Public Health 302
Introduction to Biostatistics
COURSE
SYLLABUS
Fall Quarter, 2006

Schedule:	Thursdays, 6:00 - 9:00pm Robert H. Lurie Medical Research Center, Baldwin Auditorium
Lecturer:	James M. Sinacore, Ph.D. Dept. of Preventive Medicine and Epidemiology Loyola University Stritch School of Medicine 2160 South First Avenue Maywood, IL 60153
Office Hours:	Monday through Friday 10:00am - 4:30pm Work; (708) 327-9127 / FAX; (708) 327-9009 / Home; (708) 343-2263 e-mail; JSinacore@lumc.edu
Associate Lecturer:	Daniel Garside Manager, Research Programming Northwestern University, Department of Preventive Medicine Work; (312) 908-6616 / FAX; (312) 908-9588 e-mail; dgarside@northwestern.edu
Teaching Assistant:	Jocelyn Hirschman Sinai Urban Health Institute Mount Sinai Hospital Work; (773) 257-5329 / FAX; (773) 257-5680 e-mail; hirj@sinai.org
Required Texts:	*Statistical Analysis: An Introduction to Univariate and Multivariate Methods* by Sam Kash Kachigan. Radius Press. *More Damned Lies and Statistics: How Numbers Confuse Public Issues* by Joel Best. University of California Press.
Other Readings:	As assigned.

*... de cetero, fratres, **quaecumque sunt vera**, quaecumque pudica, quaecumque iusta, quaecumque sancta, quaecumque amabilia, quaecumque bonae famae -- si qua virtus, si qua laus, haec cogitate.*

Welcome to Introduction to Biostatistics, Public Health 302. My hope is that your experience in this course will be pleasurable as well as enlightening. I realize that Latin phrases are not commonly exchanged among scholars these days, but I felt compelled to share the above with you. Here is the translation:

> *Finally, brethren, **whatever is true**, whatever is honorable, whatever is just, whatever is pure, whatever is lovely, whatever is gracious, if there is any excellence, if there is anything worthy of praise, think about these things. (RSV)*

This is an excerpt from St. Paul's Letter to the Phillipians (4:8) and I have displayed it for two reasons. First, I sincerely hope that your experience in this class will help you recognize that statistical analysis possesses a beauty and is a thing worthy of praise. Second, the phrase *quaecumque sunt vera* is one of the mottos of Northwestern University (the other is in Greek and says *The Word ... is full of grace and truth,* taken from the Gospel of John, 1:14). Given that our lives are immersed in a politically dense, data oriented, technical world that distracts us from thinking about whatever is true, honorable, just, pure, lovely, and gracious, I think that it is good for us to stop and realize that Northwestern was founded with these high ideals in mind. It will probably come as a surprise to many that the Northwestern mottos are based in Holy Scripture, given that the majority of modern day science, both here and elsewhere, is notably divorced from Revelation.

Although I hope to transmit a sense of beauty, I am patently aware of the need to be pragmatic. There is, of course, no reason to believe that we can not strike a balance between philosophy and pragmatism. Hence, it is my aim to help students acquire skills and knowledge that will enable them to manage basic statistical tasks that confront researchers in the health domain. From this, students will be better prepared to interpret research journal articles and to conduct their own data analysis for future research projects. Given this, the general course objectives for this quarter are as follows.

General Course Objectives

By the end of this class, students should be able to:

- $ classify statistical analysis as a collection of techniques that are intended to uncover information in data.
- $ discuss how statistical methods are used as a research tools.
- $ describe the purpose and proper use of the statistical techniques presented in class.
- $ define basic statistical concepts.
- $ conduct statistical tests of research hypotheses using methods presented in class.

Student Evaluation

Homework Assignments.
For this course, students will be evaluated in terms of their performance on a series of homework assignments that will be distributed during the quarter. These assignments will be evaluated critically; hence, ***students should be neat, careful, and above all comprehensive on these assignments***.

Students will submit homework assignments for grading on a prescribed basis, normally having had one week to complete the work. Each assignment will receive a numerical point value. Acquired points for each assignment will be displayed on Blackboard.

