Applied Biostatistics for Health Researchers (LHS665)

Fall 2023 Syllabus

Credit Hours   4

Prerequisites   Graduate standing or permission of instructor

Class Schedule   Wednesday 12:00 to 4:00pm in MedSci 2 building, Room 2813
(from Wednesday August 30th to December 6th)

Faculty   Matthew A. Davis, MPH, PhD
Associate Professor
426 North Ingalls, Room 2122
Phone: (734) 764-2814
Email: mattadav@umich.edu
Office hours: Wednesdays 3:00 to 4:00pm & by appointment

Course Website: http://canvas.umich.edu

Course Description

This is a PhD-level biostatistics course that covers fundamental statistical concepts and methods for researchers who need to analyze health and/or healthcare data and interpret research. Major topics include descriptive statistics, probability theory, statistical inference, hypothesis testing, correlation, regression (linear and logistic), survival analysis, reliability/validity of diagnostic tests, and epidemiological study designs. Relevance of analytic techniques to healthcare will be demonstrated via a series of 10 labs that focus on applications. Students will become proficient in basic data management and analysis using a statistical software program including data importation/exportation, management of datasets (creating new variables, merging and appending datasets), and statistical analyses. Effective presentation of quantitative results in tables and graphics will be emphasized throughout the course.

Course Objectives

Upon completion of the course students will be able to:
  (1) Identify the major types of epidemiological study designs and understand the strengths and weaknesses of each;
  (2) Recognize the types of numerical data and apply appropriate descriptive statistics to produce summary measures and graphics;
  (3) Utilize a statistical software program for basic data management and analysis;
  (4) Select and apply the appropriate statistical technique to test a hypothesis;
  (5) Interpret results of statistical techniques covered in class including the assessment of the role of chance;
  (6) Be able to evaluate the performance of a diagnostic test; and
  (7) Identify key applications of statistical techniques to clinical research.

Content Topics

- Health data & variable types
- Measures of central tendency & dispersion
- Normal distribution & Z-scores
- Epidemiological study designs
- Quantiles
- Health measures & risk adjustment
- Probability & binomial equation
- Sampling distribution & central limit theorem
- Standard error
- Hypothesis testing & p-values
- Chi-squared & McNemar's tests
- 95% confidence intervals
- t-tests (for paired & unpaired data)
- ANOVA
- Correlation (Pearson & Spearman)
- Linear regression
- Logistic regression
- Non-parametric tests
- Diagnostic test performance (validity, reliability, & ROCs)
- Indicator variables
- Publication-quality tables/figures
- Appending versus merging data
- Survival analysis

Teaching Methods and Target Audience

The primary teaching method in this course is synchronous, didactic lecture. Additional methods include labs assignments, course journal club, assigned readings, and occasional asynchronous lectures (via pre-recorded videos). Assessment of student performance will occur through weekly labs, two exams, and a course project presented in a poster session. All materials for this course are available on the Canvas LHS665 website. For questions regarding the Canvas website please contact me. *Please note that the syllabus is a progressive document which will be updated frequently throughout the semester. Please plan to check in at least weekly to review any changes in upcoming work.*

Grading and Evaluation

Grades will be based on a composite score reflecting class engagement, labs, exams, and your final project grades. The following describes how your grade will be determined:

Class engagement	10%
Labs 1-8 (3.75 points each)	30%
Exam 1	20%
Exam 2	20%
Final project ("Lab 9")	20%
Total	100%

Grading Policies

You will receive a grade at the end of the term in accordance with the Department of Learning Health Sciences guidelines. The course grade will be a letter grade (A – D, F) will be assigned using this conversion scale:

A+	97-100	B+	87-89	C+	77-79	D+	67-69
A	93-96	B	83-86	C	73-76	D	63-66
A-	90-92	B-	80-82	C-	70-72	D-	60-62
						F	< 60

Your final grade for the course will be posted within 48 hours following the end of the poster presentation on 12/06. You have until December 10, 2023 to contest any residual issues

regarding your course grade – shortly following, your grade for LH665 will be submitted to the Office of the Registrar and cannot be changed. Students who experience extraordinary circumstances and wish to request a 'incomplete' for LHS665 must request this before 12/14 and will be asked to sign documentation that details due dates for deliverables.

