

Department of Statistics  
University of Wisconsin, Madison  
Fall Semester, 2009

Stat 849:  
Theory and Application of Regression and Analysis of  
Variance-I

|                           |   |
|---------------------------|---|
| <b>Class hours:</b>       | 8:25-9:40am MW  |
| <b>Classroom:</b>         | 133 SMI   |
| <b>Instructor:</b>        | Sündüz Keleş  |
| <b>Office:</b>            | 1245B MSC   |
| <b>Phone:</b>             | (608) 263 4533  |
| <b>Email:</b>             | keles@stat.wisc.edu.  |
| <b>Course website:</b>    | <a href="http://www.stat.wisc.edu/~st849-1">http://www.stat.wisc.edu/~st849-1</a> |
| <b>Office hours:</b>      | M: 4-4:40pm; R: 12:00-1:00pm. (subject to change)                                 |
| <b>TA:</b>                | TBA   |
| <b>TA office:</b>         | TBA   |
| <b>TA office hours:</b>   | TBA   |
| <b>Discussion:</b>        | 1:00-2:15pm (F); 8:25-9:40am (F).   |
| <b>Course email list:</b> | stat849-1-f09@lists.wisc.edu  |

**Description.** The aim of this course is to provide an introduction to theory and application of regression and analysis of variance. In particular, we will cover estimation and testing theory for least squares fits under general, Gauss-Markov, and related assumptions; diagnostic tools; prediction and model selection in linear regression; and generalized linear models. The lectures will mainly focus on the theory and provide examples of applications. The homework assignments will provide the opportunity to implement many of the methods on real data and case studies.

**Prerequisites.** Linear algebra, introductory statistics & probability.

**Reading List.** All of the four books below are on reserve at Wendt library. As our main text book, we will use

- *Linear regression analysis* by G.A.F. Seber and A.J. Lee (2nd edition).

For some of the topics that will be covered, you might also find the following book useful.

- *Linear models: least squares and alternatives* by C. R. Rao and H. Toutenburg.

In the last few weeks of the course, we will be using

- *Generalized linear models* by P. McCullagh and J.A. Nelder (2nd edition).

A good source on R/S is

- *Modern Applied Statistics with S* by W. N. Venables and B.D. Ripley (4th edition).

**Grading.** Grading will be composed of following components:

- **Homework assignments: 25%.** Bi-weekly homework assignments will include a lab component, where you will be typically required to analyze a data set, and a theory component. Homework assignments will be given on a Monday and will be collected on Friday after two weeks. The main computing environment will be R (see below). Homework assignments including a lab component should be typed in latex, word, or any other text processing language. Analytical parts can be handwritten. Submission of raw computer output is not allowed. You should incorporate your results in your text as relevant Tables and Figures with informative captions and provide the source code as appendix. We will provide a basic introduction to latex and a lab report template. There will be a total of 5-6 homework assignments.
- **Midterm exam I: 25%.** An in class midterm exam is scheduled on October 12, 2009. This is subject to change, in particular to after 6pm of the same day.
- **Midterm exam II: 25%.** An in class exam is tentatively scheduled on November 16, 2009. This is subject to change, in particular to after 6pm of the same day.
- **Final project: 25%.** The final project will include analyzing a data set of your choice (everyone should be analyzing a different data set) which will be subject to my approval. You are also welcome to review a paper of your choice (should be approved by me) or a paper from the list I will provide. You are expected to evaluate and criticize the paper and repeat some of the data analysis or simulations presented. Your final output will be a detailed report of the data analysis/paper review that includes all the analysis that you will perform. We will have a poster session at the end of the semester where you will have a chance to present posters of your projects.

In order to avoid leaving things to the last minute, everyone should submit a project proposal by October 28th, 2009. The proposal should include a brief description of the data you plan to analyze and your tentative plan of analysis.

You are not allowed to collaborate on the exams. You can share ideas on homework problem solutions but all the work presented on solutions must absolutely be yours. If you discuss your solutions with a classmate (or someone else), you should write that person's name at the end of your homework.

**Computing.** We will use R (<http://www.r-project.org>), a freely available implementation of S language, as our computing environment. Discussion sessions will include basic information on related R functions.

### Tentative syllabus

| Week | Date                                 | Topic   |
|------|--------------------------------------|---|
| 1    | Sep 2, W                             | A general introduction to the concept of regression and review on matrices and quadratic forms.   |
| 2    | Sep 7, M<br><u>Sep 9, W</u>          | LABOR DAY.<br>Linear models: definition and examples.<br>Least squares estimates in full-rank models.<br>Normal equations, calculus approach, and geometrical notes.  |
| 3    | Sep 14, M<br><br>Sep 16, W           | Properties of least squares estimates. A data example in R.<br>Review on random vectors and matrices.<br>Generalized inverse: least squares estimates in less-than full-rank models.<br>Estimation of the error term and $\sigma^2$ ,<br>Linear estimators. |
| 4    | <u>Sep 21, M</u><br><br>Sep 23, W    | Introduction to Gaussian linear model.<br>Gaussian linear model: Estimates of $\beta$ and $\sigma^2$ .<br>Gauss-Markov theorem: full rank and extended case.  |
| 5    | Sep 28, M<br><br>Sep 30, W           | Linear regression models with linear restrictions on parameters.<br>Testing linear hypothesis: t and F-tests, canonical form.<br>Linear regression diagnostics: diagnostic plots, goodness of fit, transformations, and analysis of variance.               |
| 6    | <u>Oct 5, M</u><br><u>Oct 7, W</u>   | Testing linear hypothesis: t and F-tests, canonical form.<br>Linear regression diagnostics.   |
| 7    | <u>Oct 12, M</u><br><u>Oct 14, W</u> | <b>Midterm I</b> (in class).<br>Linear regression diagnostics.  |
| 8    | <u>Oct 19, M</u><br><br>Oct 21, W    | Methods for dealing with multi collinearity:<br>PCR, ridge regression, PLS.<br>Methods for dealing with multi collinearity.   |
| 9    | Oct 26, M<br><br>Oct 28, W           | Prediction and model selection in linear regression:<br>Notions of loss functions and risk; best subset selection; stepwise methods; cross-validation.<br>Prediction and model selection in linear regression.<br><b>Project proposals are due.</b>         |
| 10   | <u>Nov 2, M</u><br><u>Nov 4, W</u>   | Prediction and model selection in linear regression.<br>Prediction and model selection in linear regression.  |
| 11   | Nov 9, M<br>Nov 11, W                | Generalized linear models: logistic, poisson, log-linear models.<br>Generalized linear models.  |
| 12   | <u>Nov 16, M</u><br><u>Nov 18, W</u> | <b>Midterm II</b> (in class).<br>Generalized linear models.   |
| 13   | Nov 23, M<br>Nov 25, W               | Generalized linear models.<br>Generalized linear models.  |
| 14   | Nov 30, M<br>Dec 2, W                | Generalized linear models.<br>Generalized linear models.  |
| 15   | Dec 7, M<br>Dec 9, W                 | Generalized linear models.<br>Generalized linear models.  |
| 16   | Dec 14, M                            | <b>Final project poster presentation and paper are due.</b><br>Time: 1-4:30pm. Location: 133 SMI.   |
| 17   | Dec 15, T                            | <b>Poster critiques are due by 10:30am.</b>   |

**PS1:** Underlined dates correspond to days on which the homework will be assigned. The homework is due a Friday after two weeks of the assigned date and should be returned to the TA in the discussion session.

**PS2:** No extensions will be granted on the homework assignments. In case of an emergency, contact me. The TA is not authorized to give extensions on the homework.