**Syllabus**
**BIOSTAT 515/518**
**Applied Biostatistics II: Introduction to Regression Analysis**
**Winter 2019**
**Syllabus date: January 4, 2019**

All information is subject to change. Students are responsible for changes as announced in class and/or on the class CANVAS site.

**Instructor** Kathleen Kerr, Ph.D.
Associate Professor of Biostatistics
Email: katiek@uw.edu
Office: HSB F667 (see CANVAS for office hours)

**Teaching Assistants** Subodh Selukar
David Whitney
Hyunju Son
TA Office Hours -- see CANVAS

**Time and Place** Lecture: MWF HSB T747, 9:30-10:20
Discussion AA: Monday 8:30-9:20 in HSB T635
Discussion AB: Wednesday 8:30-9:20 in HSB T635
Discussion AC: Friday 8:30-9:20 in SOCC 301

**Textbook** Vittinghoff, Glidden, Shiboski, McCulloch: Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models, 2nd edition. Springer. In case of conflict, lecture material takes precedence over textbook material.

**Grading** Final exam: 35%
Midterm exam: 30%
Data Analysis Project: 15%
Homework: 20%

**Course Objectives**

This course provides an introduction to the principles and application of regression methods for the statistical analysis of data. The course is designed for graduate students in public health who are already familiar with basic statistical concepts, including descriptive statistics, the components of statistical inference (p-values, hypothesis tests, confidence intervals, etc.), and concepts such as confounding and effect modification.

Specific topics will include linear regression models, logistic regression models, and Cox proportional hazards regression models for censored data. We will learn how to handle covariates such as confounding variables, effect modifiers, and precision variables in the regression setting.

Learning Objectives:

1. Identify types of questions for which a regression analysis is appropriate.
2. Identify and interpret the parameters of interest in linear, logistic, and proportional hazards regression models.
3. Perform a regression analysis, including
    - Identify outcome and predictor of interest

- Identify and justify model covariates, explain their role in the analysis
- Identify appropriate transformations
- Employ dummy variables as appropriate
- Employ interactions terms as appropriate
- Interpret regression parameters, p-values, and confidence intervals
- Demonstrate proficiency with a statistical software package

4. Write a clear and concise scientific report based on a regression analysis
5. State and explain the assumptions underlying the conclusions of a regression analysis.
6. Identify when a regression analysis can address a scientific question of interest, and explain the limitations of the analysis for answering the question.

**Software**

The "official" software package for this course is *R*. "Official" means that the instructor and teaching assistants will help with questions related to conducting data analyses in *R*. However, this is not a course in *R*. Examinations may ask students to interpret software output but will not examine knowledge of the *R* coding language. Students may use any statistical package that accomplishes the tasks necessary for the homework assignments and data analysis project.

**Discussion Section**

Discussion section will be used for multiple purposes: discussing and expanding on course material, additional topics, and actively applying methods to datasets. Students may be required to conduct a first-pass analysis of specified datasets prior to Discussion Section.

**Class CANVAS Site and Lecture Recordings**

The class CANVAS site serves as an archive of homework, handouts, lecture notes, and datasets. Students should check CANVAS regularly for information. Lecture notes will be made available on CANVAS the day before class. For students' convenience, Panopto recordings of lectures will be posted when possible. However, this is not a distance learning class, and students are responsible for all material in lectures regardless of whether a recording is available.

Time-sensitive announcements will be broadcast to the class through CANVAS. You may want to set your CANVAS settings to notify you when there are announcements or other updates.

**Questions on course material and email policy**

Students who have questions on course material have the following resources: raise the question in class, raise the question in discussion section, attend office hours, ask the question on the CANVAS discussion board. The CANVAS discussion board is particularly useful for software questions.

Email is not an appropriate venue for asking questions about course material. Such emails to the instructor or TAs will not receive a response.

Please DO email the instructor if you think something is missing from CANVAS (e.g. lecture notes for the next day, homework assignment due within the next week, etc.).

**Homework**

Homework assignments will be due approximately weekly. Students may consult with each other, the instructor, and the TAs on homework, but the submitted assignment should reflect a student's own work. In other words, it is fine to discuss the homework with others but not to copy someone else's solution.

Due to the size of the class, **late homework cannot be accepted** (even for good reasons). Because of this strict policy, the lowest homework score for each student will be dropped when calculating course grades. Note: there will be "grace period" of at least 1 hour after the due date/time before students can no longer upload homework. An assignment submitted during this grace period is eligible for full credit even if CANVAS marks it late.

In order to receive credit, homework should be neat, well-organized, and written in clear, grammatically-correct English using complete sentences. Raw output from software is unacceptable. Plots should be labeled, including axis labels, and options such as scale should be chosen to make the plot as informative as possible. Homework submissions that do not meet these standards will not be credited.

Homework is viewed as part of the learning experience of this course and not as a tool for evaluating mastery of course material. Therefore, homework will be graded based on a **good faith effort** to answer **all** homework questions. Submitted assignments demonstrating a good faith effort to all questions will receive the maximum of 10 points.

