SYLLABUS & COURSE INFORMATION

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

PUBH 7406, SECTION 001

Advanced Regression and Design
Spring 2019

COURSE & CONTACT INFORMATION

Credits: 4
Meeting Day(s): Tuesday and Thursday
Meeting Time: 2:30--4:25 pm
Meeting Place: Weaver-Densford Hall W2120

Instructor: Haitao Chu and Baolin Wu
Email: chux0051@umn.edu (H Chu), baolin@umn.edu (B Wu)
Office Phone: 612-625-2138 (H Chu), 612-624-0647 (B Wu)
Fax:
Office Hours:
Office Location: A426 (H Chu), A442 (B Wu), Mayo Building
Teaching Assistants: Evan Olawsky olaws004@umn.edu and Bin Guo <guoxx617@umn.edu>

COURSE DESCRIPTION

Topics include maximum likelihood estimation, single and multifactor analysis of variance, logistic regression, log-linear models, multinomial logit models, proportional odds models for ordinal data, gamma and inverse-Gaussian models, over-dispersion, analysis of deviance, model selection and criticism, model diagnostics, and an introduction to non-parametric regression methods. R and/or SAS are used for illustration.

COURSE PREREQUISITES

- Statistics at the level of Stat 5101 or Stat 8101, and PubH 7405
- R programming experience
- Some familiarity with matrix notation
- Co-requisite: enrollment in Stat 5102 or Stat 8102

COURSE GOALS & OBJECTIVES

This course will focus on the art and science of building, fitting, and diagnosing various kinds of advanced regression models. Assignments will require reading, writing, deriving/proving, and programming. Biostatistics students are strongly encouraged to typeset their work using LaTeX or in R markdown. Computing will be done using R or SAS.

METHODS OF INSTRUCTION AND WORK EXPECTATIONS

Course Workload Expectations
PubH7406 is a 4 credit course. The University expects that for each credit, you will spend a minimum of three hours per week attending class or comparable online activity, reading, studying, completing assignments, etc. over the course of a 15-week term. Thus, this course requires approximately 180 hours of effort spread over the course of the term in order to earn an average grade.

Learning Community
School of Public Health courses ask students to discuss frameworks, theory, policy, and more, often in the context of past and current events and policy debates. Many of our courses also ask students to work in teams or discussion groups. We do not come to our courses with identical backgrounds and experiences and building on what we already know about collaborating, listening, and engaging

is critical to successful professional, academic, and scientific engagement with topics. In this course, students are expected to engage with each other in respectful and thoughtful ways.

Class meetings will be a mixture of lecture and discussion. Students are expected to attend class, participate in class discussions, and complete all reading and homework assignments. Expectations for completing homework are described below. We do not believe in straight lecturing of the material and there will be frequent opportunities for students to work out examples and investigate concepts during class. You should come prepared to be actively engaged in class. Because we will frequently use R and/or SAS in class, it would be helpful if a majority of students brought laptops to class. This course will use Canvas to post content and assignments. The Canvas page can be accessed by going to canvas.umn.edu.

Like other work in the course, all student to student communication is covered by the Student Conduct Code (https://z.umn.edu/studentconduct).

COURSE TEXT & READINGS

Lecture notes and other materials will be available on the Canvas course website. There is no required text, but:

1) Agresti, A. (2012). *Categorical Data Analysis*. Wiley.
2) Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2004). *Applied Linear Statistical Models*. McGraw-Hill/Irwin.

are highly recommended. Additionally, you may find useful one or more of the books listed below:

3) Axler, S. (1997). *Linear Algebra Done Right*. Springer.
4) Christensen, R. (2011). *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer.
5) Faraway, J. J. (2005). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman & Hall/CRC.
6) Hogg, R. V., McKean, J. W., and Craig, A. T. (2005). *Introduction to Mathematical Statistics*. Pearson Prentice Hall.
7) Maindonald, J. and Braun, J. (2007). *Data Analysis and Graphics Using R: An Example-Based Approach*. Cambridge University Press.
8) Monahan, J.F. (2008). *A Primer on Linear Models*. Chapman & Hall/CRC
9) Ravishanker, N. and Dey, D. K. (2002). *A First Course in Linear Model Theory*. Chapman & Hall/CRC.
10) Vittinghoff, E., Glidden, D. V., Shiboski, S. C., and McCulloch, C. E. (2012). *Regression Methods in Biostatistics*. Springer.
11) Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer.
12) Wichura, M. J. (2006). The Coordinate-Free Approach to Linear Models. Cambridge University Press.

