BIOST/EPI 536
Autumn Quarter 2019
Dr. McKnight
September 26, 2019

BIOST/EPI 536
CATEGORICAL DATA ANALYSIS IN EPIDEMIOLOGY
AUTUMN 2019

PREREQUISITES: BIOST 513 and EPI 514; or BIOST 518 and EPI 514; or BIOST 515; or permission of the instructor

HOURS:
Lecture: Tuesday & Thursday 1:30 - 3:20, Health Sciences T-625

Discussions: Mondays 2:30 - 3:20, Health Sciences T-639
Tuesdays 12:30 - 1:20, Health Sciences T-473

INSTRUCTOR:
Barbara McKnight, Ph.D.
Professor
Department of Biostatistics

F670 Health Sciences
phone: 206-616-8078
email: bmck “at” uw “dot” edu

Office Hours: Tuesdays 11:00 am - noon
Thursdays 9:30 - 10:30 am
or by appointment

TEACHING ASSISTANT: Emily Voldal
Email: voldal “at” uw “dot” edu

Office Hours: Wednesdays 2:00 - 2:50 pm HS T530
Fridays 11:00 - 11:50 am HS T531 (except Nov 29)

REQUIRED NOTES: Available on the class website.
`https://canvas.uw.edu/courses/1318466`.

OPTIONAL TEXTS: (on reserve in HS library)

Breslow N. and Day N., Statistical Methods in Cancer Research, Volume 1: The Analysis of Case Control Studies. IARC Scientific Publications No. 32, Lyon, 1980.

A free .pdf version of this text is available for download from the International Agency for Research Cancer here.

Hosmer and Lemeshow: Applied Logistic Regression, 3rd ed.. Wiley, 2013. Available free as .pdf through UW libraries here.

Kleinbaum D, Klein M, Logistic regression: a self-learning text. Third Edition. New York : Springer, 2010. Available free as a .pdf through UW libraries here.

RECOMMENDED BOOKS: (on reserve in HS library)

Borgan et al. Handbook of Statistical Methods for Case-Control Studies CRC Press, Chapman and Hall, Boca Raton, 2018.

Clayton & Hills: Statistical Models in Epidemiology. Oxford University Press, Oxford, 2013. Available as online ebook and for partial .pdf download through UW libraries here.

Collett: Modelling Binary Data. 2nd edition, Chapman and Hall, Boca Raton, 2003.

Harrel, F: Regression Modelling Strategies. Springer, New York, 2001. Available free as .pdf through UW libraries here.

Jewell, NP: Statistics for Epidemiology, Chapman and Hall, Boca Raton, 2004.

Keogh R.H. and Cox D.R. Case-Control Studies. Cambridge U Press, Cambridge 2014. Available free as a .pdf through UW libraries here.

Pearl, J. Causality: Models, Reasoning and Inference. 2nd Edition. Cambridge Univ. Press, Cambridge, 2009. Available as online ebook and for partial .pdf downlaod through UW libraries here.

VanderWeele. T.J. Explanation in Causal Inference: Methods for Mediation and Interaction., Oxford University Press, Oxford 2018. Available as online ebook and for partial .pdf download through UW libraries here.

Vittinghoff, E., Glidden, DV, Shiboski, SC. (2006). Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models Springer-Verlag New York. Available free as .dpf through UW libraries here.

COMPUTER SOFTWARE:

We will be using the statistical package R (https://www.r-project.org/) with front end R Studio (https://www.rstudio.com/). Both are available at no cost for Windows, Mac and Linux operating systems. Every student is required to have a laptop computer with these installed, and you are encouraged to install them yourself before the first discussion section. We will be using these for exercises in class and in discussion sections and you will use them on homework and the data analysis project. If you do not own a laptop computer, you may borrow one for the quarter from the Student Technology Fee Loan Program (https://itconnect.uw.edu/service/student-technology-loan-program/)

CLASS WEBSITE:

Homework assignments, readings, data, lecture notes, and videos of class sessions will be available on the class Canvas website:

https://canvas.uw.edu/courses/1318466.

A class discussion board, maintained by Emily, will also be available there.

If you are registered for the course, you should have automatic access through the Canvas link on your myUW web page. If you must register late and need access before you register, please email me with a request using your **UW email** address so I can use your uwnetid to give you access.

POLL EVERWHERE: Starting next week, we will be using the Poll Everywhere audience response system in class. We will be using daily mini-quizzes with responses posted to Poll Everwhere in class this quarter. There are several reasons for this: 1) to give me feedback on how much you know and what misconceptions you might have before I begin a topic, so I can better address what you need; 2) to help you practice remembering material we have covered; 3) to give you practice applying what we have learned and solving problems using the skills you are learning with immediate feedback; and 4) to communicate regularly to you my expectations for your mastery of the material we have covered.

After some Poll Everywhere questions you will be asked to pair up and discuss your answers with a neighbor, and then sometimes asked to complete a second poll for the same question. To begin a class discussion, I may also call on you to explain your pair's thinking to the rest of the class; if you do not feel prepared to describe your pair's thinking, you may pass and I will choose someone else.