Each homework assignment is worth 10 points. Assignments submitted by the deadline will receive either 10, 9, or 0 points based on whether the student demonstrates a good faith effort to answer all of the questions on the assignment. Part of demonstrating a good faith effort is following the guidelines for homework above.

10: A good-faith effort was made on all parts of all problems.
9: A good-faith effort was made on all but very minor parts of one or a few problems. (For example, omitting a small component in part of a question). Or homework was complete but was slightly below-standard in some places.
0: At least one problem, or many parts of some problems did not receive a good-faith effort. (For example, not attempting to answer part of a question, or not attempting a whole question, or pasting software output rather than answering a question)

As mentioned above, the lowest homework grade will be dropped. As late homework is never accepted, student should save this for an unexpected situation (illness, family emergency, etc.)

A solution key will be provided after homework is graded. Because it is not possible to provide a comprehensive evaluation of every student's submitted homework, students should study the keys and compare them to their own work. Material in these keys is part of the course material and may appear on exams, even if it has not been covered in class.

**Exams**

There will be an in-class midterm in early February and a final exam during finals week. Exams are closed book. Bring a pencil, eraser, and simple calculator to the exams. Calculators cannot have web or data connection (you cannot use your cell phone, tablet, etc. as a calculator).

**Group Data Analysis Project**

A Data Analysis project will be due in approximately Week 8. It will be an analysis of data using the methods covered in class, summarized in a formal report. The instructor will organize students into small groups for completing the project. More information about the project will be provided during the quarter.

**Access and Accommodation**

The experience of every student in this class is important, and 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 or uwdrs@uw.edu or disability.uw.edu. 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.

**Academic Integrity**

http://sph.washington.edu/students/academicintegrity/

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.

Here are links for the text above in boldface:
UW Student Conduct Code (WAC 478-120)
http://www.washington.edu/cssc/student-conduct-overview/student-code-of-conduct/
SPH Academic Integrity Policy
http://sph.washington.edu/students/academicintegrity/
Community Standards and Student Conduct
http://www.washington.edu/cssc/

**Classroom Climate**

The UW School of Public Health seeks to ensure all students are fully included in each course. I strive to create an environment that reflects mutual respect and a shared desire to learn. I encourage students with concerns about classroom climate to talk to me, your advisor, a member of a departmental or SPH Diversity Committee and/or the program director.

**TA Concerns**

If you have any concerns about your TA, please see the TA about these concerns as soon as possible. If you are not comfortable talking with the TA or not satisfied with the response that you receive, you may contact the Department of Biostatistics Associate Director of Academic Affairs (biostgp@uw.edu). If you are still not satisfied with the response that you receive, you may contact the Department of Biostatistics Chair (bchair@uw.edu). You may also contact the Graduate School at G-1 Communications Building, by phone at 206-543-5139 or by email at raan@uw.edu.

**Acknowledgement**

Some course material is adapted from previous instructors, in particular Tim Thornton, Scott Emerson, and Thomas Lumley.

| **Approximate Sequence of Topics**<br>Subject to change<br>1 Lecture ≠1 Day | | |
|---|---|---|
| Lecture | Topic | Related Text Reading<br>(2nd Edition) |
| 1 | Course Introduction/Review | 1.1, 1.2, 2.4.1, 3.7 |
| 2 | Motivation and introduction to linear regression | 3.3.1, 3.3.4, 3.3.9, 4.0, 4.1 |
| 3 | Classical Linear Regression and Extensions | 3.3.2, 3.3.5 |
| 4 | Linear regression and t-tests | |
| 5 | Linear regression and inference about associations | 3.3.8 |
| 6 | Linear regression and inference about group means and prediction | |
| 7 | Assumptions for linear regression: recap | |
| 8 | Regression to the Mean | |
| 9 | Multiple regression | 4.2 |
| | Multiple regression, focus on interaction terms | |
| | Weighted regression | 12.1, 12.2 |
| | Nested Models, ANOVA tables, and F-tests | 3.3.6, 3.3.7, 4.3.1, 4.3.2, 4.3.3 |
| | Regression diagnostics | 4.7.4 |
| | Introduction to logistic regression | 5.0 |
| | Logistic regression | 5.1 |
| | Inference in logistic regression | 5.2.0, 5.2.1 |
| | Multiple logistic regression | 5.2.2, 5.2.3, 5.2.4 |
| | Logistic regression: advanced topics ; matched data | 5.3, 5.6 |
| | Alternatives to logistic regression | 5.5.1, 6.5.2, 5.5.3 |
| | Prediction of Binary Outcomes | 5.2.5, 5.2.6, 10.1 |
| | Censored data and Cox regression | 3.5. 6.0, 6.1 |
| | Regression methods for survival data | 6.2 |
| | Proportional hazards regression with multiple predictors; adjusted survival curves | |
| | Introduction to Poisson regression | |
| | Analysis of variance | 3.1.4, 7.0, 7.1 |