Math 207, Probability and Statistics
Spring Semester, 2001
Dr. Evelyn Bailey

Office hours: Room 115A Seney Hall,
In general, 1:30 - 3:30 daily, and others by appointment

Text: Introduction to Mathematical Statistics, 5th ed, Paul G. Hoel

Reader: Statistics You Can't Trust, Steve Campbell

Materials: Math 207 Notes (provided in notebook), calculator (TI-83, TI-82, or equivalent type)

Content: Visual displays of data, measures of central tendency and of variability, classification of data, counting, probability, Bayes Theorem, probability functions, Chebyshev's Theorem, discrete distributions (binomial, hypergeometric, Poisson, uniform), continuous distributions (Exponential, Normal, Uniform), Central Limit Theorem, Confidence Intervals, Hypothesis Testing (for means and for proportions), Linear Regression (simple), one-way ANOVA, nonparametric tests (median, multinomial, Wilcoxon Rank Sum Test, Kruskal-Wallis Test).

Goals: At the end of this course students should be able to: categorize data, work various probability problems, understand the role of functions in statistics, describe major misuses of statistics, recognize several standard distributions, analyze interval data for which statistical tests involving difference of means and difference of proportions is needed, check for inherent assumptions of the statistical models that are included in this course, interpret relationships in bivariate data, compare distributions of responses, understand the role of statistics in analyzing data and in inference.

Grading: Grades will be determined by student performance on four different problem sets, three experiments, and a final exam. Each problem set will have 150 possible points; each experiment will have 60 points; the final exam will have 220 points, for a total of 1000 possible points:

4 problem sets @ 150	600
3 experiments @ 60	180
1 final exam	220

Total	1000 points

There will be opportunities for extra credit work, announced in class or listed on the attached assignments. These are usually not the standard type problems but those that require more thought and some original work. **You must do work independently for credit on bonus work unless otherwise indicated.** Points earned on extra credit will be added to your total points that will determine your grade.

In general,

900 points and up	A, A-
750 to 899 points	B+, B, B-
600 to 749 points	C+, C, C-
below 600 points	F

Some Policies: All problem sets will be given out at least one week before the due date. All problem sets are due at class time on the dates indicated. Fifty points will be deducted per day for late problem sets. Emergencies will be handled on an individual basis.

For work on problem sets, you may use your own notes (those you have taken in class), the notebook provided for this class, your textbook, the computer facilities, and/or your own computer or calculator; however, you may not receive help from another person or talk to anyone about the problems on the problem sets.

Policies regarding experiments will be explained in class, when the experiment is assigned. Experiments may require group work or individual work.

Homework problems (assignments attached) may be worked with other members of this class. Solutions to some homework problems are on reserve at the library.

You will need to keep your class notes and homework problems well-organized and complete so that they will be useful to you on your problem sets and final exam.

Class attendance is important. You are responsible for work done in class. There are no tutors for this course.

There is a Math 207 class conference on LearnLink. Please use the conference to check announcements and to communicate concerns and questions appropriate for the class.

THE HONOR CODE OF OXFORD COLLEGE APPLIES TO ALL WORK IN THIS CLASS! YOUR PROBLEM SETS AND ANY OTHER WORK ARE PLEDGED TO BE YOUR WORK IN ACCORDANCE WITH INSTRUCTIONS GIVEN FOR THE ASSIGNMENT.

Important dates:

Friday, February 9	Problem Set I due
Wednesday, February 19	Experiment 1 due
Monday, March 5	Experiment 2 due
March 12-16	SPRING BREAK
Wednesday, March 21	Problem Set II due
Monday, April 9	Problem Set III due
Friday, April 13	No Class

Monday, April 23	Experiment 3 due
Friday, April 27	Problem Set IV due
Monday, April 30	Last class day
Friday, May 4	Final exam, 2-5:00

Math 207
Homework Assignments

Wednesday, January 17	Introduction
Friday, January 19	

Read Chapter 1 and part of Chapter 2 (pages 4-26)

Explain how our present model of the solar system and our present model of the atom are examples of the evolutionary nature of model building. Are there other examples?

Read Introduction and Chapter 1 in *Statistics You Can't Trust*. See attached questions.

PROBABILITY

Monday, January 22	Definitions, Addition Rule, Multiplication Rule
Wednesday, January 24	(conditional), Bayes Theorem

p. 45-46: 1, 2, 3, 5-27

Definitions: experiment, sample space, certain event, impossible event, mutually exclusive, independent

Friday, January 26	Jerome Cardan, Birthday Problem

Problems on probability in the notebook, pgs. 4,5

How many people are needed so that the probability of at least two people having the same birthday is 1/2?

What is the minimum number of red and black socks in a drawer such that the probability of picking two red socks is 1/2? is 2/3? is 3/4? Is there a solution for all three?

Monday, January 29	Craps, Review

Finish previous homework

Optional Bonus 20 points, due with Problem Set I: Design a game similar to craps where $.490 \leq P(\text{win}) \leq .493$ that uses two different Platonic solids. (1) Name your game. (2) Clearly show the number assignment to the sides of the solids you select, and

the sample space for your choice. (3) Clearly provide your calculations demonstrating the probability of winning your game.

