

# **TEACHING CHANCE**

**Funded by the National Science Foundation**

**Project # USE 9156215**

**DRAFT**

**Revised July 7, 1998**

**Dartmouth College, New Hampshire**

# TABLE OF CONTENTS

<b>1. What is Chance?</b> .....	
<b>2. Why Teach a Chance course?</b> .....	
<b>3. What does a Chance course look like?</b> .....	
3.1 <i>The Dartmouth Course</i> .....	
3.2 <i>An honors seminar at Minnesota</i> .....	
3.3 <i>Chance as writing across the curriculum</i> .....	
3.4 <i>"Lies, damned lies, and statistics"</i> .....	
<b>4. The Chance Database</b> .....	
<b>5.Using the Chance database to teach a course</b> .....	
<b>6. Course details</b> .....	
6.1 <i>Discussion of news articles</i> .....	
6.2 <i>Homework</i>	
6.3. <i>Using technology in a Chance course</i> .....	
6.3.1 <i>Class Survey</i> .....	
6.3.2 <i>Software programs</i> .....	
6.3.3 <i>Video and audio resources</i> .....	

6.4	<i>Writing activities</i>	.....
6.5	<i>Guest speakers</i>	.....
6.6	<i>Chance Fair</i>	.....
6.7	<i>Assessment</i>	.....
6.8	<i>Other resources to use in a Chance class</i>	.....
<b>7.</b>	<b>Chance guide to topics</b>	.....
7.1	<i>Surveys and sampling</i>	.....
7.2	<i>Experiments and observational studies</i>	.....
7.3	<i>Measurement</i>	.....
7.4	<i>Distributions and Measures of Center</i>	.....
7.5	<i>Variability and the normal curve</i>	.....
7.6	<i>Correlation and Causation</i>	.....
7.7	<i>Time Series</i>	.....
7.8	<i>Probability</i>	.....
7.9	<i>Statistical Inference</i>	.....
<b>8.</b>	<b>Appendix</b>	.....
8.1	<i>Teaching Statistics Using Small-Group Cooperative Learning</i>	.....
8.2	<i>Beyond Testing and Grading</i>	.....
8.3	<i>Experiences with Authentic Assessment Techniques in an Introductory Statistics Course</i>	.....
8.4	<i>Other Instructional Resources on the Web</i>	.....
8.5	<i>Student Projects</i>	.....
8.6	<i>Evaluation Instruments</i>	.....

8.7 *Book Reviews*.....

"Tainted Truth" by Cynthia Crossen.....

"A Mathematician Reads the Newspaper" by John Paulos.....

"The Power of Logical Thinking" by Marilyn vos Savant.....

"Workshop Statistics" by Allan Rossman.....

"Activity-Based Statistics" by Richard Scheaffer et. al.....

"ActivStats" by Paul Velleman

"Electronic Companion to Statistics" by George Cobb

Textbook Reviews Webpage

## 1. What is Chance?

- The **Chance Project** was funded by the National Science Foundation (1992 to 1996) to develop instructional materials for a course called Chance. The Chance course teaches fundamental ideas of probability and statistics in the context of real-world questions of current interest.
- The intent of the Chance course is to help students learn how to think about statistics and probability, how to seek out for themselves the tools appropriate for studying a particular problem, and how to read and critically evaluate quantitative information presented in the media. The course format includes extensive reading and discussion of newspaper and journal articles, computer simulations and activities, writing assignments, and student projects.
- The Chance team produces Chance News, a monthly electronic newsletter distributed to e-mail subscribers. It provides synopses of current newspaper articles involving chance issues, along with suggested thought questions that might form the basis of a classroom discussion on the topic.
- The Chance database on the World Wide Web (<http://www.dartmouth.edu/~chance/>) provides many resource materials for statistics instructors, including archived issues of Chance News. The web version of Chance News includes links to related sites. You will also find syllabi from instructors of Chance courses, and write-ups of classroom activities, statistics videos and student projects that have proved successful.
- Chance is a course, a philosophy of teaching, and an adventure!

## 2. Why Teach a Chance course?

Mathematics is not primarily a matter of plugging numbers into formulas and performing rote computations. It is a way of questioning and thinking that may be unfamiliar to many of us, but is available to almost all of us.

--John Allen Paulos, *A Mathematician Reads the Newspaper*.

The most compelling reason to teach a Chance course is that it is an effective way to help students develop quantitative literacy. Students enjoy the course and appreciate the relevance of topics and the focus on current events and real world applications. Although it is challenging to teach, Chance instructors have found the course rewarding, enjoyable, and an effective model of an innovative course that incorporate suggestions from research on teaching and learning statistics.

Evaluations of previous Chance courses suggested that students improved in their ability to reason statistically and improved their attitudes and beliefs related to statistics. In particular, the evaluations have noted the following positive results:

- Analyses of pre and post measures of students' attitudes and beliefs indicated positive changes in their belief that understanding probability and statistics is important in today's world.
- Students found statistical terms encountered in the media more understandable, felt able to explain how opinion polls work, and had more confidence in interpreting statistical information.
- Measures of statistical understanding indicated that students were better able to question results (and conclusions) of an experiment and showed improved understanding of important ideas in probability and statistics.
- On post-course surveys students expressed greater skepticism of research described in media, sophistication regarding reporting results of statistical tests and methods used to design and conduct research.

### 3. What does a Chance course look like?

There have been different versions of Chance. The best way to get an idea what a Chance course is, is to look at course descriptions for Chance courses. Despite the different versions of Chance courses, all contain the following basic statistical topics:

- Descriptive statistics
- Correlation and regression
- Probability
- Experiments and observational studies
- Samples and polls
- Statistical inference

We start with the Chance course that has been taught for several years at Dartmouth, and also at Princeton, and UCSD.

#### 3.1 The Dartmouth Course

Math 5 (Chance) Fall 1996

Instructors: John Finn and Shunhui Zhu

Course Description

##### **Content**

Welcome to Chance!

Chance is an unconventional math course. The standard elementary math course develops a body of mathematics in a systematic way and gives some highly simplified real-world examples in the hope of suggesting the importance of the subject. In the course Chance, we will choose serious applications of probability and statistics and make these the focus of the course, developing concepts in probability and statistics only to the extent necessary to understand the applications. The goal is to make you better able to come to your own conclusions about news stories involving chance issues.

Topics that might be covered in Chance include:

- Health risks of electric and magnetic fields
- Statistics, expert witnesses, and the courts
- The use of DNA fingerprinting in the courts
- Randomized clinical trials in assessing risk
- The role of statistics in the study of the AIDS epidemic

- Paradoxes in probability and statistics
- Fallacies in human statistical reasoning
- The stock market and the random walk hypothesis
- Demographic variations in recommended medical treatments
- Informed patient decision making
- Coincidences
- The reliability of political polls
- Card shuffling, lotteries, and other gambling issues
- Scoring streaks and records in sports

During the course, we will choose a variety of topics to discuss with special emphasis on topics currently in the news. We will start by reading a newspaper account of the topic in newspapers such as the *New York Times* or the *Boston Globe*. We will read other accounts of the subject as appropriate, including articles in journals like *Chance Magazine*, *Science*, *Nature*, *Scientific American*, and original journal articles. These articles will be supplemented by readings on the basic probability and statistics concepts relating to the topic. We will use computer simulations and statistical packages to better illustrate the relevant theoretical concepts.

### **Organization**

The class will differ from traditional math classes in organization as well as in content: The class meetings will emphasize group discussions, rather than the more traditional lecture format. Students will keep journals to record their thoughts and questions, along with their assignments. There will be a major final project in place of a final exam.

### **Scheduled meetings**

The class meets Tuesday and Thursday from 10:00 to 11:50 a.m. in 102 Bradley Hall.

### **Discussion groups**

We want to enable everyone to be engaged in discussions while at the same time preserving the unity of the course. After a suitable time, we will ask for reports to the entire class. These will not be formal reports. Rather, we will hold a summary discussion between the instructors and the reporters from the individual groups.

Every member of each group is expected to take part in the discussion and to make sure that *everyone* is involved: that everyone is being heard, everyone is listening, that the discussion is not dominated by one person, that everyone understands what is going on, and that the group sticks to the subject.



## Text

The text for the course is *Statistics*, 2nd edition by Freedman, Pisani, Purves, and Adhikari, available from the Dartmouth Book Store and the Wheelock Book Store. Students will also learn to use the JMP statistical package that is available from the public server as a key served application.

## Journals

Each participant should keep a journal for the course. This journal will include:

- Specific assignments that you have been asked to do for your journal. These will include questions they you are asked to think and write about, related to the current day's discussion, the results of computer investigations, etc.
- General comments about the class; things you don't understand; things you finally do understand. You might describe an experience of trying to explain material from class to a friend or family member.
- Finding and commenting on news articles about topics relevant to the course; asking us challenging questions; making connections between what went on in class and experiences in your own life; going to a casino and winning a lot of money.
- Anything interesting and imaginative about a chance topic.

A good journal should answer questions asked, and raise questions of your own, with evidence that some time has been spent thinking about the questions. In addition, there should be evidence of original thought: evidence that you have spent some time thinking about things that you weren't specifically asked about. In writing in your journal, exposition is important. If you are presenting the answer to a question, explain what the question is. If you are giving an argument, explain what the point is before you launch into it. What you should aim for is something that could communicate to a friend or a colleague a coherent idea of what you have been thinking and doing in the course.

You are encouraged to cooperate with each other in working on anything in the course, but what you put in your journal should be you. If it is something that has emerged from work with other people, write down who you have worked with. Ideas that come from other people should be given proper attribution. If you have referred to sources other than the texts for the course, cite them.

