

PM518A Statistical Analysis for Epidemiological Studies

Units: 3

Spring 2021 Wednesdays:

Synchronous Session Lecture 1:00-3:00 pm

Synchronous Session Lab 3:00-4:00 pm

Location: Zoom links for each synchronous session and Zoom office hours are available on Blackboard

Instructor: Wendy Mack, PhD

Zoom Office Hours: Fridays 11am-noon

Contact Info: wmack@usc.edu

Teaching Assistant: Irene Jiayao Chen

Zoom Office Hours: Mondays 3-4pm

Contact Info: jiayaoch@usc.edu

Course Description

The course covers methods of statistical analysis for data arising from case-control and cohort studies, including analysis of dichotomous (case-control) and time to event (survival) data. Major topics include probability models for case-control and cohort data, contingency tables and logistic regression modeling for unmatched and matched case-control data, and comparison of disease rates and disease risk with lifetables, Poisson regression, proportional hazards and other survival regression models. Applications will use the Stata software.

Learning Objectives

At the end of the course, you will be able to:

1. Describe models of exposure-disease association. Name the various parameters used in epidemiology to estimate this association.
 - a. Describe the concept of disease risk and use epidemiologic data to estimate disease risk
 - b. Name and use various models which estimate the relationship between disease risk and exposures.
2. Describe the two major epidemiologic study designs: case-control and cohort studies.
 - a. Name the advantages and limitations of each design, with reference to estimating disease-exposure associations
 - b. List and apply in epidemiologic data the various biases which may distort the estimate of the disease-exposure association
3. Select and use appropriate statistical methods for the analysis of case-control data, including:
 - a. Analysis and stratification of 2 x 2 tables and 2 x K tables for unmatched case-control data
 - b. Analysis of matched case-control data

- c. Logistic regression models: Unconditional (for unmatched case-control data) and conditional (for matched case-control data)
4. Select and use appropriate statistical methods for the analysis of cohort data, including:
 - a. Life table analysis
 - b. Standardization of rates and SMR analysis
 - c. Poisson regression to model disease rates and exposure associations
 - d. Partial likelihood (proportional hazards) regression (also known as Cox regression)
 - e. Handling of competing events risks in analysis of cohort data
 - f. Parametric survival time models
5. Recognize and appropriately utilize the concepts of confounding and effect modification in the analysis of case-control and cohort data and interpretations of analysis results.
6. Given a set of epidemiologic data, appropriately analyze and write a statistical methods, results and conclusions section.
7. Use Stata software for the analysis of epidemiologic data.

MPH Competencies:

Students who complete this course will meet the following MPH competencies:

1. Advise fellow investigators on the design, conduct and data analysis for studies in the health sciences
2. Serve as data coordinators, managers and analysts for epidemiologic studies or public health-related projects, e.g., multi-center clinical trials, government surveys, outcome databases, etc.
3. Analyze basic relationships between risk factors and outcome data using data management and statistical software.
4. Determine the appropriate study design to analyze a community health problem.
5. Seek additional biomedical, statistical, and computing assistance when appropriate.
6. Write the results of data analysis in a coherent report that can be used to guide public policy.
7. Identify potential bias in research reports and evaluate the likelihood that these potential biases actually explain the findings.
8. Identify behavioral/social/cultural and epidemiological factors in a particular setting/problem and analyze how these factors affect disease.

Prerequisite: PM512

Recommended Preparation: PM511B (Analysis of Categorical Data) is not required but is recommended. Students should also have a solid foundation in linear regression methods and contingency table analysis

Course Notes and Related Materials

Lecture and lab materials will be posted on Blackboard. The course will be organized by week. Each week folder on Blackboard will include:

1. Lecture notes and a short review of major points from the lecture notes
2. Recordings of introductory materials and concepts over the lecture (to be viewed prior to class)
3. Datasets used in the lecture notes
4. Exercises and data that will be completed during the live session
5. Lab materials
6. Additional resources (optional) to supplement lectures and readings

Synchronous Live Sessions

1. Lecture: Each 2-hour synchronous session (1-3pm) will include a live lecture, a short summary of major points from lecture notes, in-class exercises, quizzes and polls.
 - a. Quizzes will cover the most recently submitted homework assignment. There will be one quiz for each homework assignment. Each quiz will count 5 points towards that homework grade (see homework assignment and grading below).
 - b. Polls will be short questions on the material covered in class. Polls will not be graded, but will count towards your class participation score.
2. Lab: Each 1-hour synchronous session will cover lecture concepts and use of Stata, orientation to homeworks when assigned, and review of homeworks after submission.

Technological Proficiency and Hardware/Software Required

The course requires use of Blackboard to deliver course content and assignments. Zoom will be used for synchronous lecture and lab sessions and for office hours.