It will be important for everyone to complete assignments on time. Remember, late assignments are a burden on students as well as the instructional staff. Thus, it is important for everyone to be timely in completing and submitting work throughout the quarter. ***A point penalty will be incurred for late assignments.***

Examinations.
There will be a midterm and final exam scheduled during this course. These tests will be designed to assess students' knowledge of statistical concepts, their skill in calculating and interpreting statistical computations, and their knowledge gained from the supplementary course readings. Each exam will cover different content; hence, the final exam will **not** be cumulative.

Computer Assignments.
At a prearranged time outside of class, students will meet with Dan Garside to learn how to conduct basic operations in data base management and statistical analysis with STATA. There will be three computer assignments that will be recorded as completed or not completed. One missing computer assignment will lower one's final grade by one grade stratum (e.g., from A to A-). Two missing computer assignments will lower one's final grade by two grade strata (e.g., from A to B+). Three missing computer assignments will lower one's final grade by three grade strata (e.g., from A to B).

Course Grades

A grade in this course will be determined from one's performance on the homework assignments and exams. Students will acquire points in each of these areas and a percentage will be computed for each. One=s overall performance will then be computed according to the following weighted average:

$$Performance = \frac{3(Exam\,\%) + 1(Assignment\,\%)}{4}$$

In essence, performance on exams is given three times the weight as that for assignments. This

is equivalent to saying that 75% of one=s grade is determined by exams and that 25% is determined by assignments. Keep in mind that one=s final grade will include information from the midterm and the final as well as all homework assignments. Performance averages will be assigned grades according to the following scheme:

Performance % Average	Grade
92 -100	A
89 - 91	A-
85 - 88	B+
82 - 84	B
79 - 81	B-
75 - 78	C+
72 - 74	C
69 - 71	C-
65 - 68	D+
62 - 64	D
Below 62	F

Remember, adjustments to this will made in the event of missing computer assignments as mentioned above.

Computing Aids

It goes without saying that students are highly encouraged to use hand held calculators for this course. In fact, this course may be the opportunity where some students will finally learn to decipher those cryptic symbols on their calculators! Bring your calculators to class. You also are welcome to bring a laptop computer to class, if you have one. ***Keep in mind, however, that I do not want laptops used during exams.***

Your Responsibility

Please be aware that it is your responsibility to keep all of your returned assignments. At the end of the quarter students will be able to check with the teaching assistant to be sure that point totals have been accurately recorded. We are extremely careful about recording student data, but

mistakes happen. Be good to yourself. Take the time to keep track of your assignments throughout the quarter.

I will always entertain disagreements from students regarding grades. However, everyone must substantiate disagreements with documentation. Every student's official grade status be that which is recorded in the class record unless there is documentation to suggest otherwise. Remember, I am on your side, but I expect that you will take as much responsibility for your grade as I do.

Course Evaluation

The Programs in Public Health administer web-based course evaluations to students for each course near the end of the quarter. ***Your completion of both the unit (course) and faculty evaluation components is required; failure to complete either of the evaluations will result in an incomplete grade until the evaluations are submitted.*** You will be sent the web link and instructions via email later in the quarter. You will have about two weeks time to complete the evaluations before grades are submitted.

A Postscript

I approach graduate level education with the expectation that students manage their own learning. Hence, do not be afraid to confront me or to ask challenging questions. Be assertive. Make sure you get from me what you need for this class. I expect that you will view Public Health 302 as ***your*** course.

∪∪∪

CLASS TOPICS

September 21 Exploratory Data Analysis: Stem-and-Leaf Diagrams and Box Plots

> The growing awareness of statistical equations and their related probability values have had a major impact on the way scientists think about the analysis of data. Unfortunately, many people have placed more attention on the technical aspects of statistical computations than on procedures for scrutinizing their data. The precipitous growth of high speed digital computers has only made this problem worse.

In tonight's class we will review basic techniques of exploratory data analysis (EDA) that have been designed by John W. Tukey. Specifically, we will learn about stem-and-leaf diagrams and box plots. These techniques are foreign to some, but they are invaluable for uncovering information in data. Tonight we also will discuss the main tenets of EDA: resistance, re-expression, residuals, and revelation.

