Advanced Epidemiological Analysis

Andreas Neophytou and G. Brooke Anderson

2021-03-23

Contents

Chapter 1

Overview

This is a the coursebook for the Colorado State University course ERHS 732, Advanced Epidemiological Analysis. This course provides the opportunity to implement theoretical expertise through designing and conducting advanced epidemiologic research analyses and to gain in-depth experience analyzing datasets from the environmental epidemiology literature.

1.1 License

Click on the **Next** button (or navigate using the links in the table of contents) to continue.

Chapter 2

Course information

This is a the coursebook for the Colorado State University course ERHS 732, Advanced Epidemiological Analysis. This course provides the opportunity to implement theoretical expertise through designing and conducting advanced epidemiologic research analyses and to gain in-depth experience analyzing datasets from the environmental epidemiology literature.

This class will utilize a variety of instructional formats, including short lectures, readings, topic specific examples from the substantive literature, discussion and directed group work on in-course coding exercises putting lecture and discussion content into practice. A variety of teaching modalities will be used, including group discussions, student directed discussions, and in-class group exercises. It is expected that before coming to class, students will read the required papers for the week, as well as any associated code included in the papers' supplemental materials. Students should come to class prepared to do statistical programming (i.e., bring a laptop with statistical software, download any datasets needed for the week etc). Participation is based on in-class coding exercises based on each week's topic. If a student misses a class, they will be expected to complete the in-course exercise outside of class to receive credit for participation in that exercise. Students will be required to do mid-term and final projects which will be presented in class and submitted as a written write-up describing the project.

Prerequisites for this course are:

- ERHS 534 or ERHS 535 and
- ERHS 640 and
- STAR 511 or STAT 511A or STAT 511B

2.1 Course learning objectives

The learning objectives for this proposed course complement core epidemiology and statistics courses required by the program and provide the opportunity for students to implement theoretical skills and knowledge gained in those courses in a more applied setting.

Upon successful completion of this course students will be able to:

1. List several possible statistical approaches to answering an epidemiological research questions. (*Knowledge*)
2. Choose among analytical approaches learned in previous courses to identify one that is reasonable for an epidemiological research question. (*Application*)
3. Design a plan for cleaning and analyzing data to answer an epidemiological research question, drawing on techniques learned in previous and concurrent courses. (*Synthesis*)
4. Justify the methods and code used to answer an epidemiological research question. (*Evaluation*)
5. Explain the advantages and limitations of a chosen methodological approach for evaluating epidemiological data. (*Evaluation*)
6. Apply advanced epidemiological methods to analyze example data, using a regression modeling framework. (*Application*)
7. Apply statistical programming techniques learned in previous courses to prepare epidemiological data for statistical analysis and to conduct the analysis. (*Application*)
8. Interpret the output from statistical analyses of data for an epidemiological research question. (*Evaluation*)
9. Defend conclusions from their analysis. (*Comprehension*)
10. Write a report describing the methods, results, and conclusions from an epidemiological analysis. (*Application*)
11. Construct a reproducible document with embedded code to clean and analyze data to answer an epidemiological research question. (*Application*)

2.2 Meeting time and place

[To be determined]

2.3 Course grading

Assessment Components	Percentage of Grade
Midterm written report	30
Midterm presentation	15
Final written report	30
Final presentation	15
Participation in in-course exercises	10

2.4 Textbooks and Course Materials

Readings for this course will focus on peer-reviewed literature that will be posted for the students in the class. Additional references that will be useful to students throughout the semester include:

- Garrett Grolemund and Hadley Wickham, *R for Data Science*, O'Reilly, 2017. (Available for free online at https://r4ds.had.co.nz/ and in print through most large book sellers.)
- Miguel A. Hernan and James M. Robins, *Causal Inference: What If*, Boca Raton: Chapman & Hall/CRC, 2020. (Available for free online at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/01/ciwhatif_hernanrobins_31jan21.pdf with a print version anticipated in 2021.)
- Francesca Dominici and Roger D. Peng, *Statistical Methods for Environmental Epidemiology with R*, Springer, 2008. (Available online through the CSU library or in print through Springer.)

Chapter 3

Time series / case-crossover study designs

3.1 Time series data

[Exploring time series data with daily measurements of health outcomes and environmental exposures]

3.2 Fitting models

[Fitting models under time series and case-crossover study designs]

Chapter 4

Generalized linear models

4.1 Splines in GLMs

[Using splines to model non-linear associations in a GLM]

4.2 Cross-basis functions in GLMs

[Using a cross-basis to model an exposure's association with the outcome in two dimensions (dimensions of time and exposure level)]

Chapter 5

Natural experiments

5.1 Interrupted time series

[Interrupted time series assessing effects of policy/intervention in specific point in time]

5.2 Difference-in-differences

[Difference-in differences application for intervention introduced in one point in time]

Chapter 6

Risk assessment

[Predict expected heat-related mortality under a climate change scenario]

Chapter 7

Longitudinal cohort study designs

7.1 Coding a survival analysis

7.2 Handling complexity

7.2.1 Multi-level exposure

7.2.2 Recurrent outcome

7.2.3 Time-varying coeffificents

7.2.4 Using survey results

[e.g., NHANES]

Chapter 8

Some approaches for confounding

8.1 Inverse probability weighting

8.2 Propensity scores

[Modeling for weights/propensity scores, involves machine learning]

Chapter 9

Mixed models

[Using a mixed modeling framework to help analyze repeated measures]

Chapter 10

Instrumental variables

Chapter 11

Causal inference