

Data Analysis (Statistics 3022)

Zack W. Almquist
Fall Semester, 2013

Class Schedule

Lecture:	MWF	11:15–12:05	Phys 166
Lab 16410:	Tu	12:20–1:10	FordH 110
Lab 16411:	Tu	2:30–3:20	FordH 110
Lab 21412:	Tu	11:15–12:05	FordH 110

Class Website

URL: <http://moodle.umn.edu>

Note: Requires UMN login and registration in class to access.

Professor

Name: Zack W. Almquist
Office: 372 Ford Hall
Office Hours: F 10:00-11:00 AM
Email: almquist@umn.edu
Telephone: 612-624-4300

Teaching Assistant

Name: Yang Yang
Office: 313 Ford Hall
Office Hours: TU 1:10-2:10PM & 3:30-5:30PM
Email: yang3175@umn.edu

Course Objectives

This course is an introduction to modern statistical methods and software. Here, we will focus first on classic statistical hypothesis testing (e.g., t -test), and then continue on to

various model based methods of analysis (e.g., the linear and generalized linear models). All mathematical and statistical concepts covered in this course will be paralleled with the necessary tools to perform the analysis in the R statistical programming environment.

Prerequisites

STAT 3011, STAT 3021 or equivalent; students are assumed to have an elementary understanding of the basic concepts of probability and statistics. English language proficiency appropriate to an upper-division university class is assumed.

Course Requirements

Computers

It is not required that students bring their laptops to lecture and lab (if one is owned), but it is *highly* recommended since both lecture and lab will make extensive use of the computer software R. Computer labs are available on campus, please consult with the TA if you have trouble finding the various locations that computer labs reside on campus.

Readings

Weekly readings are assigned on the course syllabus. All readings are assumed to be completed before each lecture.

Homeworks

There will be weekly homework assignments. These homework assignments will be a combination of problem solving exercises, computer based exercises and comprehension exercises. The purpose of the homeworks are two fold: first to review the concepts covered in class and second to test mastery of these of concepts.

Unless announced otherwise, homeworks will be made available on Wednesday and due the next Wednesday at 5:00 through Moodle. Homeworks will be graded on combination of completeness (i.e., problems attempted) and correctness (i.e., problems completed correctly).

ONLY the version of homework reports turned in through Moodle will be accepted. No late homework will be accepted. The lowest homework grade will be dropped for the final grade calculation.

This class has zero-tolerance of scholastic dishonesty. Any suspicious homeworks will be submitted to the University for further evaluation.

Labs

Labs are intended to give students practice in R and can be completed in lab section. Labs assignments will be made available on Tuesday and are to be turned in through Moodle the following Tuesday by 5:00pm. Labs will be graded on completeness.

ONLY the version of lab reports turned into Moodle will be accepted. No late labs will be accepted. The lowest lab grade will be dropped for the final grade calculation.

Exams

To assess mastery of course material, three exams will be administered over the course of the semester. Two of these will be in-class midterm exams, and the other will be administered during the final exam period (see course outline). The examinations will cover all material presented in lecture, in addition to assigned readings. Students are advised to keep up-to-date on reading assignments, and to attend lab sections regularly, so as to be prepared for exams. Students must be present for each exam to obtain credit; exams are not rescheduled due to travel or other considerations. Note that dictionaries or other language aids may not be used when taking exams (see English proficiency, under Prerequisites).

Simple calculators such as the TI-30 (e.g., <http://www.radioshack.com/product/index.jsp?productId=3572861>) will be allowed, but *no graphing calculator or phones* will be allowed. No cheat sheet will be allowed, but all equations and R syntax necessary will be provided unless specifically mentioned in lecture (e.g., you will be expected to memorize standard equations like the sample mean and sample variance and certain basic R commands such as `pnorm` and `qnorm`).

Grading

Grades for this class are based on participation in the weekly labs, homeworks, and the three exams (two midterms and a final). There may be the possibility of extra credit in lecture (this is not guaranteed) and will not be announced beforehand. This total grade is determined as follows:

Homeworks:	20%
Lab:	5%
Midterm 1:	20%
Midterm 2:	20%
Final Exam:	35%

Lectures, readings, labs, and review sessions are provided for each student's benefit. It is the responsibility of the student to take advantage of these opportunities to acquire and demonstrate mastery of course material, so as to achieve his or her desired grade.

Letter grade assignment

A	93%+
A-	90-92.99%
B+	87-89.99%
B	83-86.99%
B-	80-82.99%
C+	77-79.99%
C	73-76.99%
C-	70-72.99%
D	60-69.99%
F	<59.99%

Required Texts

Fred Ramsey and Dan Schafer (2013), *The Statistical Sleuth* (3rd Edition). ISBN-10: 1-133-49067-0 — ISBN-13: 978-1-133-49067-8.

Required Software

We will be using the R statistical programming language. R can be downloaded at <http://www.r-project.org/>.

Recommended Software

RStudio IDE (Integrated Development Environment) is a software application which facilitates interaction with the R statistical programming language. It is often preferred to the GUI (Graphic User Interface) made available through CRAN. You can download it at <http://www.rstudio.com/>.

