36-220: Engineering Statistics and Quality Control  
Spring 1998

Lectures: Monday (some) W,F  
Hamerschlag Hall B103  
9:30 – 10:20

Labs: (other Mondays)  
A: 9:30 – 10:20 BH 239  
B: 11:30 – 12:20 BH239  
C: 2:30 – 3:20 Hunt N and F

Course Policies and Syllabus

Vital Information

Instructor:  
Joseph B. Kadane  
Statistics Dept.  
Baker Hall 232E  
268-8726  
kadane@stat.cmu.edu

Deputy Instructor:  
Steve Duckworth  
Architecture Dept.  
Margaret Morrison 4126  
268-3211  
sld@andrew.cmu.edu

Joint Office Hour:  
11:00 – 12:00 a.m. Tuesday in BH232E  
or by appointment.

Consulting Engineer:  
Mitchell Small  
Environmental and Civil Engineering  
Engineering and Public Policy Depts.  
Porter Hall 123K  
268-8782  
ms35@andrew.cmu.edu

Consulting Cognitive Psychologist:  
Marsha Lovett  
Center for Innovation in Learning  
Baker Hall 345A  
268-3499  
lovett@cmu.edu

Head Teaching Assistant:  
Terra McKinnish  
Statistics Dept. & Heinz School  
Baker Hall 132B  
268-8591  
terram@stat.cmu.edu

Office Hours: (TBA)
Teaching Assistants:
Ilaria DiMatteo
Statistics Dept.
Baker Hall 232D
268-1884
dimatteo@stat.cmu.edu

Jiayin Xiang
Statistics Dept.
Baker Hall 132M
268-2724
jxiang@stat.cmu.edu

Office Hours: (TBA)

Graders:
Christine Hou
xch@andrew.cmu.edu

Alan Charness
charness@andrew.cmu.edu

Office Hours: (TBA)

Required Text

Recommended Text
MINITAB, Inc. A Beginner’s Guide to MINITAB Statistical Software

Prerequisites

- A solid understanding of first year calculus. This means that you must be able to integrate and differentiate standard functions (polynomials, exponentials, logarithms and trig functions), and you must understand the basic applications of calculus (finding areas, maximizing/minimizing functions, etc.).
- Sufficient familiarity with Andrew to write small programs in MINITAB. The recommended text on MINITAB will be very helpful; additional information will be covered in class or handouts.

Statistics and Quality Improvement

The environment in which American industry is operating has dramatically changed over the last decade, and further dramatic changes are likely. Foreign competition has intensified, and many products invented in America are no longer manufactured in America (e.g. consumer electronics). The causes for this are very complex, but one aspect is the significant improvements in product quality of foreign competitors’ products compared with comparable American products (e.g. American versus Japanese automobiles 10 years ago). Many American companies have embraced quality improvement programs (often called Total Quality Management or TQM). Moreover, some
of these companies are expecting the engineers they hire to have training in quality improvement methods and ideas.

Statistics is an intrinsically important discipline in quality improvement, because improving quality means gathering and analyzing data to assess the state of a process, experimenting with methods to improve the process and assessing the benefits of alternative process designs. This course will introduce you to the principles of statistical data analysis and modeling, statistical process control, and quality improvement. Upon successful completion of this course, you will have skills which will be useful not only for the rest of your studies at Carnegie Mellon but also in your career as a professional engineer or architect.

This course is “applied” rather than “theoretical.” We will follow the organization of the Schaeffer and McClave text, cutting and adding material from the chapters as needed. We’ll begin with simple descriptive summaries of data, which are quickly generated on the computer (Chapter 1). Next we will go over the basic rules of probability (Chapter 2), and examine a few of the most commonly used probability models (Chapters 3 and 4). In statistics there is a constant tension between variability (randomness) and regularity (law of large numbers; central limit theorem) and we will explore this tension (Chapters 5 and 6). This leads to the construction of control charts, which form a simple but powerful technique for keeping track of a process. After that we’ll move on to more formal methods for estimation and testing hypotheses (Chapters 7 and 8). Many problems in engineering and process control depend on examining the relationship between one variable and another, and we will discuss linear regression, a common statistical technique for doing this (Chapters 9 and 10). Finally (Chapter 11) we will go through a more detailed discussion of experimental design, and see how careful design of experiments can improve our ability to draw conclusions from data.