Stata software will be accessible on the virtual laboratory. As an alternative, you can purchase an individual version of Stata. Student versions are reasonably priced; there is a student version with 6-month availability that costs less than most textbooks. Refer to: <http://www.stata.com/order/new/edu/gradplans/student-pricing/>

You will learn the basics of Stata software use, as it applies to analysis of case-control and cohort data. The Stata corporation has a youtube channel with many many useful videos ([youtube.com/statacorp](https://www.youtube.com/statacorp)). I will provide links to relevant youtube videos in the “Additional resources” section of each week’s folder on Blackboard. I encourage you though to explore youtube further as needed.

Required Materials

Selvin S. Statistical Analysis of Epidemiologic Data, 3rd Edition. Oxford University Press, 2004.

Hosmer DW, Lemeshow S, May S. Applied Survival Analysis, 2nd Edition. John Wiley and Sons, 2008.

Optional Materials

Rothman KJ, Greenland S, Lash TL. Modern Epidemiology, 3rd Edition. Lippincott Williams & Wilkins, 2008. (Note: There is a brand new 4th Edition, coming out in early 2021!).

Description of Assessments/Assignments

1. Homework (15%): Seven homework assignments will be assigned via Blackboard throughout the course. Homeworks turned in past the due date will be evaluated and returned with written feedback to the student, **but no credit will be given**. Homeworks will involve applied exercises on the current lecture topics. Each homework will have a data analysis component; students will be instructed in labs on the use of STATA procedures and the concepts required to complete homeworks. **Although STATA is the software package taught in the lab and used for examples in lecture, you may use any appropriate statistical software to complete homeworks.**

Students will form groups of 4 at the beginning of the semester. You will work in these groups to complete and turn in homework assignments. **For each group, only one homework should be turned in via Blackboard. The score given on that homework will be applied to all students in your group. Homeworks that are turned in for separate groups but are found to be close copies of one another will be given no credit.**

For each homework, the total score will be comprised of **two parts**:

1. The **group** response to the homework assignment
 2. A very short (5-10 minute) quiz administered **individually** in the lecture synchronous session following the homework deadline. Each quiz will be worth 5 points.
2. Class participation (5%): This includes: (1) **attendance** at synchronous lectures and labs; (2) **participation** in data analysis exercises (completing the analysis and participating in verbal responses and discussions); (3) **participating** in general question-answer exercises and polls that are part of the synchronous lectures.
 3. The midterms (30% for midterm 1, 20% for midterm 2) will be conducted during synchronous sessions. Both will be open book/open note. Midterm 1 will cover the approximate first half of the course; you will be given 3 hours to complete this midterm. Midterm 2, along with the project (below) will cover the approximate second half of the course; you will be given 2 hours to complete midterm 2.
 4. Project (30%): Students will be provided a dataset with questions to be addressed. You will be given approximately 4 weeks to complete the project. The project will involve data analysis and a written document summarizing analyses and responses to questions.
 - a. **Students are expected to work on projects INDIVIDUALLY. Any evidence of sharing and collaboration will be at the minimum assigned a score of 0 and at the maximum result in failure in the course.**

Grading Breakdown

Grades will be recorded on the Blackboard gradebook

Homework: 15%
Participation: 5%
Midterm 1: 30%
Project: 30%
Midterm 2: 20%

Grading Scale

Course final grades will be determined using the following scale.

A	95-100
A-	90-94
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	73-76
C-	70-72
D+	67-69
D	63-66
D-	60-62
F	59 and below

Assignment Submission Policy

All homeworks as well as the project will be posted and are to be turned in through the course Blackboard. The TA will show you how to do this.

Assignments that are submitted past the deadline will be reviewed, but given no credit. If you have valid reasons why you cannot submit assignments by the deadline, please discuss this with me prior to the assignment deadline.


Additional Policies

We will be analyzing data in class, with students reporting back their results in class.

A portion of the lecture notes covering more basic concepts and methods that were covered in PM510, PM512 and other of your statistics and epidemiology courses, will **not** be covered in class lectures. These notes will be **assigned reading** and recordings covering these notes will be available on Blackboard. Unless otherwise specified, students will be responsible for all concepts covered in these notes. You will be tested on the material in these notes in homeworks and on midterms.