You may not discuss this bonus assignment with anyone. If you decide to work this optional problem, sign a statement that you have done this work by yourself, using only your notes, text, and calculator.

Read Chapter 9 in Statistics You Can't Trust. Answer the attached questions.

COUNTING

Wedensday, January 31 — Fundamental Theorem of Counting, permutations, combinations

P. 48: 28-37, 40 (Get common denominator)

How big is "30!"? How big is Avogadro's number? How many drops of water are there in all the oceans of the world? How many grains of sand are there on all the beaches of the world? How do you answer these questions?

Receive Problem Set I

Friday, February 2
Monday, February 5 — Probability and Counting

Problems on Counting in the notebook, pgs. 8, 9, 10
P. 50: 71-80, 82

FUNCTIONS (THEORETICAL)

Wednesday, February 7 — Random Variable, Density Functions

Read in Chapter 2 pages 27 to 45
page 48: 41-45
Explain the difference between a pdf and a cdf for both continuous and discrete functions.

Friday, February 9 — Joint Density Functions, Marginal Distributions

Problem Set I Due

page 49: 47-52, 58, 60, 62, 64, 68; page 52: 91, 92, 93, 97

Optional Bonus 20 points, due with Problem Set 2: Find an algorithm for determining the number of zeros at the end of N! where N is any whole number. Clearly give the steps in your algorithm. Show how your algorithm works by giving the number of zeros at the end of 200! and at the end of 1000! as examples.

Notebook problems p. 17, 18

Know the following definitions: random variable, joint probability distribution, marginal distribution, conditional distribution, independence.

Monday, February 12 Review

Experiment 1 assigned

DISCRETE FUNCTIONS

Wednesday, February 14 Discrete Functions, Binomial Distribution

Read pages 53-72
page 92: 1, 2, 5, 6, 7, 9, 12, 13, 15, 16, 17, 18, 20

Friday, February 16 Poisson, Hypergeometric, and Uniform Distributions

page 94: 22, 24, 25, 29, 33, 34
problems in notebook, pgs. 23, 24
Know the following definitions: expectation, moments (about the origin and about the mean), mean, variance, skewness, kurtosis.

CONTINUOUS FUNCTIONS

Monday, February 19 Continuous Functions, Chebyshev's Theorem, Normal
Wednesday, February 21 Distribution, empirical rule

page 95: 37, 42, 44
page 96: 46-50, 54, 55, 57, 58, 59, 62, 64
page 100: 101, 103, 104, 105, 106, 108
Experiment 1 due

Friday, February 23 Normal to approximate the binomial, Central Limit Theorem

page 97: 66, 67, 68
page 99: 83, 85, 87, 94, 95
Experiment 2 assigned

Monday, February 26 Rectangular (Uniform) distribution, Review

Finish previous homework
Problems in notebook, p. 29, 30
Answer questions about the experiment

DATA AND MEASUREMENTS

Wednesday, February 28 Descriptive Statistics
Friday, March 2

Read pages 102-109
page 117: 1-5
problems in notebook, p. 35, 36
Read Chapters 2, 3, and 10 in Statistics You Can't Trust. Answer attached questions.

Monday, March 5 Data measures and classification

Read pages 129 to 134; p. 162: 10, 11
Know the following definitions: random sample, stem-and-leaf, outlier, statistical inference, histogram, quartiles, parametric vs nonparametric statistics, types of data (nominal, interval, ordinal, ratio), measures of central tendency and measures of variability.
Read Chapters 4 and 5 in Statistics You Can't Trust. Answer the attached questions.
Experiment 2 due
Receive Problem Set II

CONFIDENCE INTERVALS

Wednesday, March 7 Confidence Intervals for means and for proportions
Friday, March 9

p. 164: 35, 36; p. 168: 72; problems in notebook, p. 40
Know the following terms: inferential statistics, point estimate, maximum error of estimate
Finish any previous homework, catch up!

Spring Break March 12-16

HYPOTHESIS TESTING

Monday, March 19 Theoretical hypothesis testing
Wednesday, March 21

p. 118: 13, 14, 16, 18, 19, 27, 28, 30
problems in the notebook, p. 44
Problem Set II due on Wednesday

Friday, March 23 — Hypothesis testing - means [to a value (large sample, small sample)]

Read chapters 6, 7, 8 in Statistics You Can't Trust. Answer the attached questions.
Know the following terms: null hypothesis, alternate hypothesis, test statistic

Monday, March 26
Wednesday, March 28 — Means continued [compare means (dependent, large sample, small sample, homogeneity of variance)]

Read pages 138 to 159; p. 162: 12, 15, 16, 25, 26, 27
problems in the notebook, p. 46, 47

Friday, March 30 — Hypothesis Testing - proportions

p. 163: 28, 29, 30, 32, 33; problems in the notebook, p. 48, 49
Receive Problem Set III