Class engagement (i.e., participation)

Notice that 10% of your final grade will be based on what we call "class engagement" in the course. Engagement grades will be awarded at the end of the term in 5-point increments (i.e., 80, 85, etc). Class engagement is based on two parts: (1) lecture attendance and (2) participation in journal club session.

It is expected that you will attend all lectures. Students who are feeling ill should not come to class in-person. Grades will not be negatively impacted by not attending class due to illness. Other than illness, students are expected to attend all classes and will only be excused for extenuating circumstances and cultural and religious holidays. These circumstances shall be discussed with the instructor prior to missing class. Attendance will be taken during class sessions. Failure to adhere to the policy may result in reduction in the student's grade for the course.

Lab due dates/times

All labs are expected on the designated due dates and must be submitted by 11:59pm Eastern Standard Time. Extensions must be negotiated with the course instructor *prior to the scheduled date of submission*. All labs are to be completed by the end of the academic term.

Unless stated otherwise, labs should be uploaded as Word document files (i.e., either file extension ".doc" or ".docx"). Late submission of labs that were not negotiated prior to the due date with the instructor *will result in an automatic grade reduction of 25% for each day following the assigned due date*.

Required Texts and Other Materials
There is no required textbook for the course. Selected readings will be assigned and are available via digital textbooks at the Taubman Health Sciences Library. For your convenience we will provide PDFs throughout the term.

Recommended (optional) textbooks
- Pagano M and Gauvreau K. Principles of Biostatistics, 2nd Edition. Brooks Cole; Belmont, CA 2000 (ISBN-10: 0-534-22902-6)

    For those of you who would prefer a more traditional statistics textbook for reference or to supplement the course textbook, Principles of Biostatistics is a very good choice.

- Gonick L and Smith W. The Cartoon Guide to Statistics. Harper Perennial; New York, NY 1993 (ISBN-10: 0-06-273102-5)

    Yes, you are not mistaken – I am recommending a cartoon book. This book was co-written by a cartoonist and a card-carrying statistician. The book is a fun (and accurate) take on statistics.

Other required materials (computer software & calculator)

You will need Microsoft Office (Excel, Word, and Powerpoint) for labs. We will use Stata BE (version "BE" for basic edition) statistical software for this course as well. There is student pricing for Stata available on the Stata website (http://www.stata.com/) and a variety of purchasing options based on the length of the license. For those of you who are not interested in having

your own copy of Stata on your computer, you are able to access the software using a virtual desktop interface (VDI).

If you decide to forgo purchasing a copy of Stata it will be at your discretion – *the instructor is not responsible for assisting students will issues pertaining to the virtual desktop interface*. You should have a calculator for this course as well. We recommend a graphing calculator (such as a Texas Instruments TI-83 or TI-84) that has the ability to perform more complex calculations.

Overview of the Course: What's the big idea?
It's been said that <u>*statistics is where mathematics meets the real world*</u>. Throughout this course we will explore the various applications of statistics to health and healthcare. Although you likely have already come in contact with some of these topics (for instance, correlation and regression) this course will make their relevance and usefulness to health research and medicine more salient. Clinical practice is dependent on data and the application of statistical analysis underpins what we know about disease etiology, prognosis, and clinical effectiveness. Striking a balance between statistical theory and application is an instructional challenge. In this course, the focus will be on application; however, theory will be used to reinforce concepts and demonstrate how/why. The course is separated into the following four modules:

Module 1: Data and Descriptive Statistics

This first module will lay the foundation for more advanced topics later in the course. We will discuss data and different types of variables. How we summarize data descriptively depends on the type of data (for instance systolic blood pressure in mmHg would be summarized differently than obesity status). You'll have the opportunity to practice analyses using Stata in the labs that will include summarizing data using measures of central tendency and creating graphics. We will also provide an overview of epidemiological study designs (that will be helpful for journal club).