Beyond all the details you should gain an appreciation for the importance of considering the variability among measurements; if you do, this most basic concept in statistics and quality control should stay with you for many years to come. Understanding variability, and its consequences, motivates all of the specific skills you must master by the end of the term.

Course Objective

1. To develop a critical approach to the evaluation of numerical data.

2. To introduce the basic ideas and methods of probability theory and to illustrate the rich diversity of applications of probability models.

3. To introduce the basic principles and methods that underlie the practice of statistics and empirical research methods with a special emphasis on methods for investigating relationships among variables, i.e., regression analysis and quality control.

4. To implement approaches to the statistical analysis of data through the use of the statistical software package, MINITAB.
Course Materials

The main course material is in the textbook and the lectures. The lectures amplify, explain, exemplify and critique the text. Therefore, you will get the most from the lectures if you read the text before hearing the lectures, and perhaps reread the text afterwards. The labs give you the opportunity to practice what you learned from the textbook and lectures in a supervised way; with help immediately available. The homework allows unsupervised practice. The quizzes and exams help you and me evaluate your progress.

Generally new topics are introduced first in lecture (on Wednesday and Friday), and practiced in lab on Monday. Homework on that topic will be due Thursday by 5 p.m. at my office, and homework solutions will be handed out in class on Friday and will be on the web.

COURSE POLICIES

Student Evaluations

There will be unannounced quizzes at random times during the lectures. These will check your current understanding of the material.

I intend to give two midterm examinations, the first on or about February 11 and the second on or about March 18. Exams and quizzes will cover material discussed in lecture and computer lab as well as readings in the texts. No makeup examinations or quizzes will be given. A student who misses an examination or quiz because of a medical reason must provide documented evidence of medical incapacitation to me. Other reasons for missing an examination or quiz must be discussed with me as soon as possible before the day of the examination or quiz. Each case will be considered on an individual basis. The overall course grade for a student who misses an examination or quiz with a valid reason will be based on that student’s remaining course work. A student who misses an examination or quiz without a valid excuse will receive a grade of zero on that examination.

Each examination and quiz will be closed-book-and-closed notes, except that you may use one side of an 8 1/2" by 11" sheet of paper with whatever formulas, facts or explanations you find helpful. Sharing of calculators and formula sheets during examinations and quizzes will not be allowed. This means that it is to your advantage to come to class with such a sheet, and to keep it current.

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 can not be changed. If you make travel plans before you know when the exam is scheduled you do so at your own risk.

There will be a final lab exercise the last week of classes that will be graded. More details about this exercise will be provided during the semester.

Homework will be assigned weekly. The purpose of these assignments is to help you learn the course material. It is okay to discuss an assignment with other students or with others. When
you do so, you must report, with your homework, whom you talked with and the content of the discussion. If you so disclose and do so accurately, you are not liable under the plagiarism rules (although your grade for that homework may be reduced if your lack of participation in solving the homework problems so warrants). Conversations with me and with the TA’s do not have to be disclosed. You must disclose your conversation whether you were the giver or the receiver (or some of both) of advice or help. The written solutions to homework problems must be your own and not copied from anyone else. *Late homework will be corrected but not graded, and will be recorded with a grade of zero, absent a valid, written medical excuse.* Solutions will be available soon after you turn your assignment in.

The homework assignments will contribute positively toward your final course grade; however, they are more valuable because they give you a chance to practice applying the concepts, and reinforce the material covered in class. How well or poorly you do on exams is usually directly related to how seriously you take the homework assignments all through the term.

Statistical modeling and data analysis are skills best learned by trial and error, and by discussion. *I encourage you to discuss homework problems with each other, provided you disclose the conversation.* However, your solutions and writeups to problems must be your own. *Do not copy derivations, batch files, output files or writeups from any other person or source without disclosure,* since I will treat this as cheating.

