period. In-class Final Exam**		

References

Ramsey, F. L. and D. W. Schafer. 2002. The statistical sleuth: a course in methods of data analysis, 2nd Edition. Duxbury Press, Belmont CA, 742 pp & data CD. {3, 8, 9, 11, 31}

Sterne, J. A. C. and G. D. Smith. 2001. Sifting the evidence — what's wrong with significance tests? British Medical Journal *322*: 226-231. *[Available online at: http://bmj.bmjjournals.com/cgi/reprint/322/7280/226]{30}*

Index