Schedule: Thursdays, 6:00 - 9:00pm, unless otherwise noted
Robert H Lurie Research Center (250 E. Superior), 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:
Jessica Jakubowski
Senior Research Data Analyst
Northwestern University
Work; (312) 503-3519 / FAX; (312) 503-3535
E-mail: j-jakubowski@northwestern.edu

Teaching Assistant:
Frank Aguilar
Phone; (312) 695-4481
E-mail: frankaguilar2013@u.northwestern.edu

Required Texts:
Fundamentals of Biostatistics (7th edition) by Bernard Rosner. Thomson

Other Readings: As assigned.
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 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). The seal of the university expresses the values of its founders (one of whom was physician John Evans, after whom Evanston is named) and recalls the university’s Methodist roots. Given that our lives are immersed in a politically dense, data driven, 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 the values upon which Northwestern University was founded. 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 Operations**
At a prearranged time outside of class, students will meet with Jessica Jakubowski to learn how to conduct basic operations in data base management and statistical analysis with STATA. There will be no computer lab assignments, per se, but students will be shown each week how to run STATA for all of the assignments throughout the course.

**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:

\[
Course\ Average = \frac{3 \times Exam\% + Assignment\%}{4}
\]
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:

<table>
<thead>
<tr>
<th>Course Average</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>92 - 100</td>
<td>A</td>
</tr>
<tr>
<td>89 - 91</td>
<td>A-</td>
</tr>
<tr>
<td>85 - 88</td>
<td>B+</td>
</tr>
<tr>
<td>82 - 84</td>
<td>B</td>
</tr>
<tr>
<td>79 - 81</td>
<td>B-</td>
</tr>
<tr>
<td>75 - 78</td>
<td>C+</td>
</tr>
<tr>
<td>72 - 74</td>
<td>C</td>
</tr>
<tr>
<td>69 - 71</td>
<td>C-</td>
</tr>
<tr>
<td>65 - 68</td>
<td>D+</td>
</tr>
<tr>
<td>62 - 64</td>
<td>D</td>
</tr>
<tr>
<td>Below 62</td>
<td>F</td>
</tr>
</tbody>
</table>

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 use a laptop computer for class and for exams. **Keep in mind, however, that I do not want anyone to run statistical programs on a laptop during exams.**

Your Responsibility

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

Academic Integrity

Every Northwestern faculty member and student belongs to a community of scholars where academic integrity is a fundamental commitment. The Program in Public Health abides by the standards of academic conduct, procedures, and sanctions as set forth by The Graduate School at Northwestern University. Students and faculty are responsible for knowledge of the information provided by The Graduate School on their Web page at [http://www.tgs.northwestern.edu/academics/academic-services/integrity/index.html](http://www.tgs.northwestern.edu/academics/academic-services/integrity/index.html)

Academic misconduct includes, but is not limited to

1. Receiving or giving unauthorized aid on examinations or homework
2. Plagiarism
3. Fabrication
4. Falsification or manipulation of academic records
5. Aiding or abetting any of the above

The PPH follows The Graduate School’s procedure for evaluating alleged academic misconduct, as outlined on the TGS website. [http://www.tgs.northwestern.edu/academics/academic-services/integrity/dishonesty/index.html](http://www.tgs.northwestern.edu/academics/academic-services/integrity/dishonesty/index.html)

Faculty reserve the right to use the “Safe Assignment: Plagiarism Detection Tool” that is part of the Course Management System to evaluate student assignments. Information about this tool can be found at [http://www.it.northwestern.edu/education/course-management/support/assessments/safeassignment.html](http://www.it.northwestern.edu/education/course-management/support/assessments/safeassignment.html)
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

Sep 27  Exploratory Data Analysis: Stem-and-Leaf Diagrams and Boxplots

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:


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. We began to discuss this last week with stem-and-leaf diagrams and boxplots. However, in tonight's class we will cover other methods for descriptive analysis. To begin, we will learn about frequency tables, histograms and polygons.

Following this, 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 often 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. We also will learn about three additional EDA techniques that are designed to reveal information in data: smoothing, letter value displays, and coded tables.

In preparation for this class read:
Rosner, Chapter 2.
Best; "Preface: People Count" & Chapter 1 “Missing Numbers.”

There is little in life that is certain. Much of what we experience is influenced by factors that are unknown or, at least, out of our control. This especially is true of tests of statistical significance. The coveted “p” value that goes along with every statistical test always reminds us that the result has a certain level of probability -- a fact that most investigators do not recognize. We will begin tonight’s class by discussing basic ideas in probability. The objective will not be to learn how to compute probabilities in all types of situations, but to learn that probability is the essence of hypothesis testing.

