

Syllabus for:  
Stat 423: Applied regression and analysis of variance  
Winter quarter 2007

January 3, 2007

## 1 Outline and prerequisites

This is a course on multiple regression, aimed at mathematically trained undergraduates. Prerequisites for the course are Stat 342, Stat/Math 390, Stat 421, or Stat/Econ 481. Math 308 is recommended. (For a slightly higher level class, please see Stat 504.) The main goals for this class are:

- That you learn to run a multiple regression analysis, and to interpret the results correctly. We will use the statistical software R to get hands-on experience with this.
- That you learn to interpret and critique regression analyses done by others. This is important because regression analysis is one of the most widely used statistical methods. We will look at newspaper and academic articles for this purpose.

You should spend about 8 hours per week on this class, outside of the lecture and lab.

## 2 Contact information

Instructor:	TA:
Marloes Maathuis Padelford B-307 / B-302 Phone: 206-685 7379 Office hours: W, Th 5:00pm-6:00pm marloes@stat.washington.edu <a href="http://www.stat.washington.edu/marloes">http://www.stat.washington.edu/marloes</a>	Kapil Phadnis Padelford B-222 Phone: 206-616-9262 Office hours: T, Th 1:30pm-2:30pm kapilp@u.washington.edu <a href="http://students.washington.edu/kapilp">http://students.washington.edu/kapilp</a>

The class website is <http://www.stat.washington.edu/marloes/teaching/stat423>.

## 3 Meeting times and places

Lectures: MWF 2:30-3:20pm, More 226  
Computer lab: W 3:30-4:20pm, Mary Gates Hall 058

## 4 Text

Required text:

- John Fox (1997), “Applied Regression Analysis, Linear Models, and Related Methods”, Sage Publications. This book is on reserve in the math research library.

Additional good texts:

- Sanford Weisberg (2005), “Applied Linear Regression, 3rd edition, Wiley. (Contents similar to the book by Fox, but more mathematical and more succinct.)
- Paul D. Allison (1999), “Multiple linear regression, a primer”, Thousand Oaks. (Mostly on interpretation, non-mathematical.)
- Peter Dalgaard (2002), “Statistics and Computing”, Springer. (On software R.)
- Julian J. Faraway (2005), “Linear Models with R”, Chapman & Hall. (On using R for regression.)

## 5 Computing

We use the statistical software R. The software is free, and available at <http://cran.r-project.org/>. To work with R outside of lab times, you can install it at home or use the MSCC computer labs in CMU B022 and CMU B027 (by logging in as the “lab” user). For more information on the MSCC computer labs, see <http://www.ms.washington.edu/mscc/lab.html>.

## 6 Homework, exams and grading

### 6.1 Homework

**Homework philosophy:** Like many subjects, statistics is learned best when you work with the material. Therefore, there are substantial weekly homework assignments, consisting of data analysis problems and theoretical problems. Please don’t view the homeworks as a test, but as an opportunity to learn and to get feedback on how well you understand the material.

**Homework format:** Your homework can be either written or typed. Make sure to have the questions in the right order, to write your name on the first page, and to *staple* the pages together. At the end of your homework, you should include all R-code with output.

**Homework collaboration:** You are free to discuss the assignments with others to understand the concepts, but you *must write up your own solutions* and you *must write your own R-code*. If solutions and/or R-code of two or more students appear to be copies, then all will receive grade zero.

**Homework schedule:** The first homework is due on Monday January 7 by 2:30pm. After that, the assignments are due each Friday by 2:30pm. Late homeworks will not be accepted, and I will not give extensions except for very special circumstances.

## 6.2 Midterm exams

There will be two midterms in weeks 5 and 6 of the quarter: one is an in-class exam, and one is a take-home exam. The in-class exam is closed book. The take-home exam is similar to a homework assignment, but a bit larger. It must be done completely independently, meaning that you cannot communicate in any way with your classmates or anybody else other than me. You can make use of books and the internet, but you must cite your sources appropriately.

## 6.3 Group projects

We will do group projects during the last three weeks of class. There are about 3 students in a group, and I will form the groups. You can choose your own data set and research question. During the last week of class, you must submit a report about your project, and you will receive the report of another group. During finals week (March 12-16), each group attends two sessions. One time you will be presenting your project, and one time you will be critiquing the work of another group. Such a session takes about 30 minutes:

- Group A presents their project (5-10 min)
- Group B gives feedback/comments/critique (10 min)
- Group A responds, and there can be discussion

## 6.4 Grading

**Grade complaints:** Complaints about the grading of homeworks or exams are only considered if they are submitted *in writing and within 1 week*.