Journals will be collected and read as follows:

Thursday	10	October
Thursday	24	October
Thursday	7	November
Thursday	21	November
Tuesday	3	December

## **Homework**

To supplement the discussion in class and assignments to be written about in your journals, we will assign readings from your text FPP, together with accompanying homework. When you write the solutions to these homework problems, you should keep them separate from your journals. Homework assignments will be assigned once a week and should be handed in on Thursdays.

## **Final project**

We will not have a final exam for the course, but in its place, you will undertake a major project. This project may be a paper investigating more deeply some topic we touch on lightly in class. Alternatively, you could design and carry out your own study. Or you might choose to do a computer-based project. To give you some ideas, a list of possible projects will be circulated. You can also look at some previous projects on the Chance Database. However, you are also encouraged to come up with your own ideas for projects.

## **Chance Fair**

At the end of the course we will hold a Chance Fair, where you will have a chance to present your project to the class as a whole, and to demonstrate your mastery of applied probability by playing various games of chance. The Fair will be held during the final examination time assigned by the registrar.

## **Resources**

Materials related to the course will be kept on our web site (<http://www.dartmouth.edu/~chance/>) and on Kiewit public server (go to Public then Courses & Support then Academic Departments&Courses then Math then Chance). In addition supplementary readings will be kept on reserve in Baker Library.

In the Dartmouth Chance Course we start each class by preparing a set of questions relating to a article in the news (usually the current news). The students then break up into groups of 3 or 4 and spend about 20 minutes reading the article and discussing the questions. We then have them report on their ideas and spend the rest of the first hour discussing with them the article. In the second hour we pursue further the statistical issues related to the article. To assist in this we might have students carry out an activity or view a part of the "Against all Odds" video series, We might illustrated, by a computer program, a basic idea such as the normal approximation for the sample means.

We find that the students can read the text FPP and do the review problems at the end of the chapters pretty much on their own. We have an extra hour discussion section for those who need help with this. We find that the basic concepts discussed in a traditional statistics course come up naturally in news articles and this, together with their reading in FPP, provides them with a pretty good understanding of basic concepts or probability and statistics. This did require some skipping around in the text but this is possible with FPP.

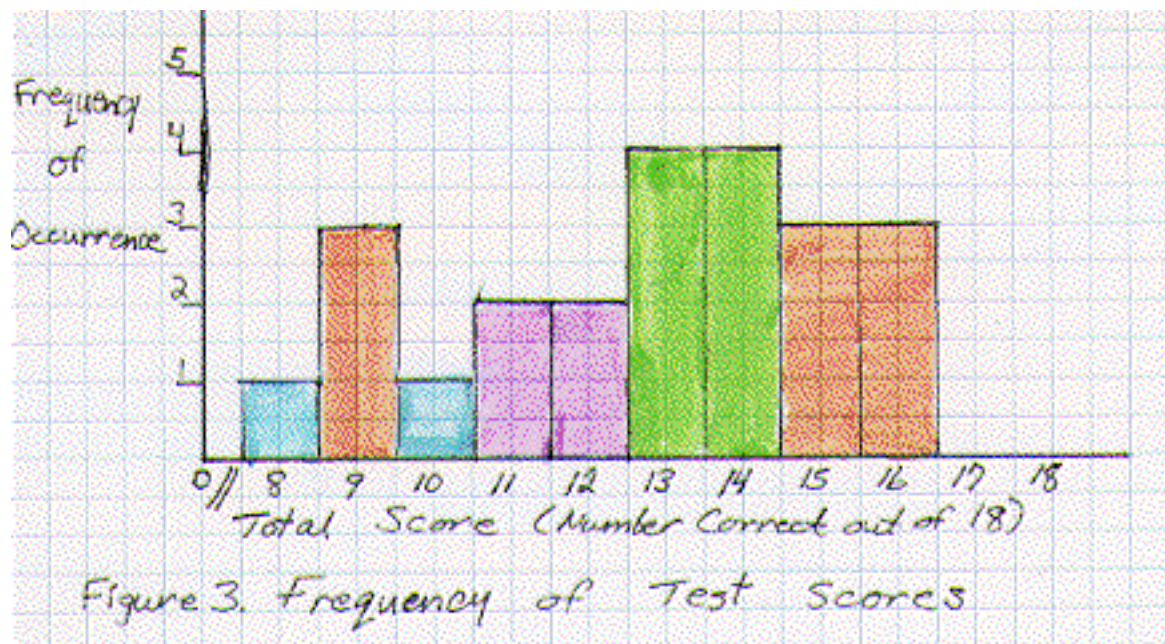
A 12-minute Chance video is available from Laurie Snell ([jlsnell@dartmouth.edu](mailto:jlsnell@dartmouth.edu)). This video includes excerpts from a Chance course taught at Dartmouth as well as interviews with the course instructors.

We will see later examples of typical handouts. A complete set of handouts used in recent courses can be found on the Chance web site.

### 3.2 A Liberal Arts Honors Seminar at the University of Minnesota

Joan Garfield taught a Chance course as an honors seminar at Minnesota. She followed the organization of the Dartmouth course except that she used David Moore's book *Statistics: Concepts and Controversies*. Instead of following current news, as was done in the Dartmouth course, she followed the text and used newspaper articles to provide examples to motivate statistical concepts as they arose in the text. She used current articles when appropriate but did not insist that the articles were current and so was able to choose from a much wider collection of newspaper articles.

Joan did not require a statistical package and it was interesting to see what the students did with presenting their data on their final projects. Here is a typical graph from a project on Polling.



You can see this project as well as two other projects from Joan's course by going to the Chance web site and looking under "Student Projects" in "Teaching Aids."

Joan has also introduced current news and other materials developed for the Chance project into her regular statistics courses. This has led to what we call "Chance Enhanced" courses.

### 3.3 Chance as Writing Across the Curriculum.

Bill Peterson at Middlebury and Tom Moore at Grinnell taught chance as a freshman seminar in their college's writing across the curriculum programs. Both Bill and Tom chose ahead of time four or five topics that were often found in the news and limited their discussions to these topics. Here is a description of a Middlebury course.

CHANCE First-Year Seminar  
FALL 1992  
Middlebury College  
Instructor: Bill Peterson

**COURSE TOPICS**

**Unit I. Public Opinion Polls; Census Undercount.**

Key Concepts

Populations and Samples.  
Variability in Estimators.  
Confidence Statements.

**Unit II. Visual Display of Quantitative Information**

Key Concepts

Displaying Distributions.  
Histograms, Bar Graphs, Scatterplots.  
Abuses of Graphical Displays.

**Unit III. Streaks in Sports**

Key Concepts

Basic Probability Rules.  
Misperceptions of Chance Processes.  
Kahneman-Tversky Research.

**Unit IV. Statistics and AIDS**

Key Concepts

Bayes' Rule (Screening Tests).  
Controlled Experiments (AZT Trial).

**Unit V. DNA Fingerprinting**

Key Concepts

Tests of Significance.  
Independence and the Multiplication Rule.

**CLASS STRUCTURE**

**Texts:**

David Moore, Statistics: Concepts and Controversies  
Victor Cohn, News and Numbers,  
Chance magazine

## Classroom Activities

- \* Discussions of newspaper articles and assigned background reading
- \* Small group work analyzing questions
- \* Physical and Computer simulation experiments
- \* Limited lecture (mechanics)

## Writing Assignments

<i>Style</i>	<i>Assignment</i>
Narrative	Describe a personal experience where you think statistical knowledge would have been helpful to you.
Expository	Describe for a non-specialist the key ideas involved in obtaining information about a population by sampling, citing examples from current news stories.
Critique	Debunk three examples of graphical abuse that you find in the popular press.
Analysis	Simpson's paradox (pitfalls of aggregation) in data on race and imposition of the death penalty.
Argument	Should there be mandatory screening of health care workers for HIV infection?
Research (final project)	Research paper on your choice of Chance topic in the news.

Finally we include an example of a Chance type course that was taught by someone not a member of the Chance Project. This is a Chance type course taught at the University of Toronto by Professor Nancy Reid. She kept summaries of all her classes on her web site. The URL for her course is <http://www.utstat.utoronto.ca/reid/199Y.html>. Here is her course description and first class assignment.

### **3.4 "Lies, Damned lies, and Statistics," Nancy Reid University of Toronto Fall 1996**

The title of the course is a quotation attributed to Disraeli, a British statesman of the nineteenth century. Does it still apply today? This course will consider how statistics and statistical thinking get used (and abused) in a variety of activities, including polling, public health, marketing, advertising, lotteries. Some questions that will be addressed are: Why do newspapers report a "margin of error" for poll results, and what does it mean? How can graphs and charts provide information (or misinformation)? What makes a good graph? How do new cancer drugs get tested, and why doesn't the same protocol work for AIDS? How do studies on mice get extrapolated to humans, and do the results make any sense? What is quality control, and why is it currently so fashionable in North American industry?