You should use the homework as an early warning system about how well you understand the material. If you have trouble with the homework, it is far better to know about that early and use it as motivation for asking questions to aid your understanding. The students who do well in this course are those that insist on clarifying what they do not understand. The students who do poorly are those that let things slip (sometimes by relying too much on help from others in doing the homework). This is why it is very important that you write up your homework yourself, and why the rules about plagiarism apply to homework.

The homework will consist of mathematical problems, data analysis problems using MINITAB, and reading and writing assignments, since all these are skills a good data analyst should have.

All homework should be done in complete English sentences, with sensible math where necessary. For data analysis problems, the written interpretation and conclusions from a data analysis are at least as important as generation of data summaries, statistics, tests, etc.

Solution sheets will be handed out on the day after a homework assignment is due to provide immediate feedback. Consequently, *late homework papers will not be accepted* (except for excused illness). You should turn in assignments even if they are incomplete.

- The computer labs for this course meet on Mondays. During these class sessions you will work through an exercise using MINITAB which will help advance your understanding of the course material. **Attendance and participation in the computer labs is mandatory.**

▷ **Attendance in Class**

It is easy to take the attitude that your job (and mine) is accomplished with your mastery of the material of the course, and consequently that I need not bother with whether you show up for
class. Realistically, I know that in general the vast majority of students who feel they don’t need to come to class are mistaken, but only find that out, to their shock, as they do poorly on exams. One part of me suggests that I give you this information, and then let you find out for yourself.

But there is another aspect to this question. A great many resources have been assembled to give us the opportunity to have this classtime together. You can divide the tuition by the number of courses you are taking, for a start. Add in your room and board, and the opportunity cost of the money you could have made had you not been studying, as well. In addition, there are all of the University’s endowed resources that are being used-for buildings, libraries, equipment, salaries, etc. The costs are enormous, and imply responsibilities to those paying the bills. Consequently, I will expect you in class and lab (and monitor your attendance with unannounced quizzes) unless you present to me, within ten days of the start of class, a letter from your parents (or whoever is principally paying the bills) stating that they don’t care whether or not you come to class.

Cheating

At the heart of any cheating or plagiarism is a lie – to yourself and to me – about how well you understand the course material. A student who is disturbed because he or she doesn’t understand the course material is someone I can teach. A student who pretends to know, but doesn’t really know, is much harder for me to engage in the learning process. The earlier you see that you don’t understand something, the earlier you can get answers and start to build your knowledge. Leaving it until the last minute does not work with the material in this course.

For this reason, I care more about teaching and reinforcing honesty in you than I do about teaching you statistics. I urge you to read the pamphlet “Promoting Academic Integrity”. If you are in any doubt about what is and is not permitted, please ask.

The formal rules and procedures are in the CMU Student Handbook. At the end of the syllabus is an example of a letter you do not want to receive. It becomes part of your record as long as you are an undergraduate at CMU. If there are two such incidents (in any courses while you are here), a hearing about whether you should be suspended will ensue.

Questions

Asking questions in class is encouraged. My experience is that sometimes students are reluctant to ask questions, for fear of holding up the class. Whether to deal with a question in class, or suggest that it be pursued outside of class, is the instructor’s responsibility, not the student’s. Often if one student is confused about something and wants to ask about it, others are too, and will benefit from the explanation.

Additionally, you are welcome to use email and the class bboard on the web (see below) to ask questions, as well as office hours.

Course Grade

The lowest homework grade will be dropped except if it is the last assignment of the semester, which is mandatory. The remaining homework grades will be used to compute the homework
average. The lowest quiz grade will be dropped; the remaining quiz grades will be used to compute the quiz grade. Each midterm will contribute 15% towards the final grade and the final exam will contribute 20%. The computer labs will contribute 15%. The final lab exercise will contribute 10% towards the final grade. The quizzes and homework together constitute the remaining 25% of your grade. Your score for this component will be either 20% quiz, 5% homework or 15% quiz, 10% homework, whichever gives you the higher average. Midterm grades will reflect your work to that point.

