BIOST/STAT 579: DATA ANALYSIS & REPORT WRITING, Sp 2019

Instructor:
• Mary Emond, Ph.D. Associate Professor of Biostatistics
• Office = 15th Floor, UW Tower, NW window area.
• Office Hours = By Appointment. Email questions kk.
• Email = emond at uw dot edu

Course Prerequisites:
BIOST/STAT 570 or permission of the instructor

Course Schedule:
Time: 10:30 a.m. to 12:20 p.m.
Days: Tuesdays
Location: Room 010, Anderson Hall (AND)
***Due dates for the three major assignments*, 5pm on:**
Ap 14 (abstract, methods, results, tables/figs)
Ap 28 (full report)
June 2 (full report)
Because a primary purpose of this class is to prepare for the Applied Exam which has a strict deadline, no late work will be accepted for these three assignments. Submit whatever you have done by the deadline. A late assignment will result in No Credit for the class except in cases of unpredictable, severe and time-consuming emergencies. (Plan ahead for weddings.)

Other assignments will comprise class presentation of intermediate analysis results and potentially summaries of background reading (scientific or statistical) that enables one to perform a good analysis.

Course Description:
The purpose of this course is to provide hands-on experience and instruction in data analysis and scientific report writing. Classroom activities will include some lecture-style instruction, group discussions and peer critique, feedback on assignments, and oral+media presentations by students (usually in pairs). Emphasis is on participation. ***Attendance is required.*** While presentations can be joint and discussion of ideas is encouraged, reports should be written by yourself alone.

Upon entering this course, you are expected to have completed courses covering 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, logistic and Cox regression models, how to interpret regression coefficients in multiple regression models and how to perform hypothesis tests about regression coefficients. You should be familiar with the concept of explained variation in regression models. You should also be familiar with analysis methods for some form of correlated data, such as longitudinal data or family data, and you should be familiar with a variety of

experimental and observational study designs, including the randomized experiment, and case-control and cohort studies. You should also understand the difference between inference and prediction.

Learning Objectives:
The overall goal of the course is to improve your skills in the analysis of scientific data and report writing. The focus is not on learning new analysis techniques, though you might find it useful to venture into extensions of methods already learned. You will improve your ability to select and implement a sound approach to addressing the scientific question(s) the data were collected to answer and *to describe the results in a meaningful and accessible way*.

Becoming a truly proficient scientific data analyst is a process that takes many years, exposure to many types of data collected to answer many different scientific questions along with continual attention to ways to improve. (Often, statisticians concentrate on particular subfields of science in order to develop a deeper understanding of the science that promotes better-suited analyses and/or new methods development.) This quarter will provide the types of exposure and the opportunity for reflection and comparison that will continue throughout your career.

After successfully completing this course, you can ordinarily expect to have improved your ability to:

• Formulate a scientific question in statistical terms, choose an appropriate target of inference and develop a statistical approach to answering the question.
• Determine whether, and/or under what assumptions, the data contain information to answer the scientific question(s) of interest.
• Recognize key features of the data, including overall features, such as the study design and correlation structure, and incidental features such as influential observations; choose and implement statistical methods that allow you to draw valid inferences in the presence of these features, if possible.
• Use descriptive statistics and graphical displays to communicate relevant information contained in the data to the statistically untrained reader. (*Aside: a picture is worth a thousand words; try to create a key figure or figures that impart the primary result.)*
• Recognize assumptions required for the validity of inferences from the statistical methods you use; check these assumptions, if possible, or conduct a sensitivity analysis of the effects of departures from the assumptions, if not.
• Organize and write clear, succinct scientific reports that communicate the results of your data analysis and the conclusions you draw from it to a statistically untrained reader; describe in the report any limitations on your ability to answer the scientific question.

Grading Policy:
This course is graded C/NC. Students are expected to attend class regularly and to actively participate in discussion to receive credit. Any absences must be discussed with the instructor, preferably in advance (e.g. your best friend's wedding; obviously, this does not apply to sudden illness.) Assignments are graded as pass/fail, but reports also will be

given scores from 1 to 5 on Science, Statistics and Writing as is done for the Applied Exam. Assignments will involve one or more of the following: reading and critiquing scientific papers, data analysis, writing and presentation.

Statistical Software:
You may use any statistical software you wish to perform data analyses in this course. We assume that most of you will be using R. Learning statistical package syntax is NOT an explicit objective of this course, though you are encouraged to share code and tips with each other.

Data:
Data sets will be made available on the course website in one or more of the following formats: text, Excel, .CSV