Examples of current events in which statistical thinking is a relevant component

- the collapse(?) of the B.C. salmon fishery
- polling results for the U.S presidential election
- standardized testing in the schools
- discovery of new treatments for AIDS
- "Cancer scare forces Bell workers to abandon third-story office" (Globe and Mail, Sept. 7, 1996, A1&A8)

What's required in this course -- what you'll do

- regular attendance, participation and discussion (15%)
- lots of reading
- copy of *The Visual Display of Quantitative Information* by E. Tufte (VDQI)
- copy of *Tainted Truth* by C. Crossen (TT)
- class presentations (2) of approximately 15 minutes (15%)
- short projects regularly (approximately 6 in total) (30%)
- final essay (40%)

What's offered in this course -- what I'll do

- provide handouts each week, in class and at <http://www.utstat.toronto.edu/reid/199Y.html>
- organize discussion around selected themes
- provide background information
- provide directions to further material for reading
- try to keep discussion lively and topical
- try not to get too technical
- provide a sample solution for the first short project, before it's due
- provide guidelines for the final essay
- hold office hours Tuesday 2-3 and Thursday 3-4 (SS6006)

## How will the course be organized

We will consider between five and ten themes for discussion, background, and further investigation. These will be selected from current topics according to interests of the class participants, and may consider any of the topics already mentioned in this outline, and any of the following: statistics in sports (is there a hot streak in basketball, how are tennis players seeded, what would have been the outcome of the 94 World Series, is figure skating judging fair, ...); how to assess risks (should bicycle helmets be required, do power lines cause cancer, ...); probability in everyday life (the Monty Hall problem, picking winning lottery numbers,...); environment vs. heredity ( The Bell Curve, twin studies, genetics and crime); .. Please feel free to suggest other topics for consideration, and be warned that you will be asked about your preferences for choice of topics.

The first theme which will be treated in depth is the use of pictures (usually graphs) to display numerical information. Because numbers are often difficult to assimilate, especially if there are a lot of them, and because statistics is often considered to be very mysterious, many popular accounts in newspapers, magazines, and so on summarize information with some type of graphic. This can be done well or badly, as we will see, and there is a scholarly field of inquiry into how to do it well.

### Sources of information

- daily newspapers: pick your favourite; fairly good discussions of science issues appear in the *New York Times*, *The Times*, and the *Globe and Mail*
- magazines: *Chance*, *New Yorker*, *Saturday Night*, *Scientific American*, *Science*, *Nature*, *J. Amer. Medical Assoc.*, *New England J. Medicine*, *Lancet*.
- books:
  - *Visual Display of Quantitative Information*, E. Tufte [required]
  - *Tainted Truth*, C. Crossen [required]
  - *Seeing Through Statistics*, J. Utts
  - *Statistics: Concepts and Controversies*, D. Moore
  - *Statistics: A Guide to the Unknown*, J. Tanur et al.
- various web sites (more later): two that I've used are
  - electronic newsstand (<http://www.eneews.com/>)
  - *J. Amer. Medical Assoc.* (<http://www.ama-assn.org/>)

More specific reading lists will be provided for each theme.

### Required for the first week

- Find a graph in a newspaper article that interests you. Bring a copy of the graph and accompanying article to class.
- Find out where you can read/borrow/buy current issues of the *New York Times*, *The Times*, the *Washington Post* and the *Globe and Mail*
- Find out where you can read/borrow/buy current issues of *Chance*, *New Yorker*, *Saturday Night*, *Scientific American* and *Lancet*

## 4.The Chance Database

Almost all of the materials and resources used in teaching a Chance Course are included in the Chance Database on the World Wide Web. The Chance Database can be accessed using any Web reader, such as Mosaic or Netscape. The URL address for the database <http://www.dartmouth.edu/~chance/>.

The Chance Database contains current and past issues of Chance news, course syllabi and various teaching aids, and search mechanisms to help locate materials. Short descriptions of some of the sections of the database are described below. These appear as links on the Chance homepage.

- **Chance News**

Chance news is a monthly newsletter that provides abstracts of articles in current newspapers and journals that involve probability or statistics concepts. Links are made to related resources at other web sites. Discussion questions are provided for many of the articles. You can receive Chance News by e-mail by sending a request to [jlsnell@dartmouth.edu](mailto:jlsnell@dartmouth.edu).

- **The Chance Course**

This section contains syllabi of previous Chance courses and articles that have been written about the Chance course.

- **Teaching Aids**

Here you will find articles and teaching aids listed by topic, descriptions of activities, datasets, information about video tapes and other resources we have found useful in teaching a Chance course.

- **Other related internet sources**

This section provides links to other servers that have materials useful for teaching a Chance course or other probability or statistics courses.

- **Fair Use**

Look here for a discussion of fair use as it applies to materials on this database.

If you click on any of the underlined items in the database you are led to the next level of detail. For example, if you click on **Chance News** you see:

- **What is Chance News?**
- **Current issue**
- **Previous issues**
- **Search Chance News**

Clicking on **Current Issue** will give you the most recent issue which itself will have links to full text of articles and to articles from other databases. For example, you might be referred to the Bureau of Labor Statistics, or the *Journal of the American Medical Association* for an article that was the basis for the newspaper article.



## 5. Using the Chance Database to Teach a Course

As we have remarked, a typical class in the Dartmouth Chance course starts with students breaking up into groups of 3 or 4 to discuss a series of questions provided relating to a current issue in the news. We often use Chance News to choose an appropriate article. As this is being written the current Chance news (27 May to 26 June 1998) has the following list of articles abstracted:

1. Two new reports: wolves and meteorites.
2. Two new Chance Lectures videos.
3. Radio programs on chance topics.
4. Life by the numbers.
5. Investing it; duffers need not apply.
6. Textbook solution to the birthday problem.
7. Ask Marilyn: What does margin of error mean?
8. A new poll website.
9. Online voters get Hankerin' for Anarchy.
10. Sampling and Census 2000: the concepts.
11. Hidden truths.
12. Midwives beat doctors in government survey.
13. Will you please just park?
14. A debate is unleashed on cholesterol control.
15. Accuracy of eyewitness recall of suspects tested.
16. Coincidences and linguistics.
17. Experts seek to avert asteroid scares.
18. Unprovoked shark attacks found to increase worldwide.
19. Dining out in L.A. comes to crunching numbers.
20. Coincidence or conspiracy?

You can click on any of these abstracts that might interest you. You will see where the article occurred, be able to read an abstract of the abstract along with discussion questions related to the article.

Our choice of article might depend on what topic we want to emphasize. If we were discussing sampling we might choose the item:

Investing it; duffers need not apply.  
 The New York Times, 31 May 1998, Section 3, p. 1  
 Adam Bryant

An investment compensation expert, Graef Crystal, carried out a study purporting to show that the major companies, whose C.E.O's had low golf scores, had high performing stocks. Crystal obtained data for golf scores from the journal Golf Digest and used his own data on the stock market performance of the companies of 51 chief executives. He created a Stock Rating which gave each company a stock rating based on how investors who held their stock did with 100 being highest and 0 lowest.

It is rare that an article in the New York Times includes the data set, but this article did. Here it is, as sent to us by Bruce King (we have saved it on the Chance Website in the data section of Teaching Aids): In the article the data is arranged as follows:

CEO	Company	Handicap	StockRate
Top 25% of golfing executives			
Melvin R. Goodes	Warner-Lambert	11	85
Jerry D. Choate	Allstate	10.1	83
Charles K. Gifford	BankBoston	20	82
Harvey Golub	American Express	21.1	79
John F. Welch Jr.	General Electric	3.8	77
Louis V. Gerstner Jr.	IBM	13.1	75
Thomas H. O'Brien	PNC Bank	7.1	74
Walter V. Shipley	Chase Manhattan	17.2	73
John S. Reed	Citicorp	13	72
Terrence Murray	Fleet Financial	10.1	67
William T. Esrey	Sprint	10.1	66
Hugh L. McColl Jr.	Nationsbank	11	64
Average		12.4	76

Middle 50% of golfing executives

James E. Cayne	Bear Stearns	12.6	64
John R. Stafford	Amer. Home Products	10.9	58
John B. McCoy	Banc One	7.6	58
Frank C. Herringer	Transamerica	10.6	55
Ralph S. Larsen	Johnson & Johnson	16.1	54
Paul Hazen	Wells Fargo	10.9	54
Lawrence A. Bossidy	Allied Signal	12.6	51
Charles R. Shoemate	Bestfoods	17.6	49
James E. Perrella	Ingersoll-Rand	12.8	49
William P. Stiritz	Ralston Purina	13	48

Duane L. Burnham	Abbott Laboratories	15.6	46
Richard C. Notebaert	Ameritech	19.2	45
Raymond W. Smith	Bell Atlantic	13.7	44
Warren E. Buffett	Berkshire Hathaway	22	43
Donald V. Fites	Caterpillar	18.6	41
Vernon R. Louckes Jr.	Baxter International	11.9	40
Michael R. Bonsignore	Honeywell	22	38
Edward E. Whitacre Jr.	SBC Communications	10	37
Peter I. Bijur	Texaco	27.1	35
Mike R. Bowlin	Atlantic Richfield	16.6	35
H. Lawrence Fuller	Amoco	8	33
AVERAGE		14.6	47

Bottom 25% of golfing executives

Ray R. Irani	Occidental Petroleum	15.5	31
Charles R. Lee	GTE	14.8	29
John W. Snow	CSX	12.8	29
Philip M. Condit	Boeing	24.2	25
Joseph T. Gorman	TRW	18.1	24
H. John Riley Jr.	Cooper Industries	18	22
Richard B. Priory	Duke Energy	10	22
Leland E. Tollett	Tyson Foods	16	20
Bruce E. Ranck	Browning-Ferris	23	15
William H. Joyce	Union Carbide	19	13
Thomas E. Capps	Dominion Resources	18	12
AVERAGE		17.2%	22

Left in the clubhouse

Scott G. McNealy	Sun Microsystems	3.2	97
William H. Gates	Microsoft	23.9	95
Sanford I. Weill	Travelers Group	18	95
Frank V. Cahouet	Mellon Bank	22	92
William C. Steere Jr.	Pfizer	34	89
Donald B. Marron	Paine Webber	25	89
Christopher B. Galvin	Motorola	11.7	3