Monday, April 2 — Review hypothesis testing

p. 164: 35, 36, 41, 42, 43, 45, 46, 47, 48, 49, 53, 55, 56
p. 169: 80, 81, 83

REGRESSION

Wednesday, April 4
Friday, April 6 — Correlation and Simple Linear Regression

Read chapter 7
page 211: 2, 3, 7, 9, 13, 19
Work problems in the notebook, p. 53
Know the following definitions: bivariate data, coefficient of determination, covariance, method of least squares, spuriously correlated
Bring Calculator

Monday, April 9
Wednesday, April 11 — ANOVA

Problem Set III due Monday
Work problems in the notebook, p. 55, 56
Read Chapter 11 in Statistics You Can't Trust. Answer the attached questions.
Experiment 3 assigned Wednesday....you must be in class

Friday, April 13 — NO CLASS

NONPARAMETRIC STATISTICS

Monday, April 16 — Chi Square Tables, Multinomial Experiments, median test
Wednesday, April 18
Friday, April 20

Read chapter 9
page 266: 1, 2, 3, 4, 5, 15, 18
Read chapter 12 in Statistics You Can't Trust. Answer the attached questions.
Work problems in the notebook, p. 59, 60
Receive Problem Set IV on Wednesday
Optional Bonus experiment, 20 points, to be described in class on Friday and due with Problem Set IV

Monday, April 23 — Wilcoxon Rank-Sum, Kruskal-Wallis
Wednesday, April 25

Work problems in notebook, p. 61
Experiment 3 due Monday

Friday, April 27 — Discuss Major Misuses of Statistics

Read Chapters 13 and 14 in Statistics You Can't Trust. Answer attached questions.
Problem Set IV due

Monday, April 30 — Review, Evaluate

notebook p. 62
Read Glossary in Statistics You Can't Trust
Receive Take Home part of final exam

Friday, May 4

Final Exam at 2:00

Statistics You Can't Trust by Steve Campbell

Answer the following questions in a clear and succinct manner.

Chapter 1

1. Find one example of a misconception from the newspaper. Explain how it is a misconception.

Chapter 2

1. What are the two meanings of the word, statistics?
2. Why is an operational definition important?
3. Find one example of a meaningless statistic from a magazine or newspaper. Clearly explain your example.
4. Find one example of a hyperaccuracy from a magazine or newspaper. Clearly explain your example.
5. Define a self-selected sample. Give an example.
6. Define an unknowable statistic. Give an example.

Chapter 3

1. What are dishonest charts? Find an example from a magazine or newspaper.
2. Is it possible, in your opinion, to have a visually accurate pictogram? Why or why not?

Chapter 4

1. What is a pseudoaverage?
2. Under what circumstances would the average (arithmetic mean) not be a good measure of central tendency? (The median is robust to what?)

Chapter 5

1. Why is the variation important in a set of data?

Chapter 6

1. Give the definitions of the following: percent, percent change, and percent points of change. How can each be misleading?
2. What is the opportunistic construction of a percent?
3. Explain why broad-base fallacy is a statistical trick. Make up an example to illustrate this "unscrupulous" behavior.

Statistics You Can't Trust by David Campbell

Chapter 7

1. Would a clear operational definition eliminate problems with "an untidy comparison" or are there other considerations?
2. Describe the *risk-your-life-and-live-longer-fallacy.* Find an example in a newspaper or magazine.
3. Why is a control group necessary to arrive at conclusions, especially in health related experiments?

Chapter 8

1. Give an example why human inclination is to make order or over explain.
2. For the three claims on pages 150-152, offer a Plausible Alternative Conclusion.

Chapter 9

1. Who was Chevalier de Méré?
2. Why were Pascal and Férmat credited with the beginnings of probability?
3. Explain why "before the fact" and "after the fact" gives us different prospectives on determining probability. Relate these approaches to coincidences.
4. What is the classical approach?

Chapter 10

1. Why should one be suspicious of sample selection?
2. What is the difference in induction and deduction? Relate to statistical inference.
3. What is the margin of error and why is it important to report this margin when giving statistical information?

Statistics You Can't Trust by Steve Campbell

Chapter 11

1. Why is it not appropriate to discuss cause and effect in relation to a regression? What is the difference between correlation and causation?
2. Find an example of *post hoc ergo propter hoc* fallacy (or *post hoc* fallacy) in a magazine or newspaper.
3. Explain the following terms: necessary cause, sufficient cause, necessary and sufficient cause, contributory cause.
4. What is a lurking variable and why should one not discount it?
5. What are some problems with the use of a small sample size?

Chapter 12

1. What does it mean to be "trapped in cell a"?

Chapter 13

1. Why is it easy to misinterpret conditional probabilities?
2. Explain the regression fallacy proposed by Sir Francis Galton.
3. Why should one be cautious of: (a) projections? (b) computer results? (c) omission of details? (d) ill-conceived ratios?

Chapter 14

1. Evaluate the four questions in bold print on page 241. Do these four questions include all the types of misinformation?
2. Pick your favorite three examples in this chapter. Give the number of your example and explain why it was selected by you.