To get ready to use Poll Everywhere, please go to the site (`https://itconnect.uw.edu/learn/tools/polleverywhere/`), click on Poll Takers: Set Up Your Account and Log In, and set up your account if you do not already have one.

BRING TO CLASS: To be able to participate in classroom exercises, please bring either a tablet or notebook and pen or pencil to class and a phone, tablet or laptop to participate in Poll Everywhere polls. Please bring a laptop to Discussion sections.

ACCESS AND ACCOMODATIONS:

Your experience in this class is important to me. If you have already established accommodations with Disability Resources for Students (DRS), please communicate your approved accommodations to me at your earliest convenience so we can discuss your needs in this course.

If you have not yet established services through DRS, but have a temporary health condition or permanent disability that requires accommodations (conditions include but not limited to; mental health, attention-related, learning, vision, hearing, physical or health impacts), you are welcome to contact DRS at 206-543-8924 uwdrs@uw.edu or http://depts.washington.edu/uwdrs/. DRS offers resources and coordinates reasonable accommodations for students with disabilities and/or temporary health conditions. Reasonable accommodations are established through an interactive process between you, your instructor(s) and DRS. It is the policy and practice of the University of Washington to create inclusive and accessible learning environments consistent with federal and state law.

RELIGIOUS ACCOMODATIONS:

Washington state law requires that UW develop a policy for accommodation of student absences or significant hardship due to reasons of faith or conscience, or for organized religious activities. The UWs policy, including more information about how to request an accommodation, is available at Faculty Syllabus Guidelines and Resources. Accommodations must be requested within the first two weeks of this course using the Religious Accommodations Request form available at https://registrar.washington.edu/students/religious-accommodations-request/.

HOMEWORK:

Homework will be assigned regularly throughout the quarter, and you will always be given a week to complete it.

I view the homework in this class as an important part of the process of your learning. You are encouraged to do the best work you can. It is fine to work together with others on written homeworks to solve problems and/or to ask Emily or me for answers to questions about how to do something in R. However, I expect that your written homework gives a summary of your personal understanding of the answers, independent of of others. **Late homework will not be accepted.**

Homework will be marked 10 points if it represents a good faith effort to answer all parts of all questions and answers are substantially correct, and 8 if all but one or two small parts of questions are answered with a good faith effort or there are substantive errors in the answer to at least one problem, and 5 if there are substantive errors on more than one problem. More incomplete homework will be scored zero.

After homework has been graded and the key posted, you are expected to read the key, and study it if necessary, so that you are able to answer the questions without reference to the key.

PROJECT:

A data analysis project will be assigned by mid-quarter. You will be assigned to a group and asked to work together to analyze data from an Epidemiologic case-control study and write an article reporting your results.

ACADEMIC INTEGRITY:

Students at the University of Washington (UW) are expected to maintain the highest standards of academic conduct, professional honesty, and personal integrity.

The UW School of Public Health (SPH) is committed to upholding standards of academic integrity consistent with the academic and professional communities of which it is a part. Plagiarism, cheating, and other misconduct are serious violations of the University of Washington Student Conduct Code (WAC 478-120). We expect you to know and follow the university's policies on cheating and plagiarism, and the SPH Academic Integrity Policy. Any suspected cases of academic misconduct will be handled according to University of Washington regulations. For more information, see the University of Washington Community Standards and Student Conduct website.

Please note that all of the following count as academic misconduct: using a cellphone or other communication device to communicate with others during an exam, talking or sharing answers during an exam (unless specifically directed to do so by the instructor during some discussion-based exam portions) .

EMAIL:

I am happy to answer email questions when I have time, but I am not often able to give prompt replies except during office hours when no students are in attendance. Sending email is usually not the most efficient option for you to get your questions answered: it usually ends up taking you less time to attend my or Emily's office hours, and we encourage you to do so.

LEARNING ENVIRONMENT:

I take seriously my role as an advocate for your learning in this class. In addition to providing information, assignments and activities that I hope will support your learning, I will do my best to help us maintain the classroom as a supportive learning environment that respects diversity: gender identity, sexuality, disability, age, socioeconomic status, ethnicity, race, nationality, religion and culture. If you have a concern, I encourage you to communicate it to me. Please let me know ways to improve the effectiveness of the course for you personally, or for other students. If you make a request and you do not feel my response has been adequate, please contact the chair of the department at `linoue ‘‘at’’ uw ‘‘dot’’ edu`.

To maintain a respectful and welcoming classroom environment, I ask that we all commit to showing respect to each other both inside and outside of class.