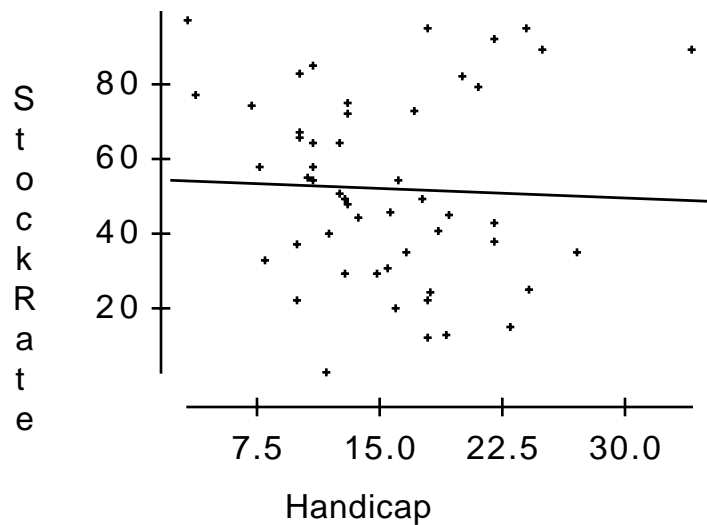
Crystal regarded the last seven as outliers and threw them out (described in the article as being scientifically sifted out). Christal notes that the averages of the stock ratings decrease as the average of the golf scores increase and concludes from this and remarks:

“For all the different factors I’ve tested as possible links to predicting which C.E.O.’s are going to perform well or poorly, this is certainly of the oddest—but also the strongest—I’ve see,” he said. “There’s got to be something here.”

We give the students copies of the full text of the article (available from the database) and ask them to discuss one or two of these, or other questions we might make up relating to the article, for about twenty minutes. For example we might use the following questions:

**DISCUSSION QUESTIONS:**

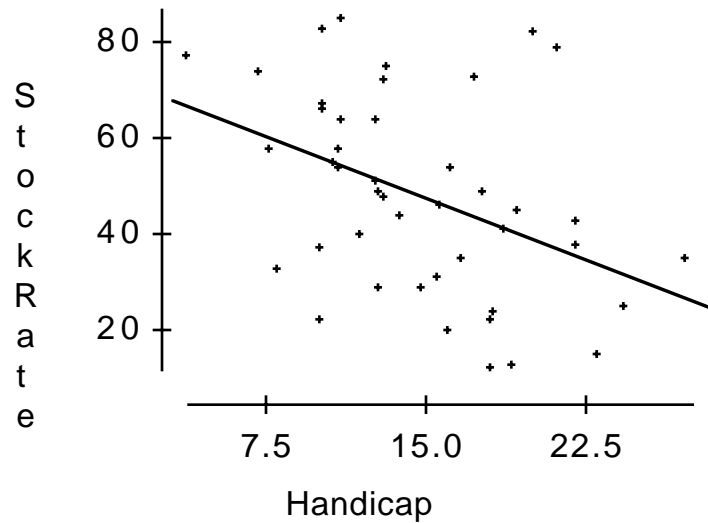
- (1) What do you think of Crystal’s method for showing correlation?
  
- (2) Consider Crystal’s remark: “For all the different factors I’ve tested as possible links to predicting which C.E.O.’s are going to perform well or poorly, this is certainly the oddest. What does this suggest to you?”
  
- (3) Crystal goes on to say: “As tidy as the statistical correlation may seem there remains the tricky matter of figuring out why better golfers also tend to be better chief executives, or vice versa.” Must there be a reason?
  
- (4) If Crystal had plotted a scatter plot of his data he would have seen:



with a correlation  $-.042$ . Do you think Chrystal looked at this?

Where are the outliers? Do they look like outliers?

If we remove the outliers we obtain the following scatterplot:



This results in a correlation of  $-.414$ . Does this establish a significant correlation between C.E.O.'s golf performance and the stock records of their company?

What is the criteria for a point to be an outlier in a scatterplot?

We then continue a discussion of the topic with the students using their responses as a basis for our discussion.

As we have remarked, Joan Garfield in her Chance course at Minnesota followed a somewhat different approach. She followed the order of the text and uses current and previous Chance news to supplement her discussion. One way articles were used was to assign students to pick an article to read and critique corresponding to a particular statistical topic. To assist in this we grouped some articles in Chance News and other resources by topic.

If you look in **Teaching Aids** you will find Articles and materials relating to specific topics. Clicking on this gives:

**Articles and materials relating to specific topics.**

- **Descriptive statistics**
- **Correlation and Regression**
- **Experiments**
- **Measurement**
- **Polls and surveys**

- **Probability**
- **Lotteries**
- **AIDS**
- **Weather Prediction**
- **Marylin vos Savant column**
- **DNA Fingerprinting**
- **Smoking**

The first topic in Moore's book is Surveys and Polls. Looking at this category we find the following suggested articles and teaching aids.

#### **Articles related to polls**

- "The Numbers Game" from Newsweek
- "The trustworthiness of Survey Research" by Judith Tanur in The Chronicle of Higher Education
- So, now we know what Americans do in bed, so? NYT
- What polls say -- and what they mean, NYT

#### **Instructional materials related to polls and survey**

- Discussion questions: how newspapers describe polls
- Discussion questions: a radio station account of an election
- The census undercount problem: Bill Peterson
- Against all Odds Video #13: Blocking and Sampling
- Against all Odds Video #14: Samples and Surveys
- Discussion questions on the meaning of margin of error
- Discussion questions related to a sex survey
- Website: PollingReport.com  
(<http://www.pollingreport.com>) includes national polling data -- from major research organizations, like Gallup, Harris, Yankelovich, and Princeton Survey Research -- on a variety of issues.

Thus if we wanted to emphasize the importance of careful wording of survey questions we could choose the article by Judith Tanur which has examples of poor wordings and suggestions for good wordings. We might want to discuss the concept of margin of error and, if so, we could use some of the discussion questions on the margin of error used in previous courses.

Another way to find articles on a particular statistical topic is to use the "Search Chance News" feature. For example, if we wanted to find a current article that mentions correlation, we could search on correlation. This yields a long list of articles in Chance News, including the following recent articles:

>>>>=====>

Global-scale temperature patterns and climate forcing over the past six centuries  
Nature, 23 April 1998, 779-787  
Mann, Bradley, and Hughes

>>>>=====>

Politics of youth smoking fueled by unproven data  
The New York Times, 20 May 1998, A1.  
Barry Meier

>>>>=====>

Colleges look for answers to racial gaps in testing.  
New York Times, 8 Nov. 1997, A1  
Ethan Bronner

>>>>=====>

## 6. Course details

### 6.1 Discussion of news articles

What is the optimal number of articles to use? How can you select the best article on a particular topic? These are challenging questions given the daunting number of articles available on any statistical topic.

We recommend choosing one good article on a topic, or two complementary articles, and posing 3-4 questions to stimulate discussion. Sometimes additional articles may be assigned to read outside of class and critiqued in their journals.

The articles may be selected by:

- Choosing an article on the same topic from the Chance Data base
- Choosing an article from the current Chance news
- Have students select an article from a newspaper, journal, or magazine of their choice.
- Give students an article (hard copy) to read outside of class.

In order to stimulate discussion of a particular article, it helps to have a mix of questions based on factual information as well as opinion. Instructors may find that students are initially shy about sharing their opinions or answers with the class. Some suggestions for stimulating discussion include:

- Wait for 30 seconds after asking a question to allow students to formulate their answers.
- Have students discuss articles in groups and assign one person in each group to be the reporter. Then, after the small groups have discussed the articles, call on the reporters to report the summaries of group discussion.
- Encourage students to respond to each others questions and comments, rather than commenting on each student's response. For example, instead of commenting on one students' response, ask: what do you think? Does anyone agree or disagree?

As an alternative to having students read articles during class time, instructors may choose to assign articles for students to read before coming to class.

## **6.2 Homework**

One of the key features of a Chance course is that students are assigned material to read in their textbook, rather than receive the material via a lecture during class. By reading the assigned material outside of class students are able to spend class time discussing articles and working on activities that promote conceptual understanding of important statistical ideas.

We suggest assigning sets of problems corresponding to each section of the reading. Solutions to problems may be handed in at each class or each week to be graded by the instructor or teaching assistant. An alternative model has students correct their own assignments and turn them in every few weeks.

Note: We believe that it is important that students keep current in the assigned reading as they will need to build on the concepts and methods in class. We have found that students appreciate having connections made between concepts in the assigned reading and activities in class.

## **6.3 Using Technology in a Chance course**

The computer has been used in Chance courses in different ways: (a) we use a statistical package to explore data, (b) we use applets or simulation programs to illustrate statistical concepts, (c) we use video and audio resources on the web to provide more detail on various topics, and (d) we use a simple language such as Basic to write programs to simulate and analyze experiments. It would also be possible to use Mathematica or Maple for this purpose.

The statistical packages that have been used in the courses we have taught are Minitab, Data Desk (with the associated CD-ROM ActivStats) and JMP, though there are many other similar packages.

ActiveStats is a CD-ROM that combines a statistical package with a complete discussion by Paul Velleman of the basic concepts of statistics. For more about ActiveStats see the review at the end of this Teaching Guide. Another CD-ROM that can be used to supplement is An electronic companion to statistics by George Cobb. More information about this can also be found in the review of this CD-ROM at the end of this Guide.



### 6.3.1 Class Survey

We often start a Chance course by asking the students to provide some data about themselves and then the students get started using the statistical package by exploring this data. It is a real challenge to think of a set of questions that will lead to a variety of interesting statistical questions. Here is a sample of a questionnaire that we have used. Other class surveys are included in the Chance Data Base.

