Instructor: Joel Greenhouse  
Statistics Department  
Baker Hall 229F, Phone: 268-8872  
e-mail: joel@stat.cmu.edu  

Teaching Assistant: TBA  

Class Schedule: Tuesday and Thursday 10:30-11:50, in DH A302  

Office Hours: Whenever I am in my office, or by appointment. Also, formal hours will be announced.  

Optional Textbooks:  

Prerequisites: A course in statistical methods, including regression and analysis of variance, and a course in statistical theory, including conditional probability and maximum likelihood estimation (e.g., 36-226 or 36-625).  

Course Webpage: Go to https://www.cmu.edu/canvas  

Overview  

Epidemiology is the branch of medicine concerned with understanding factors that cause, prevent, and reduce diseases by studying associations between disease outcomes and their suspected determinants in human populations. It thus requires measuring and comparing rates of diseases in populations in which different persons or groups have different exposures to possible determinants.  

Epidemiology involves taking measurements from subjects and making inferences about relevant characteristics of a wider population typifying the subjects. Since statistics is the science that is concerned with making inferences about population parameters using sampled measurements, statistical methods provide the basic tools for epidemiological research. Good epidemiological research thus requires both an understanding of statistics, which provides a means of assessing the probabilities associated with apparent associations between determinants and disease outcomes, and a knowledge of medicine, enabling statistically probable associations to be interpreted and placed in a biologically causal framework. (McNeil, 1996)  

This course will consider a number of statistical problems that arise in the design and analysis of research in the biomedical, public health and social sciences which are not ordinarily considered in theoretical statistics, eventhough some of these problems are of considerable generality. In this sense the course is applied, but this should not be interpreted to mean that we shall only be concerned with the application of statistical methods, such as the analysis of variance or regression, to bodies of data from the biomedical and social sciences. We will indeed discuss methods and analyze data but we will also stress the principles and assumptions behind the methods. In particular, the problems to be considered all use the basic ideas of probability and statistics, such as point and interval estimation. The aggregate
of problems of the type we shall consider is sometimes referred to as biostatistics, but this does not mean that there is a special type of statistics used in the biomedical sciences that is not used elsewhere. The class of statistical methods we shall consider are primarily methods for the analysis of categorical data.

Discrete or categorical data arise whenever counts are made instead of measurements. When we look at several categorical variables simultaneously, we say that they form a multidimensional contingency table, with each variable corresponding to one dimension of the table. Methods for the analysis of categorical data are based on analogs of regression and analysis of variance models. These methods are interesting both theoretically because of the elegance of the underlying theory, and from an applied point of view, because of the wide variety of uses of these methods that have appeared, and continue to appear. In this course, methods for the analysis of multidimensional contingency table are discussed with the purpose of learning how to apply them to data. The central statistical themes are building models, assessing fit and reliability, and drawing conclusions.

Objectives

1. To develop skills in model building, investigating model assumptions, and interpreting results from statistical models with particular applications to epidemiologic and social science data. Specifically, to learn how to analyze discrete multivariate data.

2. To introduce methodology for dealing with unique problems of the design and analysis of epidemiologic data and to develop skills in epidemiologic thinking.

3. To develop principles for critically evaluating biomedical and social science research and assessing the impact of their results on public policy.

COURSE POLICIES

Student Evaluations

1. There will be a Midsemester and Final exam for the course. Details will be given in class.

2. Homework will be assigned weekly. The purpose of these assignments is to help you learn the material. It is okay to discuss an assignment with other students but the written solutions to homework problems must be your own and not copied from someone else.
   - **You should always show all of your work.** You will not receive credit for simply writing down a numerical answer, even if the calculations seem simple enough to do in your head. Showing the method of solution is as important as the correct answer.
   - **Assignments are due at the beginning of class.** An assignment turned in before the end of class will be accepted but PENALIZED FOR BEING LATE; papers turned in by 2:00 pm on the day that the assignment is due will be accepted but will be PENALIZED 15 POINTS; Assignments turned in between 2-5 pm WILL NOT RECEIVE ANY POINTS, but will be corrected in order to give you feedback on your work. No assignment will be accepted after 5:00 pm on the day it is due. No excuses please.
   - Assignments are designed to take on average about 6 hours to complete. Please start homeworks early so that you can get help if you need it. Each homework assignment will be worth 100 points. These points will be divided approximately equally among each of the parts of the assignment.
   - Some assignments will require using a statistical software package, such as R or MINITAB. Unless requested, no computer output should be submitted. Rather you should cut-and-paste the appropriate
parts of the output and neatly insert it onto your homework paper. Please label all output, tables, variables, etc., appropriately.

