

PHC 6053 Regression Methods for the Health and Life Sciences

Spring 2008

Instructor

Yongsung Joo HPNP #3117 Phone: 273-5822 Email: yjoo@phhp.ufl.edu

Course Webpage

www.phhp.ufl.edu/~yjoo/PHC_regression

All students should read this webpage entirely every Tuesday.

Course Description

This course introduces graduate students in fields other than statistics to a wide range of modern regression methods. Emphasis is on modelling driven by actual data from studies in a variety of areas, primarily from health, biology, and ecology. The primary topics are multiple linear regression, logistic regression, and Poisson regression. A main goal is to learn to determine what approach to use among the linear and nonlinear models, and how to determine if the fit is adequate. By the end of the course, students will achieve competence in carrying out the analyses in standard statistical software, primarily the SAS language.

Course Objectives

1. Describe standard terminology and definitions for regression analysis.
2. Draw appropriate conclusions from regression-type data, for both randomized designed experiments and observational studies.
3. Test and interpret linear models for continuous outcome data (normal linear model).
4. Test and interpret models for categorical outcome data (logistic and Poisson regression).
5. Identify limits of parametric methods, by contrast with nonparametric and robust methods for regression analysis.
6. Select appropriate biostatistical methods for specific data, especially in determining whether a linear or a nonlinear approach to regression is appropriate.
7. Sharpen analytic skills and reasoning.
8. Achieve a high level of statistical judgment, with the ability to decide upon and complete a satisfactory path of analysis among many possible approaches.
9. Use statistical software for performing regression analysis, mainly the SAS language.
10. Communicate clearly to subject matter experts the purposes and results of complex statistical analysis, both orally and in writing.
11. Describe roles of biostatisticians in modelling health science data.

Prerequisite

STA6166 or an equivalent course. Some familiarity with statistical software (ex. SAS, Minitab, etc.).

Class hours

Mon 8:30-9:20am and Wed 8:30-10:25am at G307 HPNP bldg.

Office hours

Yongsung Joo: TBA

Required Textbook

Neter, Kutner, Nachtsheim, and Wasserman (2004): *Applied Linear Statistical Models, Fifth Edition*, McGraw-Hill.

Easy SAS books (probably not in campus store)

- Applied Statistics and the SAS Programming Language (5th Edition) by Ron P. Cody and Jeffrey K. Smith (Paperback - Mar 30, 2005)
- The Little SAS Book: A Primer, Third Edition by Lora D. Delwiche and Susan J. Slaughter (Paperback - Nov 2003)
- If you do a search in internet bookshops, you will find many different SAS books that might be worth ordering.

Statistical Software (required)

All students should have SAS in his/her Windows laptop. You will have exams and quizzes using your laptop. Please class webpage for detailed instruction and updated information. Mac is NOT suitable for this class. SAS stopped updating Mac version SAS about 15 years ago. Instructor will use SAS 8.0 or higher primarily. See <http://software.ufl.edu/sas/> for SAS program purchase and online documents. If you are registered as a student in an approved program at the University of Florida, you can get a CD of the SAS system and a one-year user's license from the bookstore (392-0194) of the Florida Bookstore for around \$35. You may have to collect this in person.

Grading

The course grade will be based on class participation (5%), homeworks + quizzes (25%), two midterm exams (20%+20%) and a term project (30%). Quizzes will be given whenever necessary. **Each exam will have two parts-SAS and written parts. You must use SAS in your own laptop.** In my opinion, the expectation on SAS skills is not high. But, SAS skills will be a very important part of your grade. **If you don't need or avoid learning basic computer programming, I strongly suggest taking other STA6166 section.**

To obtain the full points in exams, homework, quizzes and project, students are required to show how he/she gets the final answer. *Points will be given only for good reasoning.* You will get A with 90-100 out of 100, B+ with 85-89.99, B with 80-84.99, C with 75-79.99, etc.