**Homework grade:** Your homework scores will be converted to percentages. Your lowest homework score will be dropped, and your homework grade is then determined by averaging the remaining scores.

**Group project grade:** The grade for the group project is based on your report (1/3), the presentation/defense of your project (1/3) and your comments/critique on the project of the other group (1/3).

**Final grade:** Your final grade will be determined as follows: homework: 25%, in-class midterm exam: 25%, take-home midterm exam: 25%, group project: 25%.

## 7 Academic integrity

Please see the section about collaboration on homework (above), stating the amount of collaboration that is allowed on the homeworks. Moreover, see the section about the midterm exams (above), stating that these exams must be done completely independently. Students suspected of cheating at the exams will be reported to the Vice President for Student Affairs, who will then investigate the case and determine appropriate sanctions.

## 8 Class website

The class website is: <http://www.stat.washington.edu/marloes/courses/stat423>. The site contains: (i) this syllabus, (ii) a schedule for the class, (iii) handouts and notes, (iv) R-code, (v) homeworks and exams, (vi) information about the group projects, (vii) Epost, (viii) Umail, (ix) relevant links.

Epost is a discussion board. I encourage you to use this board to discuss issues regarding the lecture, homework, programming, or any other questions you may have. You are also strongly encouraged to respond to questions of others, since explaining something is one of the best ways to learn. If you post a question, please use informative subject lines, such as: “Lecture Jan 3: ...”, “Lab Jan 3: ...”, “HW1, exercise 3” or “R: how to make a scatter plot?”. That makes it easy for others to search the messages. It is *not* allowed to post homework solutions, or to post R-code that is needed for the homework solutions. So if you have a question that requires you to give away most of the answer to a homework problem, then please don’t use the discussion board and email the TA or me instead.

Umail can be used to send anonymous email to the TA and/or me. We welcome your comments and feedback on the course. So please feel free to tell us in person, or to send (anonymous) email about it. Please note that we cannot respond to anonymous email.

## 9 Statistics Tutor & Study Center

The Statistics Tutor & Study Center is a comfortable place to study by yourself or with friends, and to get free help from statistics graduate students. For more information, see: <http://www.stat.washington.edu/www/tutorcenter/>. The center is mainly focused on introductory statistics classes. Therefore, it may happen that the tutors need to help students of those classes first, or that they don’t know the answer to some of your questions.

## 10 Students with disabilities

To request academic accommodations due to a disability, please contact Disabled Student Services: 448 Schmitz, 206-543-8924 (V/TTY). If you have a letter from DSS indicating that you have a disability which requires academic accommodations, please present the letter to me so we can discuss the accommodations you might need in the class.

## 11 Tentative schedule

Week	Topic	Readings
1	Introduction to the class and the software R Examining data and introduction to statistical models	Chap. 3 & 1
2	Introduction to regression and transforming data	Chap. 2 & 4
3	Linear least-squares regression	Chap. 5
4	Statistical inference for regression	Chap. 6
5	Dummy-variable regression	Chap. 7
6	Logistic regression	Chap. 15
7	Model selection	
8	Diagnostics I: unusual and influential data	Chap. 11
9	Diagnostics II: Nonlinearity	Chap. 12
10	Diagnostics III: Collinearity and variable selection	Chap. 13

## 12 Miscellaneous remarks

### Attendance:

Attendance in both lectures and lab is expected. If you miss a lecture or lab, then it is your responsibility to check in with classmates and the website, in order to catch up on any developments you missed.

### Participation in class:

Asking and responding to questions during lecture or lab is strongly encouraged. If something is unclear to you, then it is probably unclear to many of your classmates as well, so please speak up.

### Resources for help:

It is very important that you keep up with the course, because once you get behind, it is hard to catch up. So if anything is unclear to you, you should resolve this right away, by figuring it out yourself, discussing it with your classmates, coming to office hours, or coming to the Tutor & Study Center. You can also post your questions on the discussion board, or email the TA or me with questions/concerns. If you are struggling in the class, please let me know as soon as possible, and I will do my best to help out.