Introduction to Statistical Reasoning

36-201   Spring, 1999

COURSE POLICIES AND SYLLABUS

```
MW, 9:30--10:20, Doherty 2315
Computer Labs Thu and Fri
```

Course Information:

Instructor:

Brian Junker, Statistics
232C Baker Hall
268-8873
brian@stat.cmu.edu

Office Hours:

Tuesdays, 2:00–4:00pm
or by appointment.

Course Web Site:

http://www.stat.cmu.edu/~brian/201/

TA's:

Head TA: Lou Mariano, ltm@stat.cmu.edu
Others TBA; See Course Web Site.

Office Hours:

TBA; See Course Web Site.

Computer Labs: *All sections meet in 239 Baker Hall.*

- Lab Section (A) meets from 09:30 to 10:20 on R
- Lab Section (B) meets from 14:30 to 15:20 on R
- Lab Section (C) meets from 09:30 to 10:20 on F
- Lab Section (D) meets from 12:30 to 13:20 on F
- Lab Section (E) meets from 14:30 to 15:20 on F

Know your lab section, and attend it each week. Lab sheets and homework are collected and returned in labs; it is your responsibility to be in the lab and to make sure the TA's in the lab have your completed lab work and homework.

Required Texts:

Moore, D. S. (1996). Statistics: concepts and controversies. 4th or current edition. NY: WH Freeman. ISBN: 0-7167-2863-X (paperback)

Siegel, A. F. and Morgan, C. J. (1996). Statistics and data analysis: and introduction. 2nd or current edition. NY: Wiley. ISBN 0-471-29332-6 (paperback)

Recommended Text:

Minitab, Inc. (1995). Minitab Mini-manual: A beginner's guide to Minitab statistical software. State College PA: Author. ISBN 0-925636-31-2

OVERVIEW

Does eating fat cause cancer? Does vitamin C prevent colds? Was Babe Ruth a more consistent hitter than Roger Maris? Can you predict someone's University grades from their SAT scores? Is it possible to win money in a Casino?

Statistics can help us answer all these questions. Broadly speaking, Statistics is a set of methods that help us collect, organize and process data. We use these tools to uncover relationships between variables, to make predictions and to discover causal mechanisms.

The aim of this course is to help you become Statistically literate. You will learn the concepts of Statistics. You will be equipped to read a newspaper and know when the quantitative information being presented is accurate or misleading. You will also know something about organizing and displaying numerical and graphical summaries of data, and drawing conclusions from these summaries.

You'll see very few equations. There will be few arithmetic calculations. Instead, we will emphasize graphical analysis of data and simple but powerful concepts that do not require difficult mathematical manipulations.

The main concepts we will talk about in this class are:

Exploratory Data Analysis: A data set consists of a collection of measurements. How can we organize them and describe them? How can we gain insight by exploring the data visually?

Experimental Design and Causation: A newspaper article reports that people who eat lots of oat bran have fewer heart attacks. Does this mean that eating oat bran prevents heart attacks? How can we be sure, when we observe that two phenomena are associated, that one "causes" the other?

Sampling Variability: If we measure a fixed quantity—like the length of a tabletop, the speed of light, or the fraction of people who support health care reform—repeated measurements will vary somewhat. Why is this? How much variability can we expect? Does the variability mean we can never tell what the right answer is?

Statistical Inference: It is rarely feasible to measure everyone in the population. At best, we can draw a random sample from the population. How can we make inferences about the population based on the sample? Pollsters do this everyday. How do they do it?

WHO SHOULD TAKE THIS COURSE

36-201: Statistical Reasoning is a freshman-level, General Education Requirement for all undergraduates in the college of Humanities and Social Sciences (H&SS). Bachelor of Science students also take *36-202: Introduction to Statistical Methods* form a two-semester introduction to Statistics for Bachelor of Science students in H&SS. Some non-H&SS programs, such as BFA, HCI, Architecture, use 36-201 to meet a Statistics requirement for their students as well.

You should not take 36-201 if you have already taken any other Statistics course at CMU, if you have taken AP Statistics in High School and passed the Advanced Placement (AP) test in Statistics with a score of 4 or 5, or if you have had some calculus and you qualify—because of a minor or additional major—to get into 36-220 (Engineering) or 36-247 (Biology), or 36-207/70-207 (Business Administration).

