Mathematics 206 — Data Analysis

Syllabus and Course Procedures — Spring 2012

Instructor:  Brian D. Jones  
Office:  303 Hayes Hall  
Phone:  5386  
e-mail:  jonesbd@kenyon.edu

Office Hrs:  Monday 9:10 - 10:00  
            Tuesday 9:10 - 11:00  
            Wednesday 12:10 - 1:00  
            Friday 9:10-10:00

Required Texts


2. Peck and Devore, supplemental sections to the above text. These are available to the public at http://www.cengage.com/statistics/peck

3. Moore and McCabe, supplemental chapters to Introduction to the Practice of Statistics, 5th edition. These are available to the public at http://bcs.whfreeman.com/ips5e/

Course Material  The following is a rough topics list. Some topics may be omitted or added, depending on our time schedule, and we probably won’t have time to cover all topics. The case studies at the end will be covered on an if-time-permits basis, and they reflect a couple areas of data analysis that are growing rapidly — biology and forensics.

- Review of binomial, normal, t distributions
- Hypothesis Testing Overview, Errors and Power
- Review of one-sample tests for mean
- Normality Tests
- One sample test for median
- Review of the two-sample setting
- Ch 13 - Simple Linear Regression and Correlation
- Ch 14 - Multiple Regression Analysis
- Ch 15 - Analysis of Variance
- Supplement - Random Effects Models
- Ch 16 - Nonparametric Statistical Methods
- Supplement - Fisher’s Exact Test and Simulation
- Moore & McCabe Ch 14 - Bootstrap Methods and Permutation Tests
- Moore & McCabe Ch 16 - Logistic Regression
- Supplement - Loglinear Models
- Supplement - Mixed Models and Longitudinal Studies
- Moore & McCabe Ch 17 - Statistics for Quality: Control and Capability
- Case Studies in Microarray Analysis
- Case Studies in Density Estimation
Your Course Grade  

Your course grade will be determined as a weighted average as follows:

- Homework 18%
- Short Quizzes* 8%
- Analysis Concepts Gateway Exam 8%
- Consulting Project 1 9%
- Consulting Project 2 9%
- Midterm Exam 1 14% Friday March 2
- Midterm Exam 2 14% Friday April 20
- Final Exam 20% May 8, 8:30-11:30

Grading scale is approximately: A (90-100), B (80-90), C (70-80), D (60-70), F (below 60)

* I will delete your lowest quiz score before calculating your final course average.

Homework  

Homework problems will usually be assigned and collected weekly, sometimes twice per week. In addition to homework exercises from the text, in-class and out-of-class computer lab and SAS work will be assigned frequently. Homework solutions should be legible and presented in a logical fashion, with problem number clearly indicated. You do not need to type your work, but please write neatly. I may give no credit to messy homework that is difficult to read. Besides the usual symbolic language of mathematics and statistics, good homework solutions should be accompanied, where appropriate, by explanations and ideas written in complete English sentences.

Homework assignments are to be handed in at the beginning of the class period at which they are due unless I specify otherwise. No credit will be given to unexcused late papers. Do not leave your homework back at your dorm room! If you have a conflict with a due date because you are a student-athlete, tell me as far in advance as possible. If you have an excused illness, send me an e-mail as soon as possible.

One of my favorite axioms is Homework is where most of the learning happens in a course, and I agree whole-heartedly. The payback from hours of hard work on home assignments will be a deeper understanding of data analysis, high homework scores, and ultimately the likelihood of a high grade in the course.

Late Policy  

Homework assignments must be turned in to me at the beginning of the class period on the assigned due date, unless I specify otherwise. No credit will be given to unexcused late papers. If you have an illness or athletic schedule conflict, let me know as soon as possible.

Short Quizzes  

I plan to give approximately four to six quizzes, so their frequency will be roughly bi-weekly. The quizzes will be about 10 minutes in duration, and may cover concepts and interpretations as well as statistical calculations. The goal of the quizzes is to serve as a feedback to me as to where difficulties lie in student understanding, in addition to serving as a motivator for students to stay current with the material. I will delete your lowest quiz score before calculating your final course average.

