Math 207, Probability and Statistics
Spring Semester, 2004
Dr. Evelyn Bailey

Office hours: Room 303 Seney Hall,
In general, 10 - 11:30 TTh; 2:00 – 3:30W, 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 a green notebook), calculator (TI-83, TI-83 PLUS, 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, for one and two samples), goodness – of fit, one-way ANOVA, simple Linear regression and correlation, 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
1 experiment @ 80	80
1 Class Experiment @ 100	100
1 final exam	220

Total	1000 points

There will be opportunities for extra credit work. 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 your individual experiment will be explained in class, on the day the experiment is assigned. The class experiment will be in the form of a survey, will be designed by the class, and will emphasize: What students think academic advisers should do and what improvements at Oxford College should have high priority.

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. Some class days and parts of class days will be used to work on the class experiment.

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

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:

Wednesday, February 11	Problem Set I due
Wednesday, March 3	Problem Set II due
March 8-12	SPRING BREAK
Friday, April 9	No Class
Monday, April 12	Problem Set III due
Wednesday, April 21	Class Experiment due
Friday, April 23	Problem Set IV due
Monday, April 26	Last class day; Individual Experiment due
	Final exam

Math 207
Homework Assignments

Wednesday, January 14 Introduction
Friday, January 16

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? Why are models important?

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

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

PROBABILITY

Wednesday, January 21 Definitions, Addition Rule, Multiplication Rule
Friday, January 23 (conditional), Bayes Theorem

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

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

Monday, January 26 Jerome Cardan, Birthday Problem, Craps, Discuss Class Experiment
Wednesday, January 28

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?

Optional Bonus (20 points) due with Problem Set I explained in class.

Read Chapter 9 in Statistics You Can't Trust.

1. Who was Chevalier de Mere?
2. Why were Pascal and Fermat 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?

COUNTING

Friday, January 30 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? Which one (30! Avogadro's number, number of drops of water in the oceans, number of grains of sand) is the largest? How do you answer these questions?

Monday, February 2 Probability and Counting/Discuss Class Experiment (and any time during class as needed from now on…)

Wednesday, February 4

Problems on Counting in the notebook, pgs. 8, 9, 10
P. 50: 71-80, 82
Optional Bonus (20 points) due with Problem Set 2 explained in class.
Receive Problem Set I

FUNCTIONS (THEORETICAL)

Friday, February 6 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 continuous and discrete functions.

Monday, February 9 Joint Density Functions, Marginal Distributions

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 100! and 400!.
Notebook problems p. 17, 18
Know the following definitions: random variable, joint probability distribution, marginal distribution, conditional distribution, independence.

DISCRETE FUNCTIONS

Wednesday, February 11 Discrete Functions, Binomial Distribution

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

Friday, February 13 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 16 — Continuous Functions, Chebyshev's Theorem, Normal
Wednesday, February 18 — 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

Friday, February 20 — Normal to approximate the binomial as a model, Central Limit Theorem

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

Monday, February 23 — Rectangular (Uniform) distribution, Review

Finish previous homework
Problems in notebook, p. 29, 30
Why is there no year 0?
Receive Problem Set II

DATA AND MEASUREMENTS

Wednesday, February 25 — Descriptive Statistics
Friday, February 27

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.
Chapter 2

1. What are the two meanings of the word, statistics?
2. Why is an operational definition important?
3. Give one example of a meaningless statistic.
4. Give one example of a hyperaccuracy.
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 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?

Monday, March 1 Data measures and classification
Wednesday, March 3

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.

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?

Problem Set II due

CONFIDENCE INTERVALS

Friday, March 5 Confidence Intervals for means and for proportions

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 8-12

HYPOTHESIS TESTING

Monday, March 15 Theoretical hypothesis testing
Wednesday, March 17

p. 118: 13, 14, 16, 18, 19, 27, 28, 30
problems in the notebook, p. 44

Friday, March 19 Hypothesis testing - means [to a value (large sample, small sample)]

Read chapters 6, 7, 8 in Statistics You Can't Trust.

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.

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.

Know the following terms: null hypothesis, alternate hypothesis, test statistic

Monday, March 22 — Means continued [compare means (dependent, large
Wednesday, March 24 — 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 26 — Hypothesis Testing - proportions
Monday, March 29

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

Wednesday, March 31 — Review hypothesis testing

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

<u>REGRESSION</u>

Friday, April 2 — Correlation and Simple Linear Regression
Monday, April 5

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

Wednesday, April 7 — ANOVA/Status of Class Experiment Evaluated
Friday, April 9 – No Class
Monday, April 12

Problem Set III due Monday
Work problems in the notebook, p. 55, 56
Read Chapter 11 in <u>Statistics You Can't Trust</u>.

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 hwy should one not discount it?
5. What are some problems with the use of a small sample size?

NONPARAMETRIC STATISTICS

Wednesday, April 14 Chi Square Tables, Multinomial Experiments, median
Friday, April 16 test/Finish Class Experiment

Individual Experiment assigned Wednesday
Read chapter 9
page 266: 1, 2, 3, 4, 5, 15, 18
Read chapter 12 in Statistics You Can't Trust.
1. What does it mean to be "trapped in a cell a"?

Work problems in the notebook, p. 59, 60
Receive Problem Set IV on Friday

Monday, April 19 Wilcoxon Rank-Sum, Kruskal-Wallis
Wednesday, April 21

Work problems in notebook, p. 61
Class Experiment due - Wednesday

Friday, April 23 Discuss Major Misuses of Statistics/Discuss Final Exam

Read Chapters 13 and 14 in Statistics You Can't Trust.
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.

Problem Set IV due

Monday, April 26 Review, Evaluate

notebook p. 62
Read Glossary in Statistics You Can't Trust
Receive Take Home part of final exam
Individual Experiment due
Final Exam according to schedule