LABS

Each week, you will attend a Lab where you get hands on experience analyzing data. Attendance is mandatory. The computer labs for this course meet on Thursday and on Friday in Baker Hall 239. Check your section letter for the time your section meets. In the Labs we will use a computer program called *Minitab*. **There are no Labs during weeks in which there is a test.**

TESTS and FINAL EXAM

There will be two in-class tests during the semester. These are scheduled during class on

- **Wednesday, February 17** and

- **Wednesday, March 31**.

The tests will cover material discussed in lecture and computer lab as well as readings in the texts.

There will be a **final examination** for the course. The registrar schedules the final examination for this course later in the semester. Once scheduled the final exam cannot be changed. **If you make travel plans before you know when the exam is scheduled you do so at your own risk.**

All three tests will be closed book except that you may use one 8 1/2" by 11" sheet of paper with whatever formulas, facts or explanations you find helpful. Sharing of calculators, notes, etc. during examinations will not be allowed.

HOMEWORK

Homework will be assigned weekly. You may discuss an assignment with other students but the written solutions to homework problems must be your own and not copied from someone else. **Homework is due Thursday or Friday at the beginning of your lab section.** Usually, the homework assignment will be posted on the web-site. **It is your responsibility to get the homework assignment from the web site,** and turn in completed homework at the beginning of your lab section each week.

The written description and interpretation of data analyses is at least as important as generation of data summaries, statistics, tests, etc. Grading of both homework and exams will take into consideration written interpretations as well as numerical accuracy. Each homework assignment will be worth 100 points.

THE WEB PAGE

There is a Web page for this course. On this page you will find: (i) homework assignments, (ii) solutions to homework, (iii) announcements. Many of these things will **not** be distributed in class. It is your responsibility to check the web page on a regular basis. You can reach the web page at:

```
http://www.stat.cmu.edu/~brian/201/
```

Often, there will be three versions of each file on the web. They will look something like this:

```
homework1.txt
homework1.pdf
homework1.ps
```

You only need to download one of the three files. Files that end with "txt" are text files that can be read on any computer. Files ending with "ps" are postscript files and can be read on most campus Unix machines and PC's. Files ending with "pdf" can be read with any Mac on campus. Sometimes, only one (.txt) or two (.txt and .pdf) versions of a file available. If you have any trouble reading one of the files, the User Consultants in the clusters can show you how to read it.

CHEATING

Cheating and/or plagiarism will not be tolerated. Please see the *CMU Student Handbook,* p. 7-8, for definitions of cheating and plagiarism, and the severe consequences of such behaviors.

COURSE GRADE

The lowest two homework grades will be dropped. The remaining homework grades will be used to compute the homework average. One lab grade will also be dropped. Your final course grade will be a numerical average of your course work as follows:

- The homework average will contribute 15% to the course grade.
- Each midterm will contribute 22.5% towards the final grade.
- The weekly computer exercises will contribute 15%.
- The final examination will contribute 25% towards the final grade.

Letter grades of *at least* A, B, C, etc. are guaranteed to anyone who gets 90%, 80%, 70%, etc. of the maximum possible final course grade.

SUPPLEMENTAL COURSE MATERIAL: VIDEO TAPES

There is a series of video tapes, called *Against All Odds: Inside Statistics*, made for Public Television that was created to complement the material in the textbook. These tapes are available for your viewing in the basement of Hunt Library. Although watching these tapes is not required they are an excellent resource and course supplement.

HELP FROM TA's

The TA's will hold office hours each week. The place and times will be announced later and posted on the web site.

"WHAT IF I MISS A TEST?" and Other Important Questions

Missing a Test

No makeup examinations will be given. A student who misses an examination because of a medical reason must provide **documented** evidence of **medical incapacitation** to Professor Junker within seven days of the test. Other reasons for missing an examination must be discussed with Professor Junker **before** the day of the examination. Each case will be considered on an individual basis. The overall course grade for a student who misses an examination with a valid reason will be based on that student's remaining course work. A student who misses an examination without a valid excuse will receive a grade of zero on that examination.

Late Homework

Homework is due Thursday or Friday at the beginning of your Lab section. Because of the large class size, a fairly rigid late homework policy is needed to keep everyone sane. Late homework will be handled as follows, **no matter what the reason is**:

- Homework turned in to a TA during lab later than 10 minutes past the beginning of lab will lose 50% of the points earned. Homework turned in to a TA outside of lab will earn zero points.