Module 2: Introduction to Statistical Inference

Things are about to get really interesting. When we go about studying some phenomenon it is usually impossible to obtain data on the entire population of interest. Instead, we draw a sample from the population (ideally at random so it is more likely to represent the population we're interested in). Statistical inference is all about how to draw meaningful conclusions based on a sample. Of course, as we never know what the true value is in the population, drawing conclusions based on data from a sample relies on theory. We'll discuss the concept of a sampling distribution and what's been called the LeBron James of statistical theory (drum roll, please): *the Central Limit Theorem*.

Module 3: Analysis of Variance, Correlation, and Regression

Much of this module will be devoted to walking through a variety of techniques for quantitatively examining relationships between an independent and dependent variable. Fundamentally, these "tests" that we use to evaluate the role of chance in interpreting relationships are designed to separate signal (the effect) from noise (random variation). We'll end this module by introducing linear and logistic regression and I'll show you some examples of how regression can be used to adjust for confounding factors. In this module too we'll squeak in sessions on evaluating diagnostic test performance and survival analysis.

Module 4: Putting it All Together

By this point in the term it may feel like statistics is just a smattering of jargon, ideas, and math tricks. Now that you're familiar with the lingo (welcome to the club, by the way) and have walked through major statistical concepts, we'll place the techniques we've learned in a broader context to see how it all fits together. To do so, we'll hang the tests we've covered on independent/dependent variable formats and you'll have the chance to conduct your own analyses by either selecting one of the

datasets provided by us or using your own data. The course will conclude with a poster presentation where you'll have the opportunity to show off your great work.

Class Schedule by Week

Date	Time	Topic	Assignment
Module 1: Data and Descriptive Statistics			
Wednesday, 08/30	*Schedule:*		Reading: Wheelan chapter 1 & 2 and Pagano chapter 2
	12:00 to 12:50pm	1.1 Introduction to data and variable types	
	1:00 to 1:50pm	1.2 Measures of central tendency and dispersion	*Course intake survey (Due 09/05 by 11:59pm)*
	2:00 to 2:50pm	1.3 Normal distribution and *Z*-scores for an observation	
	3:00 to 4:00pm	Open lab/office-hour	
Wednesday, 09/06	*Schedule:*		Reading: Motulsky chapter 10 & Weiss chapter 5
	12:00 to 12:50pm	1.4 Variable types revisited and Stata demonstration	
	1:00 to 1:50pm	1.5 Primer on epidemiological study designs (observational versus experimental designs)	*Lab 1 (Due 09/05 by 11:59pm)*
	2:00 to 2:50pm	Journal club (No. 1)	
	3:00 to 4:00pm	Open lab/office-hour	
Wednesday, 09/13	*Schedule:*		*Lab 2 (Due 09/12 by 11:59pm)*
	12:00 to 12:50pm	1.6 Quantiles	
	1:00 to 1:50pm	1.7 Health measurement and risk adjustment	
	2:00 to 2:50pm	Journal club (No. 2)	
	3:00 to 4:00pm	Open lab/office-hour	
Wednesday, 09/20	*Schedule:*		*Lab 3 (Due 09/19 by 11:59pm)*
	12:00 to 12:50pm	1.8 Applications of Z-scores	
	1:00 to 1:50pm	1.9 Probability and the binomial distribution	
	2:00 to 2:50pm	Journal club (No. 3)	
	3:00 to 4:00pm	Open lab/office-hour	

Date	Time	Topic	Assignment
Module 2: Introduction to Statistical Inference			
Wednesday, 09/27	*Schedule:*		Reading: Norman chapter 5, Altman (SD versus SE), and primer on p-values
	12:00 to 12:50pm	2.1 Sampling distribution, central limit theorem, and standard error	
	1:00 to 1:50pm	2.2 Introduction to hypothesis testing, p-values, and 95% confidence intervals	*Lab 4 (Due 09/26 by 11:59pm)*
	2:00 to 2:50pm	2.3 Tests for nominal data: Chi-squared and McNemar's tests	
	3:00 to 4:00pm	Open lab/office-hour	
Wednesday, 10/04	*Schedule:*		Reading: primer on 95% CIs
	12:00 to 12:50pm	2.4 95% confidence intervals revisited and the *t*-distribution	*Lab 5 (Due 10/03 by 11:59pm)*
	1:00 to 1:50pm	Exam review	
	2:00 to 2:50pm	Journal club (No. 4)	
	3:00 to 4:00pm	Open lab/office-hour	