COURSE OUTLINE/WEEKLY SCHEDULE

KNNL for Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2004)
CDA for Agresti, A. (2012). *Categorical Data Analysis*.

Week	Topic	Readings	Activities/Assignments
Week 1 1/21-1/25	Single-factor ANOVA	Lecture notes, KNNL Chapter 16,17,18	
Week 2 1/28-2/1	Single-factor ANOVA		
Week 3 2/4-2/8	Single-factor ANOVA		Homework 1
Week 4 2/11-2/15	Multi-factor ANOVA	Lecture notes, KNNL Chapter 19,20,23	
Week 5 2/18-2/22	Multi-factor ANOVA		Homework 2
Week 6 2/25-3/1	Random-effects ANOVA	Lecture notes, KNNL Chapter 25	
Week 7 3/4-3/8	Random-effects ANOVA		Homework 3 Midterm exam (3/7)
Week 8 3/11-15	Contingency Tables	Lecture Notes, CDA Chapter 2 &3	
Week 9 3/18-22	Spring Break, No Lectures		
Week 10 3/25-29	Generalized Linear Models	Lecture Notes, CDA Chapter 4	Homework 4
Week 11 4/1-4/5	Logistic Regression	Lecture Notes, CDA Chapter 5-7	
Week 12 4/8-12	Logistic Regression		Homework 5
Week 13 4/16-19	Models for Multinomial Responses	Lecture Notes, CDA Chapter 8	
Week 14 4/22-/26	Models for Dependent Categorical Data	Lecture Notes, CDA Chapter 11-13	
Week 15 4/29-5/3	Models for Dependent Categorical Data		Homework 6
Week 16	Final Exam on 4:00-6:00 p.m., Monday, May 13, 2019)		

SPH AND UNIVERSITY POLICIES & RESOURCES

The School of Public Health maintains up-to-date information about resources available to students, as well as formal course policies, on our website at www.sph.umn.edu/student-policies/. Students are expected to read and understand all policy information available at this link and are encouraged to make use of the resources available.

The University of Minnesota has official policies, including but not limited to the following:

- Grade definitions
- Scholastic dishonesty
- Makeup work for legitimate absences
- Student conduct code
- Sexual harassment, sexual assault, stalking and relationship violence
- Equity, diversity, equal employment opportunity, and affirmative action
- Disability services
- Academic freedom and responsibility

Resources available for students include:

- Confidential mental health services
- Disability accommodations
- Housing and financial instability resources
- Technology help
- Academic support

EVALUATION & GRADING

Homework problems will be given approximately every week on Tuesday and due the following Tuesday or with an exact due date. **Late homework is counted down 20% for each day of lateness, with the first 20% accruing to homework handed in after class on the due date,** unless you have a documented reason for missing a homework assignment which has been pre-arranged with the instructor. You are encouraged to discuss the homework problems and to work together on the computing. *However, each student is expected to write his/her submission independently. Turning in the same written answers to an assignment or the same code will not be tolerated and will result in a zero on the assignment.* Students are highly encouraged to work through the assignments throughout the week (rather than waiting until near the due date) in order to receive feedback from the instructor and the TA. Many assignments will involve computing; hand in only relevant computer output. Note that submitted code should be commented, and lengthy bits of code should be placed at the end of the document.
There will be a midterm and a final exam. The final exam will be due during the University scheduled time-slot (**4:00-6:00 p.m., Monday, May 13, 2019**) and the midterm is tentatively scheduled for **March 12**. The final exam will be cumulative but will be weighted more heavily towards material from the latter portion of the course.

