Introduction to Biostatistics (PubHlth540)

UNIVERSITY OF MASSACHUSETTS at AMHERST
Department of Public Health
Division of Biostatistics and Epidemiology Program
Fall, 2009

Time:	Tu/Th 2:30-3:45 at		
Location:	1st Week:	9/8/2009	Morrill III Room 212
		9/10/2009	Library Tower, Room 1667.
	2nd Week:	9/15/2009	Morrill III Room 212
		9/17/2009	Library Tower, Room 1667.
	3rd+ Weeks:		Arnold 103
Instructor:	Ed Stanek		
Office:	401 Arnold House		
Phone:	545-3812		
Email:	stanek@schoolph.umass.edu		
Office hours:	Tu/Th 4-5:15, or by appointment.		

Course Home Page: http://courses.umass.edu/bioep540/

Required Text: Principles of Biostatistics, Second Edition (2000). Marcello Pagano and Kimberlee Gauvreau. Duxbury Press.

Course Summary and Objectives

This course provides an introduction to statistics, illustrating ideas with example and applications in the biologic and health sciences. The course aims to:

- enable the students to summarize data in a meaningful way
- help the student acquire introductory knowledge of probability and statistics as preparation for more specialized courses in the field
- enable the student to use and interact with statistical software for statistical analysis.

Upon completion of the course, the student should understand basic statistical principles, and be able to formulate, implement and interpret one sample and two sample statistical results.

Lecture Notes and Handouts

Lecture notes and handouts are posted on the course website (i.e. Topics). Lecture notes may be assigned as readings in some cases, with class time spent on problems/small groups/assignments.

Other Material

Readings: Additional readings may be available through the website, the electronic reserves at the Library, or over the WEB. Textbook readings are routinely assigned.

Weekly Problem Sets
A series of problem sets are assigned (usually weekly) and posted on the web site. All homework assignments are due on the date posted, but not all assignments will be collected and graded. Solutions to the problems will be posted after the due date. When assignments are collected, late assignments will not be accepted. Homework will require use of a hand held calculator, use of Excel, and use of SAS.

Grading/Policies

Homework	10%	
Reports/Exams (3 during the semester)	60%	(best 2 of 3)
Final Exam	30%	

Homework Policies

Collaboration with Others: You are encouraged to contact other course members, discuss homework problems and computer problems, form study groups, and work together. However, the assignment that you turn in should be your work, and not copied from another's assignment. A grade of zero will be assigned to assignments if plagiarism occurs. Ask if you are not sure whether or not something is appropriate.

Academic Honesty: As a student, you are required to follow the University Academic Honesty policy (www.umass.edu/umhome/policies/honesty.html)

Policy and grading reports*:* All assigned reports will be due on the date specified, usually 1 week after they are assigned. Late reports will not be accepted and no credit will be given.

The report must conform to the following standards:

- Typed. Tables and figures may be printed separately from the text and pasted in the appropriate place. Formulas and special characters may be handwritten, if necessary.
- Tables and figures must be neat and clearly labeled.

Reports will be graded on a 100-point scale according to the following criteria:

- Quality of the report including introduction, description of methods, results and conclusion 60 points
- Numerical accuracy 20 points
- Overall appearance and style 20 points

Policy on Exams: Any books, notes or calculators or computers may be used. The hour exams will cover material as announced, usually through the last homework assignment that was due. All exams are cumulative. Hand held calculators will be needed for some exams.

Other Procedures: All homework assignments/exams not picked up in class may be obtained from the slot for graded homework for this section in a box in the hall on the fourth floor of Arnold house (next to Arnold 407). Copies of class handouts not picked up in class may also be obtained from the same box.

Final Exam: The final exam will cover the entire course and will take place during the final exam period at the time scheduled by the University. Please make travel plans after the final exam has been scheduled.

Calculators and Computers

All students will be expected to use hand calculators and computers. Use of computers is an important aspect of this course. Students need to use Excel and SAS, which will be illustrated in class. Alternative software may be used but will not be discussed in class.

Calculators*:* An electronic calculator is required to do calculations for homework and exams. An inexpensive hand held calculator will suffice. The calculator should have a "summation" key and be able to perform simple statistical functions (such as calculate the mean and standard deviation). The calculator should also have a "ln" and "exp" buttons to perform natural logarithm and exponentiation computations, and some memory. Inexpensive calculators (< $20) that are adequate for this course include the Sharp Models EL506, EL509, EL520, EL531, and the Texas Instrument TI30X, TI36X. They are available at the University Store on the UMASS-Amherst campus and elsewhere.

Computer Access*:* All students must have PC-compatible computer access and a UMass-Amherst Email account. Students may obtain a UMass-Amherst Email account from the Office of Information Technology (OIT) (Lederle Lowrise). This account will enable OIT computer lab use at the UMASS-Amherst campus.

Computer Labs at UMass-Amherst: Hours for OIT Computer labs are posted in the labs.

Access to the School of Public Health Computers at UMass-Amherst: Computers in the School of Public Health are located in Arnold House, rooms 413 and 417. The rooms are locked with a combination door lock. The combination can be obtained from the instructor. Arnold House is open from Monday-Friday until 9:30 PM. Late evening and weekend access is possible by requesting a building key from Ms. Gloria Seaman, Arnold 330.