Wednesday, 10/11 - Exam 1 (12:00 to 4:00pm)

Date	Time	Topic	Assignment
Module 3: Analysis of Variance, Correlation, and Regression			
Wednesday, 10/18	*Schedule:*		Reading: Pagano chapters 11 & 12
	12:00 to 12:50pm	3.1 *t*-test	
	1:00 to 1:50pm	3.2 ANOVA and correlation	
	2:00 to 2:50pm	3.3 Simple linear and logistic regression	
	3:00 to 4:00pm	Open lab/office-hour	
Wednesday, 10/25	*Schedule:*		Reading: Motulsky chapter 33
	12:00 to 12:50pm	3.4 Multiple linear regression and Spearman correlation	*Lab 6 (Due 10/24 by 11:59pm)*
	1:00 to 1:50pm	3.5 Tests for ranked data	
	2:00 to 2:50pm	Journal club (No. 5)	
	3:00 to 4:00pm	Open lab/office-hour	

Class Schedule by Week (continued from previous page)

Date	Time	Topic	Assignment
Wednesday, 11/01	*Schedule:*		*Lab 7 (Due 10/31 by 11:59pm)*
	12:00 to 12:50pm	3.6 Diagnostic test performance (validity, reliability, ROCs)	
	1:00 to 1:50pm	3.7 Paired versus unpaired data, indicator variables	
	2:00 to 2:50pm	Journal club (No. 6)	
	3:00 to 4:00pm	Designing publication quality tables and figures (poster templates)	
Wednesday, 11/08	*Schedule:*		*Lab 8 (Due 11/07 by 11:59pm)*
	12:00 to 12:50pm	3.8 Survival analysis	
	1:00 to 1:50pm	3.9 Introduction to managing data	
	2:00 to 2:50pm	3.10 Exam review (variable types, parametric, and nonparametric tests)	
	3:00 to 4:00pm	Open lab/office-hour	

Wednesday, 11/15 – Exam 2 (12:00 to 4:00pm)

| Wednesday, 11/22 | *Thanksgiving Break (no class)* | | |

Module 4: Putting it All Together

Date	Time	Topic	Assignment
Wednesday, 11/29	*Schedule:*		*Final Project (lab 9) (Due Friday, 12/05 by 11:59pm)*
	12:00 to 12:50pm	Applications of statistics	
	1:00 to 4:00pm	Open lab/office-hour (final preparation of posters)	

Wednesday, 12/06 – Class Poster Presentation (Room TBA)

How to be Successful in this Course: Some friendly advice

Schedule time each week to work on labs and *start labs early*

I have designed this course so that you should be working on a lab and/or reading every week throughout the term. You are allowed to discuss the labs with your classmates; however, the work you hand in needs to be your own (i.e., copying verbatim a classmate's work will be considered a violation of the student honor code).

As this an applied statistics course you will learn how to conduct analyses – translation: we will be programming in Stata. Programming can be challenging to learn initially and often requires trial and error. Be aware that labs (particularly those that involve Stata work) will take longer than you anticipate. *You should start labs early* (see how I underlined and italicized this again). The teaching staff is not responsible for offering assistance on computer issues/programming within 24 hours of a lab's due date/time.

If you don't get it, ask me to explain it again

Throughout the term you are strongly encouraged to bring "muddy points" (i.e., points of confusion) to my attention so we can tackle them in class. Typically, I will start each session going over muddy points.