A student's final grade will be calculated by weighting the three components of the course (homework, midterm, and final exam) as follows:

- Homework – 30%
- Midterm Exam – 35%
- Final Exam – 35%

Please refer to the University's Uniform Grading Policy and Grading Rubric Resource at https://z.umn.edu/gradingpolicy]

Grading Scale
The University uses plus and minus grading on a 4.000 cumulative grade point scale in accordance with the following, and you can expect the grade lines to be drawn approximately as follows:

% In Class	Grade	GPA
93 - 100%	A	4.000
90 - 92%	A-	3.667
87 - 89%	B+	3.333
83 - 86%	B	3.000
80 - 82%	B-	2.667
77 - 79%	C+	2.333

73 - 76%	C	2.000
70 - 72%	C-	1.667
67 - 69%	D+	1.333
63 - 66%	D	1.000
< 62%	F	

- A = achievement that is outstanding relative to the level necessary to meet course requirements.
- B = achievement that is significantly above the level necessary to meet course requirements.
- C = achievement that meets the course requirements in every respect.
- D = achievement that is worthy of credit even though it fails to meet fully the course requirements.
- F = failure because work was either (1) completed but at a level of achievement that is not worthy of credit or (2) was not completed and there was no agreement between the instructor and the student that the student would be awarded an I (Incomplete).
- S = achievement that is satisfactory, which is equivalent to a C- or better
- N = achievement that is not satisfactory and signifies that the work was either 1) completed but at a level that is not worthy of credit, or 2) not completed and there was no agreement between the instructor and student that the student would receive an I (Incomplete).

Evaluation/Grading Policy	**Evaluation/Grading Policy Description**
Scholastic Dishonesty, Plagiarism, Cheating, etc.	You are expected to do your own academic work and cite sources as necessary. Failing to do so is scholastic dishonesty. Scholastic dishonesty means plagiarizing; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; altering, forging, or misusing a University academic record; or fabricating or falsifying data, research procedures, or data analysis (As defined in the Student Conduct Code). For additional information, please see https://z.umn.edu/dishonesty The Office for Student Conduct and Academic Integrity has compiled a useful list of Frequently Asked Questions pertaining to scholastic dishonesty: https://z.umn.edu/integrity. If you have additional questions, please clarify with your instructor. Your instructor can respond to your specific questions regarding what would constitute scholastic dishonesty in the context of a particular class-e.g., whether collaboration on assignments is permitted, requirements and methods for citing sources, if electronic aids are permitted or prohibited during an exam. Indiana University offers a clear description of plagiarism and an online quiz to check your understanding (http://z.umn.edu/iuplagiarism).
Late Assignments	**Late homework is counted down 20% for each day of lateness, with the first 20% accruing to homework handed in after class on the due date,** unless you have a documented reason for missing a homework assignment which has been pre-arranged with the instructor.
Attendance Requirements	Students are expected to attend weekly lectures, submit the homeworks and finish exams on time.
Extra Credit	Some homeworks contain bonus questions that students can choose to answer and earn extra credit.

CEPH COMPETENCIES

Competency	Learning Objectives	Assessment Strategies* (*see Assessment Descriptions below this table)
Analyze quantitative and qualitative data using appropriate biostatistical models and computer software.	**General outcomes** • Run statistical analysis using software, create ANOVA tables, contingency tables, interpret results, and write summary report **Confidence Intervals** • Calculate a confidence interval for a population parameter (e.g., mean(s), variance parameters) from data or summary statistics via their chosen software. **Hypothesis Testing** • Identify proper tests and statistics to conduct the hypothesis tests, and carry out the tests via their chosen software. **Regression** • Create exploratory plots to visualize the relations among outcomes and covariates. • Create diagnostic plots via their chosen software to assess how well the model fits the data. • Carry out statistical inference via their chosen software for the correlation or the fitted regression coefficients in simple ANOVA models, and generalized linear models.	• Homework • Exams
Interpret results of data analysis for public health research, policy or practice.	**Descriptive and Graphical Summaries** • Interpret summary statistics, tables, and graphs for each variable type (e.g., categorical variable(s): bar plot, count, proportion, 2x2 table, risk, odds, odds ratio, relative risk; continuous variable(s): histogram, boxplot, mean, median, SD, IQR, difference in means). • Interpret analysis results from generalized linear models (e.g., log odds ratios, explained variance, intraclass correlation coefficient, random-effects distribution)	• Homework • Exams