BIOST/STAT 571: Advanced Regression Methods for Dependent Data

Course Information: Winter 2022

About This Course

- Advanced Regression Methods for Dependent Data:
  - Correlated data

- What does that mean? We deal with Correlated data all the time:
  - Example:
    - Case-control study: y = Cancer case/control status (response, dependent variable)
    - Predictors: Age, Sex, BMI, Education, Ethnicity, genetic markers, smoking
    - These predictors are all correlated
  - In general, when we talk about correlated and dependent data, we refer to correlation among the outcomes

Dependent, Correlated, Clustered, Longitudinal, Multivariate Data

- Longitudinal data
  - Collecting data on subjects over time
- Hierarchical data
  - Observations exist within a hierarchy
- Spatial data
  - Observations have correlation due to spatial characteristics
- Multivariate data
  - Measuring multiple characteristics of a subject as outcome
  - Repeated measures of an assay
  - Gene expression profiling experiment
  - Microbiome experiment

What this course will do for you

- Understand what are different types of correlated data
- Have a toolbox of methods for addressing correlated data that you understand well and know when they are relevant
- Know how to derive or empirically evaluate the properties of these methods
- Know how to implement the methods using existing software or through de novo coding
- Have the necessary foundations to work on developing new methods when a new problem arises

What this course will NOT do:

- Provide a solution for every possible correlated data problem
- Cover every existing method that exists for correlated data

Instructor

Michael Wu

- Professor, Fred Hutch
- Affiliate Associate Professor in Biostats

- Usual Address: M3-C821 (Fred Hutch)
- email: mcwu@fredhutch.org & mcwu2004@uw.edu

- Office Hours:
  - TBD
  - OR By Appointment

- Understudy: Ken Rice
  - Professor of Biostatistics
  - Singer
  - Scottish History Buff

Teaching Assistants

Si Cheng

- chengsi@uw.edu

Rahul Biswas

- rbiswas1@uw.edu

Office Hours:

- TBD
- OR By Appointment

Course

- Mon/Wed/Fri 1:30-2:30 PM
  - Monday January 3 through Friday March 11, Except Holidays: MLK Jr and Presidents' Days
- Most Lectures will be Live: https://washington.zoom.us/j/95493890665?pwd=MmRSRTFPeWRwOGZiRHNjTnhSK1FCQT09
- Lectures will be recorded; some lectures may be pre-recorded
- Course website: https://canvas.uw.edu/courses/1514650
- Prerequisite: BIOST/STAT 570

Texts

- No required texts

- Recommended:
  - Diggle, Heagerty, Liang and Zeger (2002). *Analysis of Longitudinal Data*
  - Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*

- Others
  - McCullagh, P. and Nelder, P.(1989). *Generalized Linear Models*
  - Wakefield, J. (2014). *Bayesian and Frequentist Regression Analysis*
  - Rencher, A. C. (2002). *Methods for Multivariate Analysis*
  - Johnson, R.A., and Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis*
  - Verbeke, G. and Mollenberghs, G. (2005). *Models for Discrete Longitudinal Data*

  - Hastie, T., Tibshirani, R. and Friedman, D. (2001). *The Elements of Statistical Learning*

Where to go for help:

1. Textbooks
2. Google
3. Peers
4. TA’s
5. Instructor

Grading

- Assignments and grading are unfortunate but required components for this course.

- Homework
  - 70% of final grade
  - Posted and Handed in via Canvas
  - **Encouraged to work with peers.**
  - Write up your own work: Please don’t just copy
  - No Late Homework and No Extensions
  - Not all homework problems may be graded (solutions will be provided)
  - Students may request regrades for homeworks

Regrade Policy

- Students that perceive that a problem on any assignment has been incorrectly graded must submit for a full regrade of the entire assignment per the regrade procedure

- Regrade Procedure:
  - Entire assignment must be submitted for regrade within 1 week of the assignment's return.
  - For each potential grading error, students must write a full paragraph (4-5 sentences)
  - clearly detailing why they believe their solution is correct/deserves additional credit.
  - Regrades will be conducted with regard to the entire assignment, i.e. all problems will be reconsidered

- Regrades may result in higher, unchanged, OR lower scores (grading errors can occur in both directions).

- Regraded scores are Final.

- Mistakes in adding up points do not require full regrade.

Final

- 30% of your final grade

- Group Project/Term Paper
  - 3-4 Students
  - Assigned part-way through the quarter
  - Last year: development of new statistical methodology
  - This year: still in development
  - More details provided in Feb

Academic Integrity

- https://sph.washington.edu/students/academic-integrity-policy
- Faculty are responsible for reporting *suspected* violations.
- Please don't cheat. If it's questionable, then talk to the TA and Instructors first.
- See syllabus for full discussion

Other Important Topics in the Syllabus

- Communication Skills
- Access and Accommodations
- Religious Accommodations
- Inclusion and Diversity
- Classroom Climate
- Pronouns
- Bias Concerns
- Cheugy References
- Sexual Harassment

Approximate List of Topics

1. Introduction to Correlated Data
2. **Linear Mixed Models**
3. **Generalized Estimating Equations**
4. **Generalized Linear Mixed Models**
5. Missing Data
6. Classical and Modern Multivariate Analysis