Approximate Course Schedule: A Weekly Breakdown

	Topics/Daily Activities
Week 1 Jan 20	<p>PLEASE VIEW LECTURE RECORDING PRIOR TO SYNCHRONOUS LECTURE: Introduction to course, requirements, grading</p> <p>SYNCHRONOUS LECTURE TOPICS: Epidemiologic research methods – objectives and study designs Measures of disease occurrence Measures of exposure-disease association Introduction to case-control studies Dynamic vs. stable cohorts and case-control sampling</p> <p>READING: Selvin Chapter 1(pp 1-22), Appendix C</p>
Week 2 Jan 27	<p>To be read and reviewed by student prior to class (Lecture notes and recording on Blackboard): 1. Overview of probability, likelihood theory, maximum likelihood estimation, significance testing 2. 2 x 2 tables: Odds ratio estimation, confidence intervals (exact and approximate) on the OR 3. 2 x K tables: Multiple exposure levels 4. Lecture notes on Blackboard and READING: Selvin Appendix E, Chapter 6 (pp 159-164), Chapter 7 (pp 190-199, 208-213)</p> <p>SYNCHRONOUS LECTURE TOPICS: Stata examples and in-class exercises with 2x2 and 2xK tables Bias, confounding, and effect modification Stratification of 2 x 2 tables</p> <p>READING: Selvin Chapter 2, Chapter 6 (pp 179-189), Chapter 7 (pp 199-208, pp 227-235)</p>
Week 3 Feb 3	<p>SYNCHRONOUS LECTURE TOPICS: Introduction to matched case-control studies Odds ratio for matched studies: estimation and confidence intervals Statistical inference on the matched OR: exact and approximate (chi-square) tests 1:M matching, 1:variable matching Matched case-control: illustrations and in-class exercise</p> <p>READING: Selvin Chapter 10 (pp 291-296, 300-313, 313-320)</p>
Week 4 Feb 10	<p>SYNCHRONOUS LECTURE TOPICS: Multivariable analysis: Logistic Regression (Introduction to logistic model) Unconditional logistic regression for unmatched case-control data: parameter estimation, inference, goodness of fit Unconditional logistic regression: illustrations and in-class exercise</p> <p>READING: Selvin Chapter 8</p>

	Topics/Daily Activities
Week 5 Feb 17	<p>SYNCHRONOUS LECTURE TOPICS: Conditional logistic regression for matched case-control data: model, parameter estimation, inference Conditional logistic regression: illustrations and in-class exercise Small sample methods: Firth logistic regression for case-control data</p> <p>Optional materials (lecture notes and recording; you will not be tested on this): Case-case designs</p> <p>READING: Selvin Chapter 10 (pp 320-334)</p>
Week 6 Feb 24	<p>Midterm 1</p>
Week 7 Mar 3	<p>To be read and reviewed by student prior to class (Lecture notes and recording on Blackboard): Introduction to cohort studies Risks, rates, standardization of rates Person-years analysis: SMR estimation, exact and approximate inference on the SMR Comparison among exposure groups: SRR estimation, exact and approximate inference</p> <p>SYNCHRONOUS LECTURE TOPICS: Lifetables: data structures, Kaplan-Meier estimation, survival curves, log rank test, graphing survival curves Illustrations and in-class exercise</p> <p>READING: Selvin Chapter 1 (pp 32-39); Selvin Chapter 11</p>
Week 8 Mar 10	<p>SYNCHRONOUS LECTURE TOPICS: Multivariable analysis: Poisson regression for disease rates Poisson regression: use of internal vs. external disease rates for unexposed comparators Poisson regression: illustrations and in-class exercise Negative binomial regression</p> <p>READING: Selvin Chapter 9</p>
Week 9 Mar 17 	<p>SYNCHRONOUS LECTURE TOPICS: Multivariable regression for individual survival time data: survival likelihoods (full and partial) Proportional hazards regression: model, parameter estimation, inference</p> <p>READING (over next few weeks): HLM Chapters 1, 2, 3, 4; Selvin Chapter 13</p>

	Topics/Daily Activities
Week 10 Mar 24	<p>SYNCHRONOUS LECTURE TOPICS: More on proportional hazards regression, including: Evaluation of proportional hazards assumption Modeling of time-varying covariates In-class exercise</p> <p>READING: HLM Chapters 6, 7</p>
Week 11 Mar 31	<p>SYNCHRONOUS LECTURE TOPICS: More on proportional hazards regression, including: Proportional hazards assumption Stratification of baseline hazard In-class exercise</p>
Week 12 Apr 7	University Wellness Day
Week 13 Apr 14	<p>SYNCHRONOUS LECTURE TOPICS: More on proportional hazards regression, including: Estimation and graphing of hazard and survival functions Tied event times Model diagnostics Prediction In-class exercise</p>
Week 14 Apr 21	<p>SYNCHRONOUS LECTURE TOPICS: Cohort sampling -- nested case-control studies, case-cohort studies Competing risks</p>
Week 15 Apr 28	<p>SYNCHRONOUS LECTURE TOPICS: Competing risks: in-class exercise Parametric survival models</p>
Finals Week May 5	Midterm 2

Statement on Academic Conduct and Support Systems

Academic Conduct

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Section 11, *Behavior Violating University Standards* <https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions/>. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, <http://policy.usc.edu/scientific-misconduct/>.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the *Office of Equity and Diversity* <http://equity.usc.edu/> or to the *Department of Public Safety* <http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us>. This is important for the safety whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. *The Center for Women and Men* <http://www.usc.edu/student-affairs/cwm/> provides 24/7 confidential support, and the sexual assault resource center webpage sarc@usc.edu describes reporting options and other resources.

Support Systems

A number of USC’s schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the *American Language Institute* <http://dornsife.usc.edu/ali>, which sponsors courses and workshops specifically for international graduate students. *The Office of Disability Services and Programs* http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, *USC Emergency Information* <http://emergency.usc.edu/> will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.