Class Survey, Grinnell College, 1994

1. Where do you live? (1 = North Campus, 2 = South Campus, 3 = Off Campus) \_\_\_\_\_
2. Which year are you? (95, 96, 97 or 98) \_\_\_\_\_
3. Are you male or female? \_\_\_\_\_
4. What is your height in inches? \_\_\_\_\_
5. What is your shoe size (length, not width)? \_\_\_\_\_
6. Do you smoke? (0 = no, 1 = occasionally, 2 = regularly) \_\_\_\_\_
7. Are you left or right handed? \_\_\_\_\_ (0 = left, 1 = right)
8. How much did you spend on your last haircut (including tip)? \$ \_\_\_\_\_
9. How many CD's do you own? \_\_\_\_\_
10. What was your verbal SAT score? \_\_\_\_\_
11. What is your current gpa? \_\_\_\_\_
12. How much exercise do you get per week (hours)? \_\_\_\_\_
13. What type of person do you consider yourself to be?  
(1 = very liberal \* 5 = very conservative) \_\_\_\_\_
14. On average, how many hours of television do you watch per week? \_\_\_\_\_
15. On average, how many hours per week do you devote to a class in your major? \_\_\_\_\_
16. On average, how many hours per week do you devote to a non major class?  
\_\_\_\_\_
17. Which division is your major (or likely major) in: humanities, social studies, science? \_\_\_\_\_ (1 = humanities, 2 = social studies, 3 = science)
18. Record your pulse (the number of beats in one minute) after measuring it in class. \_\_\_\_\_
19. What is the average class size for the classes you are taking this semester? \_\_\_\_\_

We put the results of the survey into a dataset and ask the students to see what they can learn from this data. For example, look at a scatter plot of hours watching TV and grade point average to see how these are related.

Since students often do polls for their final project it is useful to show them how to input large amounts of data. For example, when Data Desk is used it is easiest to use an application like Excel to record the data and then import it into Data Desk for analysis.

### 6.3.2 Software Programs

Programs to simulate experiments are sometimes written by the instructor and the students can run and modify the programs. Some students write their own programs as part of a final project. If you use Minitab it is possible to write procedures to simulate experiments within Minitab. For most other simple statistical packages such as Data Desk and JMP, writing some programs is possible, though not as straightforward as Minitab. On the other hand, the ActivStats CD-ROM has some very attractive built-in simulation demonstrations.

Here is a typical way that we would use simulation. One of the major problems in DNA fingerprinting centers around deciding if information on different parts of the DNA are independent. To show how to test this we ask the students if hair-color and eye-color are independent traits. It is easy to collect data from the class to test this restricted ourselves to dark and light for both traits. We obtain data for a 2x2 contingency table.

It quite intuitive that the chi-squared statistics is a reasonable way to measure the deviation of the data from that we would expect if the traits were independent. To see how this statistic varies when they are independent we simulate a large number of experiments assuming independence. This simulation is very easy to carry out using a simple language like Basic or one of the standard mathematical packages such as Mathematica or Maple. The data from this simulation is then imported into our statistical package and we look at a histogram of the data to determine how large the chi-squared value would have to be to reasonably conclude that the traits are not independent.

We also use a probability demo written by John Finn. This demo illustrates a number of important probability concepts such as the Central Limit Theorem, the Poisson distribution, geometric probability illustrated by the buffon needle problem etc.

Finally, we have experimented with using the result of another NSF project called EESEE An Electronic Encyclopedia of Statistical Examples and Exercises. More information about this project can be obtained from the EESEE web site at <http://stat.mps.ohio-state.edu/projects/eeseen/index.html>

### 6.3.3 Video and Audio Resources

#### *Videos*

We have found it very useful to show videos from either of the COMAP series: *Against All Odds* or *Decisions Through Data*. The *Against All Odds* videos can be ordered from Annenberg CPB at 1-800-Learner. (Note: if calling from overseas, use this number: 802 862 8881 ext 547). The *Decisions through Data* videos can be ordered from COMAP at 1-800-772-6627.

The *Against All Odds* series consist of 26 thirty-minute videos:

<b>Tape</b>	<b>Statistical Topic</b>
1	What Is Statistics
2	Picturing Distributions
3	Describing Distributions
4	Normal Distribution
5	Normal Calculations
6	Time Series
7	Models for Growth
8	Describing Relationships
9	Correlation
10	Multidimensional Data Analysis
11	A Question of Causation
12	Experimental Design
13	Blocking and Sampling
14	Samples and Surveys
15	What is Probability?
16	Random Variables
17	Binomial Distribution
18	The Sample Mean and Control Charts
19	Confidence Intervals (beginning of inference)
20	Tests of significance
21	Inference for One Mean
22	Comparing Two Means
23	Proportion Inferences
24	Inference for Two-Way Tables
25	Inference for Relationships
26	CASE STUDY -- AIDS

Decisions Through Data (hereafter, DTD) consists of 5 video cassettes containing twenty-one 10-15 minute units on basic topics of descriptive and inferential statistics.

<b>Unit</b>	<b>Topic</b>
1	What is Statistics
2	Stemplots
3	Histograms and Distributions
4	Measures of Center
5	Boxplots
6	Standard Deviation
7	Normal Curves
8	Normal Calculations
9	Straight Line Growth
10	Exponential Growth
11	Scatterplots
12	Fitting Lines to Data
13	Correlation
14	Save the Bay
15	Designing Experiments
16	Question of Causation

17	Census and Sampling
18	Sample Surveys
19	Sampling Distributions
20	Confidence Intervals
21	Tests of Significance

*Chance Lectures* (on the Chance Database)

Here you will find an experimental form of a series of Chance Lectures given at Dartmouth in December 1997. Viewing them requires the free RealPlayer 5.0 which can be downloaded from the website. These videos are being served from the Mathematical Sciences Research Institute (MSRI) at Berkeley. You will also find at MSRI a large and continually growing collection of videos of mathematical lectures.

Risk in Everyday Life, Arnold Barnett , MIT

National Weather Forecasting, Daniel Wilks Cornell University

Local Weather Forecasting, Mark Breen , Fairbanks Museum

Stock Market Valuation, Sheri Aggarwal, Dartmouth College

Probability and Statistics in Gaming, Olaf Vancura, Harvard University

Polls, David Moore, Gallup Organization

DNA Fingerprinting in the Courts, Bruce Weir , North Carolina State University

Statistical issues in ESP research, Ray Hyman , University of Oregon

Statistics in sports, Hal Stern, Iowa State University

Census 2000, Tommy Wright, Bureau of the Census

Bible Code, Brendan McKay, Maya Bar-Hillel, Jeffrey H. Tigay, Australian National University, Hebrew University, University of Pennsylvania

The Kemeny Lectures on Finite Mathematics, are tapes of lectures from the finite math course taught at Dartmouth. These videos can be found at Peter Doyle's website:

<http://math.ucsd.edu/~doyle/docs/kemeny/clips/cover/cover.html>

The topics are:

Compound statements and truth tables

Example: equivalence and implication

Introduction to tree diagrams

Excerpt: Professor Snell and the urn

Set operations (review)

Counting problems

Partitions and binomial coefficients

What is 0!?

Partitions and binomial coefficients (review)

How many subsets?  
Pascal's triangle  
Binomial coefficients and Pascal's triangle (review)  
Introduction to probability  
Introduction to probability (review)  
Simulation in probability  
Two examples: binomial coefficients and odds  
The birthday problem  
Conditional probability (preview)  
Advice on exam-taking  
To add, or to multiply?  
Conditional probability  
Conditional probability (review)  
Venn diagram probability examples  
Example: true-false test strategy  
Example: gerrymandering  
Example: false positives and false negatives  
Coin tossing (preview)  
Coin tossing and binomial probabilities  
Coin tossing and binomial probabilities (review)  
Example: conditional probabilities by reversing the tree  
Example: getting exactly half heads  
Example: take the negation  
Plotting binomial probabilities; introduction to CLT  
The Central Limit Theorem  
Excerpt: applications of the CLT  
GAME DELAYED--DOG ON FIELD

## Audio Resources

(From the June 1998 *Chance News*)

National Public Radio (NPR) recently had two interesting programs related to chance issues: one on the Bible Codes and another on streaks in sports. Radio programs available on the web provide an interesting source of course material, so we have added links to these on the Chance website under "Teaching Aids." Here is a list of programs today:

NPR Weekend Edition June 13, 1998

A brief but quite good discussion with Keith Devlin of coincidences with special reference to Bible Codes.

NPR Science Friday, May 29, 1998

Richard Harris interviews psychologist Tom Gilovich and Ian Stewart about math in everyday life including discussions of streaks in sports and the birthday problem. There is a particularly interesting discussion between the guests and listeners who call in with their answers to the "birthday problem".

NPR Science Friday, 21 June 1996

A two-hour report from the first World Skeptics Congress. In the first hour the guests are: Paul Kurtz, Philosopher, John Paulos, Mathematician, Milton Rosenberg, Social Psychologist, and

Phillip Adams, Broadcaster and Television Producer. They discuss the media's obsession with pseudoscience and, more generally, how radio and television do report science news and how they should report science news.

In the second hour the guests are: Kendrick Frazier, Editor of *Skeptical Inquirer*, Joe Nickell, Senior Research Fellow for the Committee for the Scientific Investigation of Claims of the Paranormal, Ray Hyman, Psychologist, and Eugenie Scott, Executive Director National Center for Science Education. They discuss what it means to be a critical skeptic. They illustrate the way that professional skeptics study and explain paranormal phenomena.