People learn in different ways and, frankly speaking, certain concepts take more to process than others. Please be aware that there is built-in redundancy in the course – therefore, don't be alarmed if you don't understand something completely first. It is very likely that we will revisit the same concept. However, if you don't understand something, feel free to stop me and ask me to explain it again or in a different way. In the classroom, there are no stupid questions. If you don't understand something I've said it is very likely someone else doesn't either – be a hero and ask for clarification. However, I reserve the right to "shelf" a question to more fully address it at a later time.

Make sure you understand the solutions to the labs

You learn statistics by doing it. The nine labs will guide you through a collection of exercises to give you the opportunity to practice what we discuss in class. These labs are important study tools and are designed to help you keep up with the material. Make sure in the end you understand the solutions to the exercises and the main takeaway points.

Policies and Disclaimers

In this syllabus I have provided as much detail as possible to allow you to see an overall roadmap of the term – the goal being to give you with a clear idea of where we are headed to see how what we will learn fits together. However, I reserve the right to make changes as needed and I will likely add or modify course reading materials throughout the term. Ok, now for the administrative stuff. . .

Communication with me

Please feel free to reach out to me if you have any questions throughout the term. To assist me in triaging emails include "LHS665" in the email subject. Do not contact me using Canvas. (Canvas has the ability to email within the course, but we do not use this feature in LHS665).

Classroom etiquette

Students should turn off ring-tones and other audible alerts on their phones/tablets/laptops before class sessions begin, whether in-person or online. The use of phones/tablets/laptops for non-course related activities, such as texting, emailing, and social media is distracting to your classmates and instructors. If there is an emergency, please excuse yourself from the classroom to attend to it.

Grade grievance policy

Regular feedback on graded assignments and timely resolution of any grading concerns is helpful for student learning and overall positive experience. Students are encouraged to discuss any assignment grade concerns with the instructor(s) within two weeks of the grades being posted on Canvas. Similarly, any concerns and grievances related to the overall course grades should be raised by contacting the faculty of record within two weeks of the course grades being posted on Wolverine Access.

For this course, the grade dispute arbitration will happen within the Department of Learning Health Sciences (DLHS). After initial discussion with the faculty of record for the course, there is an opportunity for a second look by another DLHS faculty member with appropriate expertise for resolving grade disputes. If the student and faculty are not able to come to an agreement about the grade dispute, the Associate Chair for Educational Programs or assigned designee will adjudicate the dispute. Per Rackham's Academic Dispute Resolution Process, if the concern is

not resolved, the student may seek a formal resolution conference with the Medical School's Rackham Resolution Officer (see https://rackham.umich.edu/academic-policies/section9/ for full policy and the contact information for the current resolution officer).

Academic integrity of students

The academic community, like all communities, functions best when its members treat one another with honesty, fairness, respect, and trust. Your instructors expect students to work and study together to foster learning and understanding of the material. However, direct copying of homework, copying of homework from existing solutions, cheating on an exam, and other conduct that violates the academic integrity and ethical standards of the Rackham Graduate School community cannot be tolerated and will result in serious consequences and disciplinary action.

All written submissions must be your own, original work. Original work for narrative questions is not mere paraphrasing of someone else's completed answer: you must not share written answers with each other at all. At most, you should be working from notes you took while participating in a study session. Largely duplicate copies of the same assignment will receive an equal division of the total point score from the one piece of work.

You may incorporate selected excerpts, statements, or phrases from publications by other authors, but they must be clearly marked as quotations and must be attributed. If you build on the ideas of prior authors, you must cite their work. You may obtain copy editing assistance, and you may discuss your ideas with others, but all substantive writing and ideas must be your own or be explicitly attributed to another.

If we suspect you have cheated (including plagiarism), at the very least you will receive a zero on the assignment. Rackham policy dictates that we must report every instance of academic dishonesty, no matter how small. Suspected academic misconduct will be handled by Drs. Piatt and Donnelly. https://rackham.umich.edu/academic-policies/section8/

COVID policy

We each have a responsibility for protecting the collective health of our community. Applicable safety measures are described on the U-M Health Response website - https://healthresponse.umich.edu/ Please bookmark this site for policies, guidance, and FAQs.

