

36-402/608: Advanced Data Analysis II

Course Policies and Syllabus

Spring 2010

Course Information

Location: A51 Baker Hall

Times: TR 10:30-11:20

Instructor:

Teaching Assistants:

Howard Seltman

Gary Klein

Jason Waddell

hseltman@stat.cmu.edu

gklein@stat.cmu.edu

jwaddell@stat.cmu.edu

232H Baker Hall

268-3938

Office Hours:

Howard: Mondays 3:00-4:00, Wednesdays 8:30-9:30 (BH232H)

Gary: Wednesdays 3:00-4:00 (FMSB 320)

Jason: Tuesdays 3:30-4:30 (FMSB 320)

Please feel free to drop in, or e-mail to arrange another time.

Web page:

[Blackboard](#) (Read announcements weekly!)

Backup page: <http://www.stat.cmu.edu/~hseltman/402/>

Physically disabled and learning disabled students

The Office of Equal Opportunity and Disability Services provides support services for both physically disabled and learning disabled students. For individualized academic adjustment based on a documented disability, see <http://www.cmu.edu/hr/eos/disability/students/index.html>.

Textbook

The textbook for this course is Fred L. Ramsey and Daniel W. Schafer (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis* Duxbury Press.

Its home page is <http://www.proaxis.com/~panorama/home.htm>. Data from the enclosed CD will be used for homework assignments.

Prerequisites

I assume that:

- You have (more or less) mastered the material in 36-401, especially the material about regression models and diagnostics.
- You have a working knowledge of R, and have it installed (at least version 2.9.0) on your computer (update or install from <http://cran.r-project.org/>).
- You have a strong desire to analyze data and a willingness to participate in class.

Course Description

Advanced Data Analysis (ADA) I and II are key courses in the Statistics undergraduate program.

In ADA I (36-401) you should have learned to apply the theory of linear regression analysis—including confidence intervals and hypothesis testing, basic distributions such as the normal, t , χ^2 and F distributions, basic facts and techniques for linear regression, including the material about regression models and diagnostics—in order to investigate model assumptions, and take appropriate action if the assumptions don't hold. In addition you should have learned something about using R, and about writing literate data analysis reports.

In ADA II (36-402), we will further develop those skills in exploring data, building and fitting models, investigating model assumptions, interpreting results from statistical models, and report writing, while learning a variety of more advanced models including PCA, logistic regression, Poisson regression, mixed models, and mediation analysis. We will also learn SAS.

Grading

Participation in class:	20%
Homework:	80%
<hr/> Total	<hr/> 100%

Each class will include “breakout” exercises done in small groups. Full class participation credit will be given to students who sign at least 80% of the breakout sheets *and* contribute periodically to class discussions.

There will be eleven homework assignments, worth ten points each, with the lowest three dropped. Homeworks will be due by the beginning of class on Thursdays. Homework turned in between Thursday 10:30 and Friday 10:30 will lose one point, homework turned in between Friday 10:30 and Monday 10:30 will lose 2 points. No credit is given for homework turned in after that time. Cutoffs for letter grades A, B, C, and D are 90, 80, 70, and 60%.

Academic Integrity

I encourage you to discuss the homework assignments with each other, but the work that you hand in must be your own. You must not copy mathematical derivations, computer output and input, or written descriptions from anyone or anywhere else, without reporting the source with your work. Please review the CMU policies on cheating and plagiarism.

Communications

- Homework, data sets, and R code for this class may be found at Blackboard or the backup web page.
- The TAs (gklein@stat.cmu.edu and jwadell@stat.cmu.edu) and the instructor (hseltman@stat.cmu.edu) may be contacted by email. Please feel free to send us email with questions, comments, etc., anytime. Also, please feel free to drop by Howard's office or schedule special appointments by email anytime.

Other Sources of Information

A list of useful statistics texts by topic is at <http://www.amstat.org/sections/cnsl/BooksJournals.cfm>.

One good place to search the statistics literature is the Current Index of Statistics <http://www.statindex.org/CIS/psqlQuery>.

More information about R can be found at my R web page: <http://www.stat.cmu.edu/~hseltman/RTips.html>.

Tentative Schedule

Week of	Tuesday	Thursday
Jan 11	Statistics review	R review
Jan 18	Simulation and Robustness	Alternatives: Permutation tests and classic non-parametrics
Jan 25	Outliers and Missing data	Power and Model Selection
Feb 1	ANOVA I	ANOVA II
Feb 8	Models with structural assumptions	Nested ANOVA
Feb 15	Serial correlation I	Serial correlation II
Feb 22	Repeated Measures I	RM II
Mar 2	Multivariate Responses I	Multivariate Responses II
Mar 9	Spring Break	Spring Break
Mar 16	SAS	Mixed Models in SAS
Mar 23	Binary Outcomes	Tables of Counts
Mar 30	Logistic Regression I	Logistic Regression II
Apr 6	Poisson Regression I	Poisson Regression II
Apr 13	Some Experimental Design Issues	Carnival
Apr 20	Coding a new test I	Coding a new test II
Apr 27	Mediation I	Mediation II