- **Solutions** will usually be available at the following lecture or posted on the course web page. The solutions will provide you with examples of what a complete answer should look like. Read them.

3. **Lecture Checkpoints**: After each lecture and due before the next lecture, you will be asked to answer 4-5 questions about the material covered in lecture. These will be administered on Canvas and will contribute to your course grade. The purpose of these questions is to help you self-assess your understanding of the lecture material.

4. **In-class exercises**: Approximately each week, you and a partner will be given one or two problems to work on in-class. The purpose of these exercises will be to give you practice using concepts and methods that we have discussed in class and will help advance your understanding of the course material and prepare you to do the homework and exam problems. During these exercises I will be available to help you and to answer questions if you get stuck.

   Note: The Lecture Checkpoints and the In-class Exercises are important assessment tools that provide you and me with feedback on how well you understand the course material. Please take them seriously.

**Course Web Page**: Handouts and assignments will generally be available on Canvas within a day of being distributed in class. Course announcement, policies, etc. will always be available.

**Final Course Grade**

The lowest homework grade will be dropped, except for the last assignment. The last assignment is required. The remaining homework grades will be used to compute the homework average.

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>25%</td>
</tr>
<tr>
<td>Midterm Project</td>
<td>32.5%</td>
</tr>
<tr>
<td>Final Project</td>
<td>32.5%</td>
</tr>
<tr>
<td>Lecture Checkpoints</td>
<td>10%</td>
</tr>
</tbody>
</table>

You should expect the letter grade cutoffs for this course to be: 90-100 A, 80-89 B, 70-79 C, 60-69 D, below 60 F. There will be NO curving of individual exam grades. Any adjustment to the cut-offs for letter grades will be done after final exams are graded based on the distribution of scores.

**Academic Integrity**

We encourage students to work together. However, there is a difference between good collaboration and academic misconduct. We expect you to read over this list, and you will be held responsible for violating these rules. We are serious about protecting the hard-working students in this course. We want a grade for 36-461 to have value for everyone. We hold accountable both the student who cheats and the student who allows or enables another student to cheat. Make sure that you are doing everything you can to protect the value of your work on homework, quizzes, and exams.
Examples of Good Collaboration:

- After working on every problem yourself, getting together with a small group of other students who have also worked on every problem themselves to discuss results.
- Discuss ideas for how to solve the more difficult problems.
- Finish the homework problems on your own so that what you turn in truly represents your own understanding of the material.
- Writing up homework problems on your own. It is appropriate to discuss in groups, but when it comes to writing up results you should be alone.
- Discuss concepts or practice problems in the group.
- Explain concepts or practice problems to each other.
- Ask a T.A. for help on a problem related to a homework problem, but do the actual homework problem yourself. Problems done in class are great examples that will help you with homework problems.

Examples of Academic Misconduct (but not limited to):

- Divide up the problems among a group. (You do #1, I’ll do #2, and she’ll do #3: then we’ll share our work to get the assignment done more quickly.)
- Attend a group work session without having first worked all of the problems yourself.
- Start the problem yourself but then copy somebody else’s solution for the rest of the problem after you got stuck.
- Read someone else’s answers before you have completed your work.
- Have a tutor or T.A. work through all (or some) of your HW problems for you.
- Not protecting the integrity of your work, for example, letting someone copy your computer output or not keeping your exam covered.
- Doing the work for another student on any aspect of the course.

CMU students are expected to follow the ethical guidelines and cheating and plagiarism policies defined in the Student Handbook. See, for example http://www.heinz.cmu.edu/academic-resources/student-handbook-forms/index.aspx. This material is available in hardcopy and on the CMU web site. Please read it carefully! You will be held accountable for violations of these guidelines and policies that come to my attention. While we encourage you to be helpful to your classmates and to even work together, you must understand that the work you turn in must be your own. Any student who turns in work for credit that is identical, or similar beyond coincidence, to that of another student will face appropriate disciplinary action at the department, college, or university level. Cheating and/or plagiarism will not be tolerated.
**Attendance**

You are strongly encouraged to attend class on a regular basis. Since there is no assigned textbook all the material will be covered in lectures and in handouts. Only the combination of class lectures, homework sets, in-class exercises, and your own reading will give you full exposure to the material and prepare you adequately for examinations.