Course Outline and Objectives

After completing the course, the student should be able to do the following:

I. Introduction

- recognize the need to summarize data
- differentiate between a summary of data and statistical estimate
- recognize the different goals of hypothesis testing and estimation

II. Summarizing Data Using the Computer and Word Processors

- deduce the scale of measurement for a variable
- choose and apply appropriate descriptive visual data displays including bar charts, histograms, and cumulative distributions, and use statistical software to summarize data
- calculate measures of location and spread using summation notation
- manipulate summation notation, and vector notation
- organize a set of descriptive figures and tables for some data
- write a descriptive report to accompany some data

III. Simulate and Describe Samples of Selected from Distributions and Populations

- generate sets of observations from a specified distribution
- describe characteristics, such as the mean, standard deviation, and characteristics of the data in the set.

IV. Use rates in a set to evaluate rates in a larger population accounting for differences in the rates.

- represent the rates via a simple non-stochastic model
- construct summary standardized rates using indirect and direct standardization

VI. Probability and Applications

- define probability, sample space, sample points, and events
- apply probability terminology to describe a sampling distribution
- define joint probability and marginal probability
- use independent events to construct a tree diagram

VII. Populations, Samples, and Random Variables

- define and explain the terms population, parameter, random variable, subset, probability sample, and statistic
- illustrate the sampling distribution for a statistic using simple random with and without replacement sampling, and generate independent samples using software
- discuss the properties of a sampling distribution with appropriate terminology
- describe the role of the sampling distribution in estimation
- evaluate the role of the sample size on the shape of the sampling distribution for the sample mean, and statistics
- classify a sample by the sampling method, differentiating systematic, cluster, and stratified samples

VIII. More Probability

- use conditional probability to define sensitivity and specificity
- apply tree diagrams for independent yes/no variables to obtain a binomial distribution
- define and use sensitivity, specificity, positive predictive value
- describe Bayes theorem and apply it to a simple setting.

IX. The Normal Distribution and Other Distributions

- understand and use normal distribution tables for the standard normal distribution and other normal distributions
- the central limit theorem
- use standardization to convert normally distributed variables to a standard normal distribution, and reverse for determining percentiles
- justify, and criticize the use of the normal distribution when variables are binomially distributed

X. Estimation

- calculate a point and interval estimate for a target parameter in a variety of settings
- describe principles for determining a 'best' estimator. Describe and illustrate properties of estimators in different settings.
- put into your own words the interpretation of an interval estimate
- write an interpretation of a point and interval estimate for a non-technical audience
- discuss the importance of a pivotal quantity in distributions interval estimation
- critically discuss an application of estimation in an article

- discuss the information needed and be able to calculate the sample size needed to obtain a confidence interval of a specified width
- apply rules of expectation to determine the mean and variance of a linear combination of random variables

XI. Hypothesis Testing

- describe terminology used in hypothesis testing, including the null and alternative hypothesis, one-sided and two sided tests, Type I and Type II error, power and p-values
- analyze data from one or two samples using a hypothesis test
- evaluate the Type II error and power for a hypothesis

XII. Contingency Tables and Chi Square Tests

- analyze contingency table data using a chi square test

XIII. Correlation and Regression

- calculate and interpret the covariance and correlation between two random variables
- construct and interpret a scatter plot
- explain the statistical assumptions that underlie a simple linear regression
- discuss the least squares criteria commonly used to estimate regression coefficients
- calculate estimates of the slope and intercept for a simple linear regression

XIV. Repeated Measures

Additional References:

Basic Statistics for the Health Sciences, Jan W. Kuzma, 1984, Mayfield Publishing Company. (excellent readable introduction) Biostatistics, Rosner, 1984.(Good medical examples)

Introduction to Statistical Analysis, Dixon, W. J. and Massey,F.J.,1983, McGraw Hill. (classic reference, but limited explanations)

Think and Explain with Statistics, Moses, L.E., 1986, Addison-Wesley Publishing.

Biostatistics for the Health Sciences, Daniel, W. 1991, 5th Edition.

Elementary Statistics, Neil A Weiss, Addison-Wesley Publishing Company (1993) 2nd edition. (Used in Res Ec 211, Introductory Statistics taught by Dr. Morzuch)

Statistics, Freedman, Pisani, Purves, and Adhikari (1991) W.W. Norton and Company, Inc. 2nd edition. (Used in Statistics 111, by Dr. Lane)

Statistical Thinking for Managers, Hilderbrand, D. K. and Ott, Lyman , PWS-Kent, 3rd edition (used in Statistics 140)

Doing Social Research, Baker, Therese L., McGraw-Hill, Inc., New York, 2nd Edition (used in Sociology 211 by Dr. Fisher)

Elementary Statistics, Johnson, Robert (1992), Duxbury Press (6th Edition) (Used in Psych 240A, Statistics in Psychology by Drs. Berthier and Reichle, and Well)

Statistics, Hays, William L., 5th Edition (used in Soc 711 by Dr. Anderson)