Car Talk: Week of 5/23/98

The cartalk brothers discuss the infamous two-boys problem: Given that a family with two children has at least one boy, what is the probability the family has two boys.

Car Talk, Weeks of 10/18/97 and 10/23/97

The cartalk brothers discuss the Monty Hall problem. You will find more about this problem, including a historical discussion and a chance to play the game at Monty Hall Puzzler.

## 6.4 Projects

Student projects in a Chance course typically fall into one of two categories.

In one type is students ask a question, devise a plan to collect data to answer the question, collect and analyze the data, and present the results in written and oral reports. A second type is a research report on a topic of interest that may have come up in class, discussion, Chance News, or on a video. This report involves reading articles in magazines, journals, and newspapers, summarizing the issues involved, and drawing conclusions. Samples of student projects are included in the Chance Database.

### Sample guidelines given to students from U of M Honors Seminar

#### 1. What is a project?

A project is a report and poster display that focuses on a topic involving Chance, and is completed instead of a final exam.

*Projects may be completed individually or by a small group (2-3 students).*

You may choose to do a project on any topic that interests you that involves Chance. Your project may be one of the following:

a. A research report on a topic that interests you, that may have come up in class, discussion, Chance news, or on a video. This report would involve reading articles in magazines, journals, and newspapers, summarizing the issues involved, and drawing conclusions. The paper would be 5-10 pages in length, typed, with a reference list.

Examples of possible topics include:

The use of DNA fingerprinting in the courts, The reliability of opinion polls, the controversy on second hand smoke and cancer, the roll of statistics in studying the AIDS epidemic, the use of

statistics in weather forecasting, the statistics of gambling, the census undercount, card shuffling, etc.

b. A project that involves gathering and analyzing data. If you have had a statistics course already, or if you haven't, but would like to apply techniques you've read about in the text book, you may want to design and conduct a statistical study. This study may involve developing a survey or set of interview questions on a topic of your choice, selecting a sample, and administering your instrument. The responses will be summarized and a report written, describing what you did and what you found out. Or, you might want to design an experiment (a series of taste tests, or some other experiment of interest). Again, you would gather data, analyze it, and write up the results.

c. If you have another idea for a project that sounds different than the two types listed above, see me.

## **2. How can I help you?**

First, I will ask you to write up a proposal for your project, telling me if you are planning to work alone or in a group, and what you propose to do. I will help you refine your ideas into a workable project.

As you proceed with your project, I will share materials I have with you, help you locate resources, and help you enter and analyze data (if this is part of your project).

I will distribute guidelines for writing the final project reports, and criteria to be used in evaluating the projects.

I will be glad to review drafts of papers or portions of papers as they are written.

## **Chance Project Format**

*You will submit a written report for your project as well as make a poster display for our in-class Chance Fair. Both forms of your project should include the following components:*

- 1. Statement of the Problem:** Purpose of your project.  
What problem(s) or question(s) did you set out to solve? What were the key issues raised?
- 2. Background:** Preparation for conducting the project.  
Describe how you prepared for your project. What type of background reading did you do? What information did you use in order to better conceptualize your project and frame a design?
- 3. Method:** What you decided to do and how you did it.  
How did you gather information (via experiment, survey, or other data collection method)?
- 4. Results:** The summary and presentation of data gathered.  
This may include tables, graphs, and/or verbal summaries.
- 5. Conclusions:** What you learned about the problem(s) or question(s) you set out to solve.
- 6. Critique:** What you learned about the process of doing your project. What went wrong? What would you do differently next time? What advice would you give future students in this class?

**Posters:** Large print should be used, and information on each component may be fairly brief. Try to use a catchy title that captures the nature of your project.

**Evaluation criteria for posters:**

1. Does the poster include each of the 6 components?
2. Is the material clearly displayed?
3. Does the poster convey the most important aspects of the projects?

Note: The ASA introduction to the Poster Competition book gives nice guidelines for developing a poster.

**Papers:** should be double-spaced and 5 to 10 pages in length (group papers may be longer). Any standard format is fine (e.g., APA, U of Chicago, etc.). An appendix should include a list of your actual data and a copy of the survey or data recording form you used (if you used one). A brief reference list should include any of the resources included in your background reading.

**Evaluation criteria for papers:**

1. Does the paper include each of the 6 components, with each component clearly labeled with a title?
2. Is the paper clear and easy to read, with correct usage of statistical terms?
3. Does each component include the appropriate material and make sense?
4. Is an appendix included containing appropriate materials?
5. Does the reference list include appropriate references?

**Chance Project Progress Report 1**

1. What is the topic of your project?
2. What are the main issues or problems you plan to address?
3. What are your plans for obtaining information? What resources do you plan to use in developing your project?
4. If you are planning to gather data, please describe your data gathering plan, including sample size and number of variables measured.
5. If you are planning to develop and give out a survey, please list below the questions you are planning to use.

**Project Progress Report: 2**

1. How far along are you on your project in terms of each of the following:
  - a. Background reading
  - b. Data collection
  - c. Data analysis
  - d. Reporting of results



e. Planning and design of poster

2. What questions and/or concerns do you have about your project?

Guidelines for both types of project are included in the appendix.

## 6.5 Group activities

There have been many proposals in the educational literature for using cooperative group activities to enhance student learning. Small groups may be used in different ways in a Chance course: to discuss an article in the news, to generate data for an in-class activity, or to work on a project outside of class.

We have found that groups of three or four students work best. Groups may be formed in many different ways, including allowing students to form their own groups, randomly assigning students to groups, or having students number off to form the designated number of groups. Groups may be kept the same or changed throughout the course.

Suggestions for helping groups to work include:

- a. Make sure each group has a clear sense of what they are to do and accomplish and how they are to demonstrate or report the group result.
- b. Assign group roles for in-class activities to make sure that all members participate and to help students going on their assigned task. These roles might include the jobs of leader/moderator, reporter, recorder, summarizer, and encourager.
- c. Monitor the groups to see how well they are working by walking around, observing, and listening. Instead of answering questions asked by an individual group member, first ask if anyone in the group knows the answer to that questions.

For more information on groups, an article on Cooperative Learning is included in the Appendix.

## 6.6 Writing activities

We have seen that one of the formats in which Chance has been successfully offered is the Freshman Writing Seminar. The curricular goal of such courses is to provide first-year undergraduates with a small, discussion-based (as opposed to lecture-based) course where there is significant focus on the writing process. We have found the Chance course to provide an ideal context for such a seminar.

Writing in statistics courses is in general gaining respectability. Gudman Iversen ("Writing Papers in a Statistics Course." 1991 *Proceedings of the Section on Statistical Education* -- American Statistical Association (1991):29-32) in rationalizing his use of writing in his statistics courses says, "... any writing is good for you." Noreen Radke-Sharpe ("Writing as a Component of Statistics Education. *The American Statistician* 45(1991):292-293.) gives (and then expands upon) advantages of required writing in a statistics course:

- (1) it improves writing skills,

- (2) it focuses internalization and conceptualization of material,
- (3) it encourages creativity, and
- (4) it enhances the ability to communicate methods and conclusions.

In her paper Radke-Sharpe gives a variety of suggested writing assignments.

Many of the technical terms from probability and statistics have counterparts in everyday speech, where they are used less carefully than would be the case in scientific discourse. It is obvious that people frequently talk about chances, odds and likelihood without relying on a formal probability model. But also, people invoke a so-called "Law of Averages", which is often used to defend conclusions that don't follow from the statistician's Law of Large Numbers (see the article on the "Law of Averages" by Ann Watkins in the Spring 1995 edition of *Chance* magazine). Similarly, most people have a nodding acquaintance with the idea of a "bell curve", without understanding conditions under which appeals to the Central Limit Theorem might make sense, or when other data models might be appropriate. Writing assignments requiring the exposition and application of such concepts offer valuable lessons on the precise use of language. Students find they must find a balance between using their own voice and the risk of blurring sometimes subtle technical distinctions.

Journals may be useful in the seminar as informal or "free writing" opportunities to record thoughts on the day's discussion in class, pose questions for the instructor, and record solutions to homework exercises. However, the seminar format also requires more structured writing assignments.

We have found that the topics in a *Chance* course lend themselves to a number of different writing formats. Listed in the appendix are details and comments on writing assignments from two different *Chance* courses.

### **Comments by Bill Peterson on writing assignments.**

Recall that the writing assignments for my course at Middlebury were:

<i>Style</i>	<i>Assignment</i>
Narrative	Describe a personal experience where you think statistical knowledge would have been helpful to you.
Expository	Describe for a non-specialist the key ideas involved in obtaining information about a population by sampling, citing examples from current news stories.
Critique	Debunk three examples of graphical abuse that you find in the popular press.
Analysis	Simpson's paradox (pitfalls of aggregation) in data on race and imposition of the death penalty.

Argument	Should there be mandatory screening of health care workers for HIV infection?
Research (final project)	Research paper on your choice of Chance topic in the news.

The first five assignments here ranged between 1 and 4 pages in length; the last was an 8-10 page research paper. The narrative paper was assigned on the first day of class, to get the students immediately involved in writing.