- Homework turned in to my office (slide it under my door if needed) will be timestamped by me when I see it. If I see it less than 4 hours after the beginning of your lab section, we will still record 50% of the points earned; if I see it less than eight hours after the beginning of your lab, we will record 25% of the points earned. Later hw will earn zero points.

- If you miss a homework due to **medical incapacitation**, you will not be penalized as long as you bring documented evidence of your medical problem within seven days.

- If you do miss a homework, you should still get the assignment and do it, so you don't fall behind in the class.

Attending the Right Lab

You must attend the lab you are registered for. You will receive no credit if you attend any other Lab. If you want to switch into another Lab, you must ask the receptionist in the Statistics department (132 Baker Hall).

Missed Labs

There are no make-up Labs. You are allowed to miss one Lab without penalty. If you miss a Lab due to **medical incapacitation**, you will not be penalized as long as you bring documented evidence of your medical problem within seven days. If you miss a Lab for any other reason, you will receive zero for that Lab. If you do miss a Lab for any reason, you should get the Lab from the web-site and do it in your own time so you do not fall behind.

APPROXIMATE COURSE OUTLINE, SPRING 1999

Following is an approximate course outline for the semester. Since this is the first time these books have been used in this course, I may adjust this a bit week-to-week, so please pay attention to in-class announcements on what to read and study each week.

"M:"   =   Moore: *Statistics: Concepts and Controversies*
"S&M:"   =   Siegel and Morgan: *Statistics and Data Analysis*

	Dates	Topic	The Basics	The Details	Lab
1	Jan 11–15	Describing data	M: 4.1–4.4	S&M: Ch 1, 2, 3	Intro to Minitab
2	*Jan 18–22*[*]	Comparing Distributions; The Normal Distribution	M: 4.5	S&M: Ch 4	Comparing Distributions
3	Jan 25–29	Transformations, Outliers, Robustness	S&M: Ch 5, 6	—	Interpreting Standard Deviations, Transformations
4	Feb 1–5	Relating two continuous (quantitative) variables	M: 5.2, 5.4	S&M: 15.1–15.4	Scatter Plots and Regression
5	Feb 8–12	Relating two categorical (qualitative) variables	M: 5.1	S&M: 14.3, 14.4	Time of Death in Vietnam
6	*Feb 15–19*	Association, Causation, Confounding	M: 5.1, 5.3	—	**WEDNESDAY IN-CLASS EXAM**[*]
7	Feb 22–26	Unbiased Samples and Experiments	M: 1.1, 1.2, Ch 2	M: Ch 3	Treatment of Depression
8	*Mar 1–5*[*]	Surveys and sampling variability	M: Ch 1	S&M: Ch 9	Sampling Variability and Surveys
9	Mar 8–12	Probability	M: 7.1, 7.2, 7.4; S&M: 8.1, 8.2	S&M 7.1–7.4; 8.7	Probability
10	Mar 15–19	The Central Role of the Normal Distribution	S&M: 8.3–8.6	—	Normal Approximation and the Central Limit Theorem
	Mar 22–26	**SPRING BREAK**			
11	*Mar 29–Apr 2*	Confidence Intervals I	M: 8.1, 8.2	S&M: 10.2, 10.3, 10.5, 10.6	**WEDNESDAY IN-CLASS EXAM**[*]
12	Apr 5–9	Confidence Intervals II; Hypothesis Tests	S&M: 10.1, 10.7	M: 10.3–10.5	Confidence Intervals and Hypothesis Tests
13	Apr 12–16	The Consumer Price Index	M: 5.1–5.3	—	LABS DO NOT MEET - SPRING CARNIVAL
14	Apr 19–23	Time Series	M: 5.4	S&M: 15.1, 15.2	Time Series
15	Apr 26–30	"Catch up" and Review for Final Exam	—	—	TBA

[*]*Notes:*
  Week 2:   Class meets as usual on Martin Luthor King Day.
  Week 6:   Exam Wednesday; No Lab This Week.
  Week 8:   No Class Monday: Mid-Semester Break.
  Week 11:   Exam Wednesday; No Lab This Week.