Tentative schedule for exams, project and quizzes

- Midterm 1: evening exam on 3/17
- Midterm 2: evening exam on 4/21
- Project proposal: 3/24
- Poster session and project manuscript: 4/23
- Quiz: dates will be announced at least 2 days in advance.

Homework and quizzes

New homework will be announced in class and posted on the class web page. No late homework will be accepted. Everyone is expected to do every problem. Each homework and quiz will have the same weight in calculating your total grade. If your answer in homework is almost or exactly same as other student's, you can be asked to redo homework anytime after homework is turned in. When computer analyses are due, please submit a complete write-up of the assignment and documented

programming code used to analyze the dataset.

Project

This is a two-people team project. Students should collect their own data or use student's research project data. The proposal (outline description) for this project will be due about a month before the last day of class. The term project is to be a complete analysis of a dataset. Students are required to turn in a final report, which will be shared with the class on the last day as poster presentations. The final report should include text, tables, and figures in the main texts and documented SAS code in the appendix. On average, 10 page long main texts are expected. The grade for the project will depend on presentation in "manuscript", presentation during poster section, and the correctness and adequacy of the analysis. Students are encouraged to apply various statistical techniques to their own data set. In addition, each student will be asked to write brief critics on others' project during poster session.

Class announcement

Important class announcements will be announced in class. Some, but not all, of them will be posted on the web also. All students should read this webpage every Tuesday.

Class attendance

All students are required to attend all classes.

Tentative Course Schedule–Topics (Chapter) and useful exercise problems

- Week 1 Simple Linear Regression (1); Data management and simple linear regression (1); 1.6, 1.7, 1.12, 1.18, 1.29
- Week 2 Computation for simple linear regression (1); Inferences and Correlation (2); 1.30, 1.37, 2.3, 2.10, 2.11, 2.12, 2.19, 2.57, 2.54, 2.56
- Week 3 Diagnostics and remedial measures (3); 3.15, 3.16, 3.20, 3.24
- Week 4 Simultaneous Inference (4); Matrix algebra (5); 4.1, 4.21, 4.22, 1.27, 4.6, 4.10, 5.4
- Week 5 Multiple Regression I (6); 6.2, 6.3, 6.4, 6.22, 6.25
- Week 6 Multiple Regression II (7); Diagnostics I (10); 7.1, 7.23, 7.32
- Week 7 Special Regression Models (8); 8.3, 8.4, 8.5, 8.12, 8.17, 8.21, 8.27, 8.33 Week 8 Multiple regression and diagnostics II (10); 10.2, 10.3, 10.4; Both quantitative and qualitative predictors (8, 9); 9.3, 9.5
- Week 9 Remedial Measures for Multiple Regression (11); Variable selection procedures (9); 11.1, 11.2, 11.3, 11.4, 11.5
- Week 10 Non-Linear Regression (13); Remedial measures (11); 13.1, 13.2, 13.4, 11.25
- Week 11 Logistic Regression (14); 14.2, 14.3, 14.4, 14.47
- Week 12 Bivariate logistic regression models with quantitative and qualitative predictors (14)
- Week 13 Interaction effects; interpretation of Logistic Regression Models (14); Model selection in Logistic Regression (instructor's notes)
- Week 14 Poisson Regression (14); Logistic compared to Poisson regression (14); 14.12, 14.18, 14.24, 14.39
- Week 15 : Review and Student Presentations

Statement of University Honesty Policy (cheating and use of copyrighted materials)

Academic Integrity: Students are expected to act in accordance with the University of Florida policy on academic integrity (see Graduate Student Handbook for details). Cheating or plagiarism in any form is unacceptable and inexcusable behavior.

*We, the members of the University of Florida community,
pledge to hold ourselves and our peers to the highest standards of honesty and integrity.*

Policy Related to Class Attendance

Students are required to attend all classes.

Policy Related to Make-up Homework and Exams

Neither make-up homework nor make-up exam is allowed, unless there is a very reasonable justification, such as student's medical emergency or conference.