Course Policies

Missed Exam

Exams can be made up for legitimate (documented) absences, such as varied illness with a letter assigned by a physician, jury duty, military service, and religious observances. If you

must miss the exam for legitimate reasons, you have to CONTACT THE INSTRUCTOR AT LEAST ONE WEEK BEFORE THE TIME OF THE EXAM. In that case, makeup exams may be arranged to be taken any time before the exam is returned to the class. If you have a legitimate reason, but fail to take a makeup exam, an incomplete may be granted. If you miss any exam without legitimate reason, you will receive a zero for that exam. Note that social/vacation plans are not legitimate reasons for missing an exam.

Incompletes

An incomplete will only be given if: The student has a documented case of extreme hardship. The student has, up until the point of the request, been completing coursework and exams. The student's average at the point of the request is a 70% or above. If these conditions are met, the student must request the incomplete from the instructor and it is still within the instructor's rights to refuse the request. The student who is granted an incomplete must take the initiative to finish the course or the grade will revert to an F.

Academic Integrity

From the OSCAI Website: Student Academic Integrity and Scholastic Dishonesty Academic integrity is essential to a positive teaching and learning environment. All students enrolled in University courses are expected to complete coursework responsibilities with fairness and honesty. Failure to do so by seeking unfair advantage over others or misrepresenting someone else's work as your own, can result in disciplinary action. The University Student Conduct Code defines scholastic dishonesty as follows:

Scholastic Dishonesty: Scholastic dishonesty means plagiarizing; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; altering forging , or misusing a University academic record; or fabricating or falsifying data, research procedures, or data analysis.

Within this course, a student responsible for scholastic dishonesty can be assigned a penalty up to and including an "F" or "N" for the course. If you have any questions regarding the expectations for a specific assignment or exam, ask.

In addition, I will file a claim with OSCAI if I have evidence of cheating. I understand that people often end up in difficult situations beyond their control, but these situations are no excuse for scholastic dishonesty. If you find yourself in a difficult situation, please come talk to me about options for the course. I will keep all conversations confidential.

Disability Accommodations

Every attempt will be made to accommodate any disabilities. Please contact Disability Services at either ds@umn.edu or 612-626-1333 and submit any paperwork at the start of the semester so that accommodations can be made.

Course Outline: Readings and Exams

- Week 1 : Univariate statistics and hypothesis tests: Z-test, t-test and CI
• Reading: 1-2.2.4
- Week 2 : Two-sample t-test, p-value and Type 1 and Type 2 error
• Reading: 2.3-3.4 and 3.5.1-3.5.2
- Week 3 : Comparisons among several samples (ANOVA)
• Reading: 5.1-5.3, 5.5, 5.6.1; 6.1-6.2
- Week 4 : Comparisons among several samples continued
• Reading: 6.3-6.5.2; 6.6
• **Midterm #1**
- Week 5 : Linear regression
• Reading: 7.1-7.4.3, 7.5-7.6; 8.1-8.5.2, 8.6.1, 8.6.2, 8.6.4; 9.1-9.2, 9.5
- Week 6 : Linear regression: Inference
• Reading: 9.3-9.4; 10.1-10.3, 10.4.1, 10.5
- Week 7 : Linear regression: Model checking and variable selection
• Reading: 11.1-11.4
- Week 8 : Linear regression: Model checking and variable selection continued
• Reading: 12.1-12.4
- Week 9 : Analysis of variance for two-way classification
• Reading: 13.1-13.4.3, 13.5.1
- Week 10 : Introduction to time-series: Autocorrelation
• Reading: 15.1-4, 15.5
• **Midterm # 2**
- Week 11 : Odds ratio
• Reading: 18.1-18.3
- Week 12 : Introduction to count data
• Reading: 18.4; 19.1-19.3
- Week 13 : Logistic regression
• Reading: 19.6; 20.1-20.3
- Week 14 : Logistic regression continued
• Reading: 20.4-20.6; 21.1-21.4
- Week 15 : Review
- Week 16 : **Final Exam**

Calendar

MONDAY	TUESDAY	WEDNESDAY	FRIDAY
Sep 2nd 1	3rd LAB 2	4th 3	6th 4
9th 5	10th LAB 6	11th homework 1 due 7	13th 8
16th 9	17th LAB lab 1 due 10	18th homework 2 due 11	20th 12
23rd 13	24th LAB lab 2 due 14	25th homework 3 due 15	27th 16 Midterm # 1
30th 17	Oct 1st LAB 18	2nd 19	4th 20
7th 21	8th LAB lab 3 due 22	9th homework 4 due 23	11th 24
14th 25	15th LAB lab 4 due 26	16th homework 5 due 27	18th 28
21st 29	22nd LAB lab 5 due 30	23rd homework 6 due 31	25th 32
28th 33	29th LAB lab 6 due 34	30th homework 7 due 35	Nov 1st 36
4th 37	5th LAB lab 7 due 38	6th homework 8 due 39	8th 40 Midterm # 2
11th 41	12th LAB 42	13th 43	15th 44
18th 45	19th LAB lab 8 due 46	20th homework 9 due 47	22nd 48

MONDAY	TUESDAY	WEDNESDAY	FRIDAY
25th 49	26th 50 LAB lab 9 due	27th 51 homework 10 due	29th Thanksgiving
Dec 2nd 52	3rd 53 LAB lab 10 due	4th 54 homework 11 due	6th 55
9th 56	10th 57 LAB lab 11 due	11th 58 homework 12 due	13th
16th	17th Final Exam: 10:30 a.m.–12:30 p.m.	18th	20th