The expository paper on polls proved surprisingly difficult for the students. Formulating a clear statement of what a "margin of sampling error" means is not an easy task: it brings to the fore exactly the issues described above regarding colloquial connotations vs. technical meanings of key terms. Critiques of newspaper reports were illuminating here. The *New York Times* commonly includes a side-bar with an article describing a sample survey, entitled "How the Poll was Conducted." There one finds statements such as "In theory, in 19 cases out of 20 [the 95% confidence statement] the results based on such samples will differ by no more than five percentage points [the margin of sampling error] in either direction from what would have been obtained by seeking out all voters in the country." In one instance, it was reported that this "latest poll was conducted by telephone interviews...The sample of telephone exchanges was selected by a computer from a complete list of exchanges in the country. The exchanges were chosen so as to assure that each region of the country was represented in proportion to its population [stratified sampling]." Other polls are less clear in reporting their methods or the proper interpretation of their results. A perhaps disturbing number are now including disclaimers of the form "This is not a scientific poll, but *represents a sampling* of reader opinion...[emphasis added]". It would appear that, even as they confess their use non-probability sampling, these polls drop terms that attempt to sound like the statistician's goal of a "representative sample". Sorting out language issues like this is an important exercise.

The critical paper on graphs was adopted in the spirit of Edward Tufte's classic *The Visual Display of Quantitative Information* (Cheshire, Connecticut: Graphic Press, 1983). The idea is to contrast the exploratory data analysis techniques introduced in the course to the grossly misleading data displays so often found in the popular press. As with the polling paper, this led to interesting class discussions on why these abuses persist.

The analysis paper on Simpson's paradox was an eye-opener for many students. Inevitably, the first presentation of an aggregation paradox in class leads to a stunned silence. The phenomenon sheds light on what students recognized as a common occurrence in political debates: both sides are able to quote summaries of the same data that appear to support their own point of view. It is instructive to see that there is a rational way out of the paradox, and that one is not forced to abandon hope and conclude that "you can prove anything with statistics."

The assignments grew more involved as the course progressed. In the paper in HIV testing, students were required to take a stand on a controversial issue (Kim Bergalis' compelling testimony before Congress was in the news at this time), and to assemble evidence to support their position. The final paper required them to bring all of the analysis ideas from the course to bear on a story that they found personally interesting. In addition to the statistics techniques, we had scheduled classes with the library on research techniques, including electronic database searching.

### Tom Moore's Writing Assignments

Here is a second set of assignments, for a different offering of the course (Moore, Grinnell, Fall 1993). This perhaps comes closest to the original conception of the course, which was to select five major topics for the syllabus. Here each topic gives rise to a writing assignment.

<i>Course Unit</i>	<i>Writing Assignment</i>
I. Public Opinion Polls	Find and critique the reporting of a poll by the press using the outline on p. 41 of Moore's <i>Concepts and Controversies</i> book.
II. Clinical Trials and Other Kinds of Studies	Find an article from the popular press that is about some new scientific study. Briefly describe the study in the new article. Discuss the extent to which the article tells you enough to assess the validity of the study, and generate a list of questions that would help you assess validity but are <i>not</i> answered by the article.
III. Coincidences	Describe a coincidence in your own life and why you think it is a coincidence and what the likelihood of such an occurrence would be.
IV. Data Analysis and Reporting Numerical Information	Find and critique a graph in the news.
V. Deming and Quality	Based on you brief tour of Company X (the class had gone on a field trip to a local area business), what would be Deming's three primary suggestions for improvement?

### **Additional Resources on Writing**

Connolly, P. and T. Vilardi. *Writing to Learn Mathematics and Science*. New York: Teachers College Press, 1989.

Iverson, Gudmund R. "Writing Papers in a Statistics Course." 1991 *Proceedings of the Section on Statistical Education* -- American Statistical Association (1991):29-32

Radke-Sharpe, Norean. "Writing as a Component of Statistics Education. *The American Statistician* 45(1991):292-293.

Sterrett, Andrew (ed.). *Using Writing to Teach Mathematics*. MAA Notes, Number 16.

Zinsser, William. *On Writing Well: An Informal Guide to Writing Nonfiction*, 3<sup>rd</sup> Edition. New York: Harper & Row, 1985.

## 6.7 Guest speakers

Sometimes it is helpful if a guest speaker can be arranged to visit the class and present information on their work in a particular area related to the current topic being studied. Speakers invited in the past have included:

- A meteorologist (or your local weather reporter) to discuss how the probability of rain is determined and what it means.
- A statistician from the Gallup organization to discuss some of the real problems involved in polling.
- A medical researcher who works on clinical trials.
- A biologist to discuss the DNA background for DNA fingerprinting.
- A psychologist to discuss statistical reasoning and intuition (e.g., Kahneman and Tversky theories about judgments under uncertainty).

It is obvious that many of the topics that come up in a Chance course use concepts that are outside the field of probability and statistics and so there an expert in the appropriate field can give valuable background information. It is important that you tell the speaker what your course is all about and the background of the students. Here is a success story.

In a previous class we carried out an activity by breaking the class up into groups, asking each group choose a member who claims to have extra sensory perception, and design and carry out a test using coin tossing to see if this is the case. The students were disappointed that no one succeeded in demonstrating this ability. We mentioned this to our guest lecturer from the Psychology department. He cheered up the class by showing the class as a whole did have extra sensory perception. He did this by putting five coins on the table in such a way that the students could not see the coins. He then asked each of them to think very hard and try to guess what the coins were and to write down the sequence of heads and tails that represented their guess. He then recorded all their answers on the board. The clear winner was HTHHT. He asked a student to come up and see what the concealed sequence was and indeed it was HTHHT. Of course this really was a lesson in Tversky's representativeness theory but it led nicely into a discussion of this theory.

You would be surprised how easy it is to get a colleague in another department to talk in your class. Of course, if it is an outside speaker you have to have some funds to pay travel expenses. It is nice, also, to offer a small honorarium, say \$200.

## 6.8 Chance Fair

The last day of class is usually reserved for a Chance Fair. Students are asked to bring a poster display of their project so that classmates may walk around, reading and finding out about the different projects. (This poster may be instead of or in addition to a written report.) Posters may be decorative and imaginative but they should make sure that they convey the main information about the project: what problem was investigated, how information was gathered, what was learned, etc.

In addition instructors may choose to bring in and play different games of Chance. The choice of games may depend on individual class members' or colleagues' expertise in running a roulette wheel or dealing blackjack as well as available resources. Typically, students receive play money to cash in for chips to use in playing the games and a prize is given to the student who makes the most money.

## 6.9 Assessment

Assessment is often regarded as testing and grading. However, a broader definition of assessment includes the gathering and utilization of information in order to improve student learning. Instructors vary in how they use and weight different requirements of the course in order to determine course grades, typically assigning a percentage of the grade to homework, group activities, and a project. Some instructors assign brief papers or quizzes that are also used to contribute to course grades.

### Assessment to inform the instructor of a Chance course

In addition to types of measures listed above, valuable information may be gathered about how well students are learning statistical ideas and how they are responding to the challenges of a Chance course. Two methods for obtaining this information are journal entries and minute papers. Journal entries may be used to have students discuss the articles they read outside of class, to explain a particular statistical idea, or to comment on their reactions to a particular class or activity. Minute papers differ from journal entries in that they are anonymous and are typically administered at the end of a class but submitted before students leave. Minute papers may ask students to explain their understanding of a concept, to describe any difficulties they are having in learning material, or providing feedback on their reactions to the course. Here is an example of a minute paper used to gather feedback from students during the middle portion of a Chance course:

#### **Minute Paper**

1. What do you like best about this class so far?
2. What do you like least about this class?
3. What would you like to see more of?
4. What would you like less of?
5. Any other comments you care to add:

### Assessment to determine the impact of a Chance course on students

Various instruments have been developed and used to assess student learning in Chance courses. These instruments include:

- A test of general reasoning about probability and statistics that may be used as a pre and post assessment of students' thinking and problem solving. This test also

includes 10 items that assess students attitudes and beliefs about statistics that may be administered as a separate instrument.

- A set of questions given to students to see how well they are able to read and 'critique newspaper articles, and a rubric used to evaluate these critiques.
- A rubric to use in scoring student projects, when projects are used as a method of assessing student learning
- A survey used to evaluate students' reactions to having taken a Chance courses.

Copies of all instruments are on the Chance Database and are also included in the appendix. For additional information on assessment, there is an article on this topic in the appendix.

### **6.10 Other resources to use in a Chance class**

There are a number of recent books that are very useful in teaching a Chance Course. Reviews for several books are included in the appendix as well as on the Chance database. They are: *Tainted Truth: the Manipulation of Fact in America* by Cynthia Crossen, *A Mathematician Reads the Newspaper* by John Allen Paulos, and *The Power of Logical Thinking* by Marilyn vos Savant, *Workshop Statistics: Discovery with Data* by Allan Rossman and *Activity-based statistics* by Schaeffer, Gnadiseken, Watkins, Witmer.

Of course, we use the leading newspapers as a source for daily articles to use in class. Particularly good papers for Chance News are the *New York Times*, *The Los Angeles Times*, *Boston Globe*, and the *Wall Street Journal*. Of course it is useful to use the college newspaper or local newspapers also since these will often be the newspapers that the students regularly read.

When science writers report on articles from journal such as the *New England Journal of Medicine*, they do a good job of given technical jargon ordinary names that can be understood. They also often interview other leading researchers in the area to try to get different points of views relating to the study they are reporting. For this reason, it is often useful to read these reports before reading the technical article. The journals that are particularly useful for a Chance Course are:

*Chance Magazine*, *Science*, *Nature*, *The New England Journal of Medicine*, *The Lancet*, *Journal of the American Medical Association (JAMA)* and *The Economist*.

## 7. Chance Guide to Topics

While our Chance courses have differed significantly in the way that they are taught, we have all found that the same basis set of statistical topics come up in the course. The next section is organized by statistical topics. For each topic, one or more articles are included along with questions to be used to generate class discussion. We have also included a video to show, a hands-on activity to use, and a journal entry to assign. Keep in mind that these articles, activities and journal entries only illustrate a particular combination of selections, from many possible choices, and may not be the selections you would choose to use.