COURSEWORK:	Homework	Regularly
	Midterms: During the last 1.0 to 1.5 hours of class (closed book)	Th Oct 24, Th Nov 21
	Data Analysis Project:	Thursday, December 5 (9:00 am)
	Final Exam: (closed book)	Fri December 13, 2:30-4:20 pm

GRADING: Numerical class grades will be based on:

Final Exam (30%),
Midterms (40%),
Data Analysis Project (20%)
Homework (after dropping lowest score) (10%)

BIOST/EPI 536
Autumn Quarter 2019
Dr. McKnight
September 26, 2019

BIOST/EPI 536
CATEGORICAL DATA ANALYSIS IN EPIDEMIOLOGY
AUTUMN 2019
LEARNING OBJECTIVES

Upon entering this course, you are expected to have completed courses in introductory statistics or biostatistics, multiple regression, and categorical and censored survival data analysis. You should understand the basic statistical concepts of sampling variation, parameter estimation and statistical hypothesis tests. You should know how to fit multiple linear regression models, how to interpret regression coefficients in multiple linear regression models and how to perform hypothesis tests about regression coefficients. You should be familiar with case-control, cross-sectional and cohort study designs. You should know how to use simple statistical techniques for analyzing data from the binomial distribution including odds ratio estimation in 2×2 tables and a series of 2×2 tables, the Mantel-Haenszel test, and tests about the odds ratio from 2×K tables including tests for heterogeneity and tests for linear trend and departure from linear trend. After successfully completing this course, you can ordinarily expect to be able to:

1. Fit appropriate logistic regression models to data from epidemiologic case-control studies using R and evaluate the fit of these models.
2. Fit appropriate Poisson regression models to data from epidemiologic cohort studies using R and evaluate the fit of these models.
3. Interpret regression coefficients from logistic and Poisson regression models fit to case-control and cohort data and test hypotheses about them.
4. Describe the difference in interpretation between an odds ratio adjusted using regression and an odds-ratio adjusted using marginal methods.
5. Explain when logistic regression methods should be replaced by conditional logistic regression methods.
6. Conduct an appropriate analysis of epidemiologic case-control study data and document it so that someone else with the same data could reproduce the results.
7. Present results of analyses using logistic regression to readers who are not familiar with logistic regression.

BIOST/EPI 536
Autumn Quarter 2019
Dr. McKnight

BIOST/EPI 536
CATEGORICAL DATA ANALYSIS IN EPIDEMIOLOGY
COURSE OUTLINE

TOPIC (approximate # days)	**TEXTBOOK READING**
1. INTRODUCTION AND REVIEW (2)	
Measures of Disease Occurrence & Association	BD 2
Confounding and Collapsibility	BD: 3.4
Marginal and Conditional Confounding Control	
2. LOGISTIC REGRESSION (9)	
Simplest Models in Cohort Studies	BD: 6.1; 6.2; HL: 6.2, 3.2
Extension to Case/Control Studies	BD: 6.3; HL: 6.3
General Logistic Model	BD: 6.2; 6.3; HL 1.1-1.4; 2.1-2.2, 3.1
Using R to Fit Logistic Regression Models	
Estimators and Test Statistics	BD: 6.4; HL: 2.3, 2.4,2.5; KK 5
Categorical Confounder Adjustment Using Regression	BD: 6.5; 6.6 ; KK 6, 7.III
Continuous Confounder Adjustment Using Regression	BD: 6.7; 6.8; 6.9; HL: 3.5, 4.2.1
Exposure Variable Models	BD: 6.9; 6.10; 6.11; HL: 3.2, 3.3, 3.4; KK 6
Precision Variables	
Effect Modification on the Multiplicative Scale	BD: 6.12; HL: 3.5; KK 6; 7.II
Advanced Coding	
Variable Selection & Model Choice	
Regression Diagnostics	HL: 5.3; Collett

TOPIC (approximate # days)	**READING**
3. LOGISTIC REGRESSION USING A CONDITIONAL LIKELIHOOD (2)	
Data and Models Requiring Conditional Likelihood	BD: 7.1; 7.6; HL: 7.1; KK 11.IV
Differences from Regular Logistic Regression	BD: 7.2; HL: 7.1
Applications: Matched Analysis as a Special Case	BD: 7.3, 7.4; HL: 7.3, 7.4 KK 11.V, 11.VI, 11.VII
Using R to Fit Models Using the Conditional Likelihood	
Comparision to Ordinary Logistic Regression when Conditional Logistic Regression Not Needed	
Goodness of Fit; Regression Diagnostics	
4. OTHER METHODS FOR CONTROLLING CONFOUNDING (2)	
Causality and Conditional vs. Marginal Models	
Confounder Adjustment Using Standardization	
Propensity Score Adjustment	HL 10.2
Stratification Score Adjustment	
5. SPECIAL TOPICS (2) (some but not all of these, depending on time)	
Logistic Regression for Prediction	HL 4.3, 5.24; KK 10
Polytomous Logistic Regression	HL 8.1; KK 12
Sample Size / Power Calculations	HL 10.5
Effect Modification on the Additive Scale	HL 10.9
Mediation analysis	HL 10.8
Two-phase Designs	
Mendelian Randomization	
Counter-Matching	
6. REVIEW (1)	