▷ Bboard, Data Directory, Old Assignments

- There is a WWW bboard for this class at the URL:

  http://www.stat.cmu.edu/~kadane/220/wwwboard/

Announcements and discussions of problems and other issues on the bboard are part of the class; please read it regularly. If you want to be sure that I see a question you post quickly, send a copy to kadane@stat.cmu.edu also. If you have a private question or comment, or you are unsure whether your response to a bboard post is “giving the answer away”, send your response to me alone and I will post your answer to the bboard only if it is appropriate.

- Postscript and PDF copies of class handouts, assigned homeworks and solutions can also be found in the Web at the URL http://www.stat.cmu.edu/~kadane/220/

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 of the course. These tapes are available for your viewing in Instructional Technology in Building D. The programs corresponding to the material in the course are described below. Although watching these tapes is not required they are an excellent resource and course supplement. Please use them.


PROGRAM 2: Picturing Distributions  Presenting and interpreting the distribution of a single variable. Techniques taught include stemplots, frequency tables and histograms.

PROGRAM 3: Describing Distributions: Numerical Description of Distributions  Numerical measures of specific aspects of a distribution: center (mean, median), spread (percentiles, the five-number summary, boxplots, and the standard deviation). Resistance and its lack.

PROGRAM 4: Normal Distributions  Topics: density curves as smoothed histograms: mean, median, percentiles for density curves; the normal distributions (general shape, locating the mean and standard deviation, the 68-95-99.7 rule).
PROGRAM 5: Normal Calculations Standardization and calculation of normal relative frequencies from tables; assessing normality by normal quantile plots.

PROGRAM 6: Time Series From the distribution of a single variable we move to an examination of change over time. Topics: statistical control, inspecting time series for trend, seasonal variation cycles; smoothing by averaging, either over many units per time or over time by running medians.

PROGRAM 7: Models For Growth Mathematical models for the overall pattern of simple kinds of growth over time. Topics: linear growth, with review of the geometry of straight lines and an introduction to the least squares idea; exponential growth, and straightening an exponential growth curve by logarithms; prediction and extrapolation.

PROGRAM 8: Describing Relationships Topics: scatterplots and their variations, smoothing scatterplots of response vs. explanatory variable by median trace; linear relationships, least squares regression lines, and comment on outliers and influential observations.

PROGRAM 9: Correlation Correlation and its properties; the relationship between correlation and regression.

PROGRAM 10: Multidimensional Data Analysis The impact of computing technology on statistics, especially graphics for displaying multidimensional data. A case study in data analysis will employ techniques discussed in previous programs.

PROGRAM 11: The Question of Causation Association between categorical variables displayed in a two-way table; Simpson’s paradox; the varied relations among variables that can underlie an observed association; how evidence for causation is obtained.

PROGRAM 12: Experimental Design Advantages of planned data collection over anecdotal evidence or available data. The idea of an experiment. Basic principles of design: comparison, randomization, replication.

PROGRAM 13: Blocking and Sampling: Experiments and Samples Further principles of design: two or more factors and blocking. Introduction to sample surveys: the danger of bias, random sampling.

PROGRAM 14: Samples and Surveys: Sampling and Sampling Distributions More elaborate sample designs: stratified and multistage designs. The practical difficulties of sampling human populations. The idea of a sampling distribution.

PROGRAM 15: What is Probability? Probability as a model for long term relative frequencies or personal assessment of chance. Sample space, basic rules of assigning probability: 0 ≤ P(A) ≤ 1, P(S) = 1, addition rule for disjoint events.
PROGRAM 16: Random Variables  Independence and the multiplication rule for independent events. Discrete and continuous random variables. Mean and variance of a random variable.

PROGRAM 17: Binomial Distributions  The law of large numbers. Addition rules for means and variances of random variables. The binomial distributions for sample counts. Normal approximation to binomial.

PROGRAM 18: The Sample Mean and Control Charts  The sampling distribution of $\bar{x}$. The central limit theorem $\bar{x}$ control charts and statistical process control.

PROGRAM 19: Confidence Intervals  The reasoning behind confidence intervals, $z$-intervals for the mean of a normal distribution. Behavior of confidence intervals.