Student academic dispute procedures

Rackham's Academic Dispute Resolution Policy and Procedures are available to Rackham students who have a dispute or disagreement with faculty or staff about the equity and fairness of decisions or procedures that affect their academic standing, the conduct of their research, and progress toward the degree. Such issues may arise regarding fair and equal treatment in the conduct of a class, in the pursuit of the student's research, and in the grading or evaluation of academic work and research. Other issues may concern the equity and fairness of program, department, or Rackham policies. https://rackham.umich.edu/academic-policies/section9/

Diversity, equity, and inclusion

The Department of Learning Health Sciences (DLHS) is committed to developing the institutional mechanisms and norms necessary to promote the values of diversity, equity, and inclusion, both inside and outside our classrooms. To this end, DLHS upholds the expectations that all courses will: (1) be inclusive, (2) promote honest & respectful discussions, (3) follow multicultural ground rules and (4) abide by UM policies and procedures. Inclusive courses are those in which teachers and learners co-create and co-sustain environments that support and encourage all members to participate equitably. https://rackham.umich.edu/rackham-life/diversity-equity-and-inclusion/

I strive to create a safe, equitable, and inclusive learning environment in all my courses. Each class I teach brings new perspectives and experiences that I use to inform future LHS665 administrative policies and course content regarding diversity, equity, and inclusion. I'd like to make you aware of several policies in LHS665: all names of students will be masked in grading (to avoid any potential biases), video cameras will not be required for any online sessions, and the course in-take survey will gather (for those who wish) information on preferred names/pronouns as well as provide the opportunity to make me aware of any personal circumstances. If you have any other suggestions of ways of making LHS665 more equitable and/or inclusive please bring them to my attention.

Accommodations for students with disabilities

Students should speak with their instructors before or during the first week of classes regarding any special needs. Students seeking academic accommodations should register with Services for Students with Disabilities (SSD). SSD arranges reasonable and appropriate academic accommodations for students with disabilities. Please visit https://ssd.umich.edu/accommodations for more information on student accommodations.

Student mental health and well-being

The University of Michigan is committed to advancing the mental health and wellbeing of its students. As such, throughout the term we'll often start class with a 'Wellness Spot' segment that highlights something fun in the area to do. The focus of many of these will be on outdoor and/or physical activities.

If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, contact Counseling and Psychological Services (CAPS) at (734) 764-8312 and https://caps.umich.edu/ during and after hours, on weekends and holidays, or through its counselors physically located in schools on both North and Central Campus. You may also consult University Health Service (UHS) at (734) 764-8320 and https://www.uhs.umich.edu/mentalhealthsvcs, or for alcohol or drug concerns, see www.uhs.umich.edu/aodresources.

Sexual misconduct/sexual harassment reporting

As faculty, one of our responsibilities is to help create a safe learning environment on our campus. Title IX prohibits sex discrimination to include sexual misconduct: harassment, domestic and dating violence, sexual assault, and stalking. If you or someone you know has been harassed or assaulted, you can receive confidential support and academic advocacy at the Sexual Assault Prevention and Awareness Center (SAPAC). SAPAC can be contacted on their 24-hour crisis line, 734-936-3333 and online at sapac.umich.edu. Alleged violations can be reported non-confidentially to the Office for Institutional Equity (OIE) at institutional.equity@umich.edu. Reports to law enforcement can be made to University of Michigan Police Department at 734-763-3434.

Support for food insecurity

Students across the country experience food insecurity at alarming rates. If you or someone you know is struggling with food security, please check out the Maize and Blue Cupboard. It provides food, kitchen, and cooking supplies, personal and household items, and support services. For information about accessing their services, please visit https://mbc.studentlife.umich.edu/

Confidentiality and Mandatory Reporting

As instructors, one of our responsibilities is to help create a safe learning environment on our campus. Instructors are required to share information regarding sexual misconduct or

information about a crime that may have occurred on the university's campus. Students may speak to someone confidentially by contacting SAPAC's Crisis Line at (734) 936-3333.