**Regrades**

Although we strive for consistency and accuracy in grading, we know that grading mistakes can occur. These mistakes can help you or hurt you. You should focus on the overall quality of your work, and not waste energy in minor grading arguments that ultimately have no significance or material consequences. Since you will be evaluated on your achievements throughout the semester, your overall performance will be accurately assessed.

No homework regrades will be considered if the resultant change in score is expected to be 5% or less, unless there is an obvious error in tabulation or overlooked material. We will gladly correct all errors in tabulation. Beyond that, however, it is not our responsibility to guess your intentions or to imagine how your unstated assumptions, ambiguities or omissions are to be interpreted. We only grade what is given to us; not what is to be inferred or assumed. Requests for regrades based solely on your opinion that too many points were deducted for an incorrect answer will not be considered. All assignments and exams are graded using a common grading scheme so that the same number of points are deducted for the same incorrect answer.

All regrading requests must be accompanied by a written statement carefully highlighting and explaining the items you feel were misgraded within one week of when the assignment or examination is returned. You must explain how the contested items meet both the spirit and the letter of the assignment or effectively answer an examination question. When we receive your request for regrading, we may review the entire assignment or examination. Your ultimate score may therefore increase, decrease, or remain unchanged when we review it.

**Use of Email**

Please be advised that sending email to me or the TA does not create a responsibility or obligation to respond to it. Sending us email does not shift any responsibility from you to us; you are still responsible for the on-time, high quality completion of assignments. Do not send complicated questions or requests to us via email. Replies will not be given for email questions or problems requiring lengthy (more than a couple of sentences) or complicated responses. These types of communications should be done in person.

**Office Hours**

The purpose of office hours is to provide you with an opportunity for additional conversation, guidance or help. We look forward to meeting with you and getting to know you during our office hours. Come early with questions and problems. If you wait until the last minute before an assignment is due, you may not get the quality help that you need.

If you have problems related to an assignment, please come prepared when you seek advice or help from me or the TA. The more specific your question and the more documentation of your attempted solution the better able we will be to help you. Statements such as “I don’t get it” or “I am clueless” are not specific and are not helpful.
Study Skills and Time Management

Planning and time management are essential skills for academic success. It is essential that you make every effort to keep up with the lectures, readings and assignments. Manage your time effectively and bear in mind that you are ultimately responsible for your own performance in this courses. Plan to start each assignment early so that you finish on time. This is your best strategy to have time for questions and to react to any problems.

Special Needs:

If you have been certified by the Office of Disability Resources (ODR) as someone needing a course adaptation or accommodation because of a disability, or if you need special arrangements in case the building must be evacuated, please contact me during the first week of classes. The request for special arrangements should be turned in and signed no later than one week in advance of exams but I would prefer to know as early in the semester as possible. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at access@andrew.cmu.edu

The website for the ODR is: www.cmu.edu/disability-resources

Statement of Support for Students Health & Well-being

Please come see me if you are experiencing any difficulties with this course, your academic program in general, or if you are experiencing any difficult or challenging life events. As hard as it might seem, it is much better to deal with these sorts of issues earlier rather than later. I want to assure you that I take your health and well-being seriously.
# Statistical Methods in Epidemiology

## COURSE SYLLABUS

<table>
<thead>
<tr>
<th>Week</th>
<th>Topics</th>
<th>Optional Readings from Agresti</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction Ch. 1.1 Epidemiologic Studies Notes Analysis of 2x2 Tables Ch. 2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Measures of Association Ch. 2 Scientific Method &amp; Causal Inference Notes</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Thurs 9/21: No Class</strong></td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>Statistical Theory Ch. 1.2-1.4 Notes Maximum Likelihood Estimation Notes</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Goodness-of-Fit Notes</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Approximate time for the Midterm</strong></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Combining $K$ 2x2 Tables Notes</td>
<td></td>
</tr>
<tr>
<td>7-8</td>
<td>Logistic Regression Ch. 3-4</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Fitting Models Ch. 5</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Applications Notes</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Introduction to Clinical Trials Notes</td>
<td></td>
</tr>
<tr>
<td>12-15</td>
<td>Introduction to Survival Analysis Notes</td>
<td></td>
</tr>
</tbody>
</table>

**Do NOT make travel plans before you know when our final exam is scheduled for**