Analysis Concepts Gateway Exam  

All of you are familiar with the usual exam and grade paradigm—if a student gets a 70% on a test, she earns approximately a C, a score of 85% will earn her a B, etc. But some core topics are too important for an understanding level below, say, 90%. For such “you must know these cold” topics, a gateway exam is often used. A gateway exam assesses whether or not a student has a near-perfect understanding of some vital concept. For this course, some definitions, concepts, and interpretations need to be known with uncompromising accuracy and clarity. Therefore, a gateway exam on these will be given about the forth week of the semester. More specific details concerning the gateway exam will be provided about a week before the exam, but the following is a list of likely test items.

- definitions of and contrasts between parameters and statistics
- null and alternative hypotheses
- type I and type II errors in context
- type I and type II error probabilities, significance level, and power
- hypothesis rejection rules and conclusions in context
- definition and interpretation of \( r^2 \) in linear regression models
- definition and interpretation of p-value in hypothesis testing
- interpretations of a confidence interval and the confidence level
- central limit theorem
Consulting Projects There will be two data analysis consulting projects in this course. For each project you will work in pairs. You and your partner will be given a data set, a written description of how and why the data were collected, and a list of analysis questions the owner of the data is interested in. Based on knowledge gained from this course, student teams will analyze the data, draw conclusions, and write up a formal report.

Your final formal report should be of professional quality — typed, having a cover sheet, introduction, body, and results and conclusions. It should be well-written with good use of both technical and non-technical language, and displaying appropriate graphics. The grade on your report will be weighted, giving 70% weight to mathematical and statistical correctness, and 30% to presentation. Each team member should contribute to all phases of the consulting projects, and each team member will receive the same grade on the project.

More detailed requirements will be described in a separate handout early in the semester.

Exams There will be two midterm exams, and one comprehensive final exam. The dates for these exams are:
- Midterm Exam 1 — Friday, March 2
- Midterm Exam 2 — Friday, April 20
- Final Exam — May 8, 8:30-11:30 am

For each midterm exam you may use one 8.5x11.0 help-sheet (both sides usable); and for the final exam you may use two such sheets.

Software There will be a considerable amount of work done with the aid of the software packages SAS and Minitab.

I am assuming that most of you have some familiarity with Minitab, since it is our analysis staple in Math 106. If you have never used Minitab, don’t panic! Minitab is so easy to learn and use that you can pick up the menu navigation basics during class activities. I would also be happy to show anyone Minitab basics in office hours.

The SAS analysis software is popular in industry due to its statistical power and its ability to store, manipulate, and merge very large data sets. Simply put, SAS is huge! In the 1980's the SAS manuals used to form a stack about 4 feet high! We will only scratch the surface of SAS’s capabilities. Having some SAS experience will be very attractive on your resume when you seek a job involving analysis.

Minitab, on the other hand, is limited in its analysis capabilities but is very easy to use for some of the more simple procedures. Minitab also is a fast and friendly program for running simulations, which are instrumental in demonstrating the underlying statistical principles of analysis.

Fast and powerful statistical software is a somewhat bittersweet pill. With today’s powerful software, we can perform a plethora of analyses quickly and generate a ream of output, whether or not these analysis procedures are appropriate for the given data setting, and whether or not key validating assumptions are met. Knowledge and discipline must match the awesome computational power and speed of modern software!

Academic Honesty Any work you submit for credit in this course must result directly from your own understanding, thoughts, and ideas. Presenting the work of others as your own is strictly prohibited.

Disabilities If you have any disability and therefore may have need for some type of accommodation(s) in order to participate fully in this class, please feel free to discuss your concerns in private with Erin Salva, Coordinator of Disability Services, by calling her office at phone number 5453, or by sending her an email at salvae@kenyon.edu.