For each topic we have used the following organization scheme:

1. Discussion of the topic in the typical CHANCE textbooks
2. One or more articles or abstracts along with discussion questions
3. Video or video segment
4. Hands-on activity
5. Journal entry

We have not included suggested homework problems for the textbook. There are many excellent exercises to select from in either text and we leave this to the discretion of the individual instructor.

### Texts

Those of us who have taught a Chance course have used a supplementary textbook and find this helpful. It is useful for students to have a resource where they can get a systematic treatment of statistical concepts at an elementary level. It is important that this book be elementary and very clear so that students with no formal background can read it on their own. It is useful if this book is non-mathematical and organized into chapters that may be understood if read in a non-linear order. Two books that have met these requirements are David Moore's *Statistics: Concepts and Controversies*, 4<sup>th</sup> edition, 1997, W.H. Freeman) [Moore] and Freedman, Pisani, and Purves *Statistics*, W.W. Norton) [FPP]. *Seeing Through Statistics* by Jessica Utts, Duxbury Press (1996) is a book especially written for a Chance type course and would also be an excellent choice. Unlike the other two books, this book itself is centered on the use of newspaper and other media to motivate the study of statistics.

The following comments are based on earlier editions of the books by Moore and FPP.

Moore is the more elementary of the two books that we have used. It contains 3 major parts: I "Collecting Data", II "Organizing Data", and III "Drawing Conclusions from Data". Part I has very lively and intuitive chapters on sampling, experiments, and measurement with lots of real life examples. Part II has good explanation of what are now standard descriptive and EDA statistics, including an excellent chapter on relationships including both qualitative variables (cross-classifications, controlling for extraneous variables), and quantitative variables (regression, correlation). All of this is done at the descriptive level with the emphasis being on what relationships mean. Part III treats probabilities using a frequentist interpretation, and views their computation as a job for simulations. This paves ground for a very intuitive final chapter on



statistical inference, with the emphasis on what a confidence interval is and what a statistical test is and their uses and abuses

FPP has much the same material as Moore but at a slightly deeper level. For example, FPP discusses residuals and the root-mean-square-error for the regression line while Moore does not. FPP does a lot more with probability than Moore. FPP discusses conditional probability, teaches computing probability by a few simple rules, and has a chapter on the binomial distribution. FPP then devotes 3 entire chapters to computing probabilities for sums (and averages) of draws (with replacement) from a box of numbered tickets--the so-called "box models." This discussion culminates in the central limit theorem. It is done in a highly conversational, non-technical way, free of mathematical proof or notation. Yet it clearly demands much more of the student than does Moore.

Both Moore and FPP contain excellent exercises. We recommend assigning some of these to the students on a regular basis to be counted in the course grade. Some of us assign these exercises on a daily basis to be handed in and graded like normal homework. Others have had the students keep these in a loose-leaf journal, to be self-graded and commented on by the students, and handed in 3 or 4 times during the semester. Since class discussion rarely revolves explicitly around these assignments we are still feeling our way around the notion of getting our students to understand this material.

What follows is an identification of the various topics with portions of these two textbooks that pertain to those topics. We have chosen to list statistical topics followed by Chance topics. In the list of Chance topics you should find most of the topics that at least one of us has taught.

NOTE: The articles included in the following topic areas were selected in 1996. Current articles on these topics may be used in their place.
--

## **Statistical topic: Surveys and sampling**

Chance topics: Public opinion polls and survey sampling;  
Census undercount

### ***Text:***

MOORE: The Introduction to Part I (pages 1 - 3) plus Chapter 1, sections 1 - 5 comprise the essential reading. This covers the need for good sampling, investigates basic concepts and terminology, without going beyond simple random samples. Moore gives lots of real life examples of good and bad surveys, and discusses concepts of bias, precision, and confidence interval. The remaining sections of Chapter 1 give the rest of the story including other kinds of probability sampling, as well as good discussions of policy and ethical issues. In Chapter 8, sections 1 and 2, Moore discusses confidence intervals in more detail and includes formulas for confidence intervals. To understand these sections it helps if the student knows the rudiments of the normal curve, which he/she can get from section 5 of chapter 4. It also helps to have read the probability material from Moore, however, neither of these prerequisites is absolutely necessary.

FPP: Chapter 19 has an excellent overview of survey samples with good historical examples of good and bad samples to motivate the need for proper design. For example the US Presidential election polls of 1936 and 1948 are discussed at length. This chapter covers the concept of chance error in samples but does not explicitly introduce confidence intervals. To get the full story the reader must read somewhat more technical chapters 20, 21, and 23. A hearty student could read sections 1 - 3 of chapter 21, ignoring the few technical references and thereby gain insight into the meaning of a confidence interval. To properly understand chapters 20, 21, and 23 it is necessary to have read and understood at least chapters 13, 16, and 17 on basic probability and on setting up and working with box models. Chapters 13, 16, and 17 in fact form the basic information for any Chance topic about probability, and as such it might be good to assign them early in the term.

### ***Article to discuss:***

Read the article "In a first, 2000 census is to use sampling", the editorial "No need to count every last person", and the letter to the Editor "Forget the census; let's have doughnuts".

### **In a First, 2000 Census Is to Use Sampling**

The New York Times  
February 29, 1996  
By STEVEN A. HOLMES

#### ***The article begins:***

To cut costs and improve accuracy, the Census Bureau said today that it would actually count only 90 percent of the United States population in 2000 and rely on statistical sampling methods to determine the number remaining...

## Discussion questions:

- 1) How could a census done using sampling be more accurate than a census done by actual enumeration?
- 2) As best as you can figure out, what is the Census Bureau really going to do?
- 3) The first New York Times article says that "Critics of the bureau's use of sampling argue that it is unconstitutional because the Constitution calls for an 'actual enumeration'. But decisions in lower Federal courts have approved the use of sampling as long as it supplements, and does not supplant, an actual count". Do you think the current plan is constitutional? If not, should the Constitution be changed?
- 4) If constitutionality is not an issue, do you think it would be a good idea to just take a 10% sample of the entire population? What are some the difficulties involved?
- 5) Would you expect that the figure of 44% marijuana smokers in the second editorial is pretty accurate, too low, or too high?

**Video:** Decisions through Data Unit 17 (on the Census undercount) and Chance Lecture by Tommy Wright.

**Activity:** Capture-Recapture with Goldfish Crackers

### Capture-Recapture

A common method of estimating the number of fish in a lake or pond is the capture recapture method. In this method,  $c$  fish are caught, tagged, and returned to the lake. Later on,  $r$  fish are caught and checked for tags. Say  $t$  of them have tags. The numbers  $c$ ,  $r$ , and  $t$  are used to estimate the fish population.

- 1) What is your estimate for the fish population in terms of  $c$ ,  $r$ , and  $t$ .  
(It may help to think about actual numbers first.)
- 2) What if some of the tags fall off your fish? Will your estimate be too big or too small?
- 3) Do you think that fish caught the first time are more likely to be caught the second time (or less likely to be caught)? If so, how would this affect your estimate?
- 4) What other assumptions do you need to make for your estimate to be reasonable?
- 5) The Census Bureau uses capture recapture to assess the number of people who were not counted by the Census ("the Census undercount"). Can you think how they might do this?

Use capture recapture to estimate the number of Pepperidge Farm goldfish in a package. So that we can compare results, we ask that everyone capture and tag 50 goldfish initially, and then recapture 40. Since physically marking or tagging goldfish renders them unappetizing, Laurie prefers to replace the captured goldfish with pizza flavored goldfish instead. Linda doesn't mind eating magic marker. You can use whatever method of tagging you like.

**Journal Entry:** Read the articles "High Court Rules Results are Valid in Census of 1990" or "Leadership by the numbers." Comment in your journal.

HIGH COURT RULES RESULTS ARE VALID IN CENSUS OF 1990  
The New York Times , March 21, 1996  
By LINDA GREENHOUSE

*The article begins:*

The Supreme Court today upheld the validity of the 1990 census, ruling unanimously that the Federal Government had no constitutional obligation to adjust the results to correct an acknowledged undercount in big cities and among minorities. While the decision, written by Chief Justice William H. Rehnquist, ends a long-running lawsuit, it almost certainly will not resolve a continuing policy debate over the best way to count the nation's population. New York City and a coalition of other big cities brought the lawsuit in 1988 to challenge the Bush Administration's refusal to use statistical sampling methods to adjust the 1990 census figures...

Leadership by the numbers: Having brought polling to new heights, will the Clinton administration reduce government to a new low?

*The Boston Globe*  
May 29, 1994  
By David Shribman, Globe Staff

*The article begins:*

The polls show that crime is the preeminent issue, and so all of Washington rushes to get tough on crime. The polls show that the public actually thinks President Clinton, the most domestic-oriented president since Calvin Coolidge, spends too much time on international affairs, so administration policies on Haiti, North Korean and the future of Russia and the Western alliance continue to drift. The result is that Washington has a lot of polls and is exercising very little leadership...

**Suggested readings.**

(1) The article "Pseudo-Opinion Polls: SLOP or Useful Data?" Dan Horvitz, Daniel Koshland, Donald Rubin, Albert Gollin, Tom Sawyer, and Judith Tanur, *Chance* magazine, p. 16, Spring 1995 (v.8, n.2). This article is a debate/discussion of the usefulness (or lack thereof) of self-selected polls (e.g., 900 phone polls) and their effect on the public. Some argue that they are not only worthless