Computing in Statistics, 22S:166
Fall 2007

1 General Information

Instructor: Kate Cowles, 374 SH, 335-0727
kcowles@stat.uiowa.edu

Office hours: T 1:30 - 2:20 p.m.
W 12:30 - 1:20 p.m.
Th 1:30 - 2:20 p.m.
Please feel free to make appointments to see me outside of office hours, and to send me questions by e-mail.

Department: Statistics and Actuarial Science, 241 SH
DEO: Luke Tierney, 241 SH, 335-0712
luke-tierney@uiowa.edu

Lectures: M, W, F 30 SH 2:30 - 3:20
Lab: some Fridays (instead of lecture) 41 SH or 346 SH

Web page: http://www.stat.uiowa.edu/kcowles/s166_2007
Handouts, homework assignments, datasets, etc. will be posted on the web page for you to download.

Required readings: See “Web Resources” section of course web page

Mathematical Sciences Library, 125 MLH, Givens and Hoeting, Computational Statistics

2 Course goals and objectives

Through hands-on experience with real problems, students will learn computing skills essential in applied statistics and in research in methodological and theoretical statistics. Topics include the Linux operating system; R and SAS (statistical computing environments); LATEX (mathematical document preparation language); reproducible research; database management; WinBUGS (software for fitting Bayesian models); simulation methods (Monte Carlo studies, bootstrap, MCMC); statistical computing algorithms (Newton’s method, EM algorithm); interfacing to cyberinfrastructure resources.

3 Evaluation of students

3.1 Homework

In general, homework will be assigned each Fri. and will be due in class the following Fri. Exceptions to this schedule will be announced in class.

Show your work when solving written homework problems. For computer problems, turn in printouts of your commands or programs and their output.

You are encouraged to study with others. However, if you do work with others on homework assignments, please: a) write up your own assignment and make sure you completely understand all solutions that you submit, and b) write the names of the others in your study group on your assignment.

Late homework is accepted only as required by university policy, i.e. due to “illness, mandatory religious obligations, or other unavoidable circumstances or University activities.”

3.2 Exams

There will be two 2-hour hands-on computing midterm exams and one comprehensive final. The midterms will be scheduled in the computer lab outside of the regular class period and will replace the lab session for that week. The final will be given during the scheduled final-exam period. The computing exams are open book and open notes. Students may bring one 8-1/2 x 11 in. sheet of paper with notes to the final exam.

Midterm 1 week of 10/1, 346 SH (replaces lab on 10/5)
Midterm 2 week of 11/5, 41 SH (replaces lab on 11/2)
Final exam 7:30 A.M., Tues, Dec 18 2007

Missed exams may be made up only with documentation of reasons required by university policy (see “Late Homework” above).

3.3 Projects

Students will work in pairs to carry out projects involving application of the statistical computing methods covered in the course to problems of their own choosing. I will be happy to work with you at each stage of your project. Examples of possible types of projects are:

• Design and carry out a simulation study to compare the properties of two or more statistical procedures
• Code an MCMC sampler in R and use it to fit a model to a real dataset. Use WinBUGS to fit the same model to the same dataset. Use the R package BOA to analyze the output of both your sampler and WinBUGS. Verify that the results match.
• Learn to use two or more R packages that we have not studied as a class. Apply them to perform useful analyses of a real dataset.
• Choose a research question and find data to address it. Use SAS to “clean” and prepare the data and to carry out an analysis.
Projects will be carried out in three phases. Please meet with me at least once while you
are working on each phase.

- Project proposal (due 11/05) This is a detailed description of what you plan to do,
  including question(s) to be addressed, software to be used, methods to be applied.
- Project interim report (due 11/28) This informal report will indicate that your project
  is “on track.” All computing should be completed at this point. The interim report
  will include results obtained thus far and a brief summary (hand-written is O.K.)
  of what they mean and what remains to be done. In addition, each member of the
  project team will list which tasks they have performed for the group.
- Project presentation (must be posted or submitted by 12/10)
  The final form of the project must be prepared in L\TeX . This can be either
  - a paper to be posted on the course web page (send PDF file to me for posting)
  - slides to accompany an oral presentation to the class (use computer and projector
    in the classroom)

  Presentations will be given in class during the week of 12/10.

3.4 Grading

The course components will be weighted as follows:

- Homework 10%
- Midterms 35% (17.5% each)
- Project 20%
- Final 35%

4 Additional information

Academic Fraud All forms of plagiarism and any other activities that result in a student
presenting work that is not his or her own are academic fraud. All academic fraud is
reported to the departmental DEO and then to the Associate Dean for Academic Programs
and Services. See Academic Fraud at http://www.clas.uiowa.edu/students/academic_handbook/ix.shtml for the complete
policy. *

Making a Suggestion or a Complaint Students have the right to make suggestions
or complaints and should first visit with the instructor, then with the course supervisor if
necessary, and next with the departmental DEO. All complaints must be made as soon as
possible. For more information visit, Student Complaints at http://www.clas.uiowa.edu/students/academic_handbook/ix.shtml#5.

Accommodations for Disabilities Under the Americans with Disabilities Act and Sec-
tion 504 of the Rehabilitation Act of 1973, instructors provide reasonable academic accom-
modations for qualified students with disabilities. Students seeking academic accommoda-
tions first register with Student Disability Services and meet with a counselor in that office
who reviews documentation and determines eligibility for services. Students approved for
accommodations arrange to meet privately with course instructors. Visit Student Disability
Services at http://www.uiowa.edu/~sds/.

Understanding Sexual Harassment Sexual harassment is reprehensible and will not
be tolerated by the University. It subverts the mission of the University and threatens the
well-being of students, faculty, and staff. Visit this site (http://www.sexualharassment.uiowa.edu/)
for definitions, assistance, and the full University policy.

College administering this course This course is given by the College of Liberal Arts
and Sciences. This means that class policies on matters such as requirements, grading, and
sanctions for academic dishonesty are governed by the College of Liberal Arts and Sciences.
Students wishing to add or drop this course after the official deadline must receive the
approval of the Dean of the College of Liberal Arts and Sciences. Details of the University
policy of cross enrollments may be found at:

Resources for additional help The office of the Department of Statistics and Actuarial
Science maintains a list of private tutors.

5 Syllabus

This approximate schedule will be updated as needed during the semester.
08/27 - 08/31 Intro to Linux
09/05 - 09/07 Intro to L\TeX
(no class on 9/3; holiday)
09/10 - 09/14 R
09/17-09/21 R packages and reproducible research
09/24-09/28 Statistical algorithms:
The Bootstrap and Newton’s Method
10/01 - 10/05 The EM algorithm
Midterm 1
no class on 10/05
10/08 - 10/12 Simulation studies
Intro to MCMC
10/15 - 10/19 MCMC continued; WinBUGS
10/22 - 10/26 Database concepts and Microsoft Access
Reading data into SAS
lab in 41 SH
10/29 - 11/02 File handling in SAS; arrays
no class on 11/02
11/05 - 11/09 Formats; reports; proc tabulate, etc.
project proposals due 11/05
Midterm 2 (replaces lab on 11/02)
11/12 - 11/16 Data checking and validation
11/19 - 11/23 Thanksgiving break
11/26 - 11/30 SAS macro language
project interim reports due 11/28
12/03 - 12/07 Parallel and grid computing; review
12/10 - 12/14 Project presentations
Projects due 12/10
Final exam Tues. 12/18, 7:30 a.m.