PROGRAM 20: Significance Tests  The reasoning behind significance tests illustrated by the simple case of tests on a normal mean with known standard deviation. Null and alternative hypotheses and $P$-values, and cautions on the limited information provided by tests.

PROGRAM 21: Inference for One Mean  Inference about the mean of a single distribution, with emphasis on paired samples as the most important practical use of these procedures. The $t$ confidence interval and test.

PROGRAM 22: Comparing Two Means  The two-sample $t$ confidence intervals for comparing means; brief mention of the sensitivity of the corresponding procedures for variances to nonnormality and their consequent impracticality.

PROGRAM 23: Inference for Proportions  Confidence intervals and tests for a single proportion and for comparing proportions based on paired and independent samples.

PROGRAM 24: Inference for Two-Way Tables  Chi-square test for independence/equal distributions in two-way tables.

PROGRAM 25: Inference for Relationships  Inference for simple linear regression, emphasizing slope and prediction.

PROGRAM 26: Case Study  A case study that illustrates the major aspects of statistical thinking: planning data collection, analysis by graphs and informal inference, more data collection in response to partial success.
Study Tips for 36-220
Engineering Statistics and Quality Control
or
How to Use Your Study Time More Efficiently

1. **COME TO CLASS.** I can’t teach you if you don’t show up. [Furthermore, you may miss a quiz, which would negatively impact your grade.]

2. Read your lecture notes over within 24 hours of lecture (or at least once before the next lecture).
   - Highlight or make marginal notes for important words and concepts. This will help fix ideas and will help you to actively learn the material. This review takes about 20-30 minutes and really yields a large return.
   - Re-do examples yourself, step by step, with pencil and paper. Examples often look easy when explained in class, but often turn out to be much harder when you do them yourself.
   - Write down questions about things you do not understand. Bring these questions to lecture, lab and to office hours and ask them.

3. Readings are assigned for each class. Read them. Being well prepared will help you understand the lecture better. Also, as you read, highlight, re-work examples yourself, and write down questions, as suggested above.

4. **DO HOMEWORK PROBLEMS.** Actively doing problems is the *only* way to learn the material. Exam and quiz questions will often be similar to homework problems.
   - Start early. Do not leave assignments until the night before they are due.
   - Try doing the problems yourself before discussing them with other people.

5. *Use office hours* productively. Ask thoughtful questions about things that you do not understand. In other words, if you do (1)-(4) above, it will be much easier to isolate what is giving you trouble.

6. *Review solutions* to assignments and exams. Just because you do not lose points on a homework question does not necessarily mean you fully understand the question and answer. Also, the solutions should serve as a model for how to write, using proper sentences and paragraphs, discussions and interpretations of data analyses.

7. We will make every effort to help you learn the course material, but you must also make an effort to utilize the resources that are made available to help you. Please come talk to us - not only when you are having trouble but also when things are going well.
AN EXAMPLE OF A LETTER YOU DO NOT WANT TO RECEIVE

Dear Student:

This is to follow up on our conversation concerning Homework Assignment 5 in 36-220. The TA’s noticed startling similarities in your paper and that of another student, in particular the same very unusual wrong answer to question 7. When I asked you about it, you agreed that you did work together on it, and, in fact, had each copied from a common sheet.

The course policies, as stated in the syllabus, include “The written solutions to homework problems must be your own and not copied from anyone else.” They also provide that “It is okay to discuss an assignment with other students or with others. When you do so, you must report, with your homework, whom you talked with and the content of the discussion. If you so disclose, and do so accurately, you are not liable under the plagiarism rules.” In this case you did not disclose your collaboration concerning question 7.

After examining the homeworks and talking with you, I find that you have violated the plagiarism rules of the course. I have decided to record a grade of 0 for you for the homework in question. You have the right to appeal this decision, as specified in the CMU Student Handbook.

After discussing the matter with you, I am satisfied that you now understand the policies of the course and will not repeat this behavior.

Good luck in the rest of your academic career at Carnegie Mellon.

Sincerely,

Joseph B. Kadane

cc: Head, Dept. of Statistics
    Dean’s office of the student’s college
    Head, Student’s Department
    Dean of Student Affairs
    Dean of H&SS

JBJhrs

11