_Caution! Chapter 3 contains many technical statements. Please read with the objective of learning basic ideas. We will cover specifics in class._
The goal of the second part of class will be to learn how to use the Gaussian or normal distribution to compute the probabilities (i.e., the \( p \) values) of numerical outcomes that are continuous in nature. We will learn about the use of standardized (i.e. \( z \) scores and their application to variables that are normally distributed. All of this is intended to lay the foundation for hypothesis testing, which will be covered in next week’s class.

*In preparation for this class read:*

Rosner, Chapter 3,
Chapter 4 (sec. 4.1 - 4.3, 4.10, 4.11),
Chapter 5 (sec. 5.1 – 5.5).

Best: Chapter 2 "Confusing Numbers."

**Oct 18**  The Logic of Hypothesis Testing: One-Sample Tests of Statistical Significance

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 for a single measurement. *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. The \( z \) test and Student's \( t \) test for a single sample mean will be presented. We also will learn about the \( z \) test for a single sample proportion and confidence intervals.

*In preparation for this class read:*

Rosner, Chapter 6 (sec. 6.1, 6.2, 6.5, 6.6),
Chapter 7 (sec. 7.1 – 7.4, 7.12).

Best;  Chapter 3 "Scary Numbers."

**Oct 25**  Hypothesis Testing: Two-Sample Tests of Statistical Significance

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 one-tail vs. two-tailed tests, the Satterthwaite adjustment for heterogeneity of variance, and the use of transformed data. We will conclude the evening with a discussion on the Mann-Whitney U test and the Wilcoxon signed rank test as non-parametric alternatives to the Student t-tests for independent and related samples, respectively.

In preparation for this class read:

Rosner, Chapter 8 (sec. 8.1, 8.2, 8.4, 8.6 – 8.8), Chapter 9 (sec. 9.1, 9.3, 9.4, 9.5).

Best; Chapter 4 "Authoritative Numbers."


Nov 1 Midterm Exam

Nov 8 Hypothesis Testing: Multiple Samples and the Analysis of Variance
(6:00pm to 9:30pm)

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 an unacceptable amount of statistical error. For example, we would need to conduct six t-tests if we wanted to examine all pairwise comparisons among
four study 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 study groups. In addition, we will learn about a priori or planned comparisons, which allow investigators to test specific patterns of group means (hence, using multiple groups as a means to an end). We also will learn about post-hoc comparisons, which provide ways for comparing a reference group with all the other study groups (i.e., the Dunnett procedure) or for conducting pairwise comparisons among all group means (e.g., the Tukey and Scheffe procedures).

In the second part of the class, 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:

Rosner, Chapter 12 (sec. 12.1 – 12.4, 12.5 [pp.538 – 541], 12.6 [pp. 548 – 553]).


Nov 15  Hypothesis Testing: 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 ± 1.00 (you might have wondered about this if you had a statistics class in the past). We also will learn about the Spearman Correlation Coefficient, which is a non-parametric measure of association.

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:

Rosner, Chapter 11 (sec. 11.1 – 11.4, 11.7, 11.8 [pp.455 – 462], 11.12).


Best: Chapter 5 "Magical Numbers."
Nov 29  Hypothesis Testing:  Categorical Data and the Chi-Square Test

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. The chi-square goodness of fit test will be covered in the first part of class and students will learn how this procedure helps an investigator test for specific patterns of frequencies (including randomness). In the second part of class students will then learn how the chi-square test for association is used to explore relationships among categorical variables. This will be followed by a discussion of measures of categorical association. The phi and Cramer V coefficients will be discussed as well as relative risk (or benefit) and odds ratios.

In preparation for this class read:

Rosner, Chapter 6 (sec. 6.7 [pp.176 – 179]), Rosner, Chapter 10 (sec 10.1, 10.2 [pp. 357 – 367], 10.4 [pp. 373 – 377], 10.6 [pp. 390 – 394]).

Best; Chapter 6 "Contentious Numbers"


Dec 6  Statistical Power and Sample Size Estimation
(6:00pm to 9:30pm)

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 “How many subjects do I need?”

In preparation for this class read:

Rosner, Chapter 7 (sec. 7.5, 7.6), Chapter 8 (sec. 8.10).

Best; Chapter 7 "Toward Statistical Literacy?"

(continued on next page)


Dec 13 Final Exam