In preparation for this class read:

Dawes, R.M. (1993). Finding guidelines for tough decisions. *Amstat News*, No. 203, 3,4.

Sinacore, J.M., Chang, R.W., & Falconer, J. (1992). Seeing the forest despite the trees: The benefit of exploratory data analysis to program evaluation research. *Evaluation and the Health Professions, 15,* 131-146.

Velleman, P.F., & Hoaglin, D.C. (1981). Applications, basics, and computing of exploratory data analysis. Boston: Duxbury. Chapter 1 "Stem-and-Leaf-Displays" pp.1-16.

Velleman, P.F., & Hoaglin, D.C. (1981). Applications, basics, and computing of exploratory data analysis. Boston: Duxbury. Chapter 3 "Boxplots" pp.65-73.

September 28 Frequency Distributions and Measures of Central Tendency & Variability

In order to analyze data from a scientific study, the researcher begins to organize numerical observations in a meaningful way by examining the manner in which scores are distributed. Therefore, in tonight's class we will learn a method for constructing a frequency distribution. Specifically, we will learn about frequency tables, histograms and polygons.

In addition, we will study the properties of frequency distributions. We will cover the fundamental property of central tendency, which reflects the magnitude of the numbers in a group of measurements. Central tendency is typically expressed as some type of "average." Different types of averages will be discussed. We also will explore the property of variability. This is defined as the extent to which measurements differ from central tendency. In other words, variability is a measure of the degree to which measurements deviate from the "average." In tonight's class we will discuss the standard deviation.

In preparation for this class read:

Kachigan; Chapters 1, 2, 3, 4, & 5 (sec. 1-5, 8).

Best; "Preface: People Count" & Chapter 1 "Missing Numbers"

Lord, F.M. (1953). On the statistical treatment of football numbers. *American Psychologist, 8*, 750-751.

October 5 Relative Standing: Percentiles, z Scores and the Gaussian Distribution

There are times in which we need to know how one measurement in a distribution relates to all the others. For example, we are pleased to know when our score on an exam is better than that of 90% of our peers. Therefore, in tonight's class we will learn how to compute percentile points and ranks when we are working with a sample or group of measurements. We also will learn a method of computing percentile points and ranks when we are working with a distribution of scores that is normal (i.e., bell shaped) in form. Our class will cover z-scores and the Gaussian (aka the normal) distribution.

In preparation for this class read:

Kachigan; Chapter 5 (sec. 6, 7).

McCall, R. (1990). *Fundamental statistics for the behavioral sciences.* New York: Harcourt, Brace, and Jovanovich. Chapter 4 Measures of Relative Standing, pp. 92-102.

Best; Chapter 2 "Confusing Numbers"

October 12 The Logic of Hypothesis Testing: The One-Sample z and t Tests

Scientists are in the business of making educated guesses about the world. These guesses are called hypotheses and are usually made in the context of a scientific study. In tonight's class we will discuss how scientists rely upon statistical analysis to help them test their hypotheses. The primary focus of this class will be upon the logic or steps of hypothesis testing. We will begin to learn about this process by discussing the z test. This will be a very important class because the logic of hypothesis testing underlies the work for the remainder of the course. We will extend our discussion of hypothesis testing by learning about sampling distributions. An extension of the z test and Student's t test for a single sample will be presented. We also will learn about confidence intervals, which are formed by the equations used in hypothesis testing.

In preparation for this class read:

Kachigan; Chapters 7 (sec. 1-5), 9 (sec. 1-3) then read 7 (sec. 8), & 9 (sec. 7).

Best; Chapter 3 "Scary Numbers"

Tukey, J.W. (1969). Analyzing data: sanctification or detective work? *American Psychologist, 24,* 83-91.

October 19 Student=s t-Tests for Independent and Related Samples

The simplest form of an experiment compares two groups -- one that has received a treatment and one that has not. In tonight's class we will learn how scientists test hypotheses within this type of design. Two forms of Student's t test will be presented. The first assumes that the two groups are comprised of different individuals. The second assumes that the same set of individuals have participated in two experimental conditions. We also will discuss confidence intervals for the difference in means.

In preparation for this class read:

Kachigan; Chapter 7 (sec. 6, 7); 9 (sec. 5, 6) then read 7 (sec. 9); & 9 (sec. 8).

Best; Chapter 4 "Authoritative Numbers"

Skipper, J.K., Guenther, A.L., & Nass, G. (1967). The sacredness of .05: A note concerning the uses of statistical significance in social science. *American Sociologist, 2,* 16-18.

O'Brian, T.C., & Shapiro, B.J. (1968). Statistical significance --What? *Mathematics Teacher, 61,* 673-676.

Kimmel, H.D. (1957). Three criteria for the use of one-tailed tests. *Psychological Bulletin, 54,* 351-353.

Goldfried, M.R. (1959). One-tailed tests and Aunexpected@ results. *Psychological Review, 66,* 79-80.

October 26 Midterm Exam

November 2 Analysis of Variance

There are times when we need to employ three or more groups in our research. Using Student=s t- test to compare these groups could result in a large number of tests, leading to statistical error. For example, we would need to conduct six t-tests if we wanted to examine all pairwise comparisons among four groups. In tonight=s class we will discuss the principles of the analysis of variance (ANOVA), which we can use to analyze data from multiple research groups. In addition, we will see how ANOVA can be used to analyze interactions, which reveal how two or more factors influence the variable in our study.

In preparation for this class read:

Kachigan; Chapter 12 (sec. 1-5, 8 App. 304 to the top of 311,@ 9).

November 9 Linear Correlation and Regression

The most fundamental activity of a scientist is to look for relationships. To help us with this, we have the techniques of statistical correlation and regression which provide a methodology that scientists find very useful. In tonight's class we will study the origin of the Pearson correlation coefficient and will learn why it cannot range beyond $\pm$ 1.00 (you might have wondered about this if you had a statistics class in the past).

One of the main goals of regression analysis is prediction. This follows directly from correlation. If we know that two variables are related, then it makes sense that we can predict one variable from our knowledge of the other. In tonight's class we will explore the rudiments of this type of prediction.

In preparation for this class read:

Kachigan; Chapter 5 (sec. 9), 10 (sec. 1-8, 10), 11 (sec. 1-7, 15, 16).

Sherin, K.M., Sinacore, J.M., Li, X-Q., Zitter, R.E., 7 Shakil, A. (1998). HITS: A short domestic violence screening tool for use in a family practice setting. *Family Medicine, 30,* 508-512.

Best; Chapter 5 "Magical Numbers"

November 16 Chi Square Test For Goodness of Fit and Association

The tests of significance that we have discussed to this point have all required interval or ratio level data. There are times, however, when our data are nominal in nature. The chi-square test is uniquely designed to help scientists test hypotheses with data that are at this level.

In preparation for this class read:

Kachigan; Chapter 13 (sec. 1-6).

Best; Chapter 6 "Contentious Numbers"

Schulman, J.L., Kupst, M.J., Suran, B.G. (1976). The worship of "p": Significant yet meaningless research results. *Bulletin of the Menninger Clinic. 40,* 134-143.

Bross, I.D.J. (1964). Prisoners of jargon. *American Journal of Public Health, 54,* 918-927.

November 23 Thanksgiving B No Class

November 30 Statistical Power and Sample Size Estimation

Our final class will be devoted to the topic of sample size estimation. In the past, this aspect of data analysis was sorely neglected. It has, however, recently been receiving proper attention among scientists. In this class we will discuss the concept of effect size and Type I and Type II errors. The focus of the class is to help students learn about the type of information they need to bring a statistician so that they can avoid the blind question AHow many subjects do I need?@

In preparation for this class read:

Kachigan; Chapter 9 (sec. 4 , 9, 10).

Best; Chapter 7 "Toward Statistical Literacy?"

Tversky, A., & Kahneman, D. (1993). Belief in the law of small numbers. In G. Keren & C. Lewis (eds). *A handbook for data analysis in the behavioral sciences: Methodological issues*, pp. 341-349, Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Prentice, D.A., & Miller, D.T. (1992). When small effects are impressive. *Psychological Bulletin, 112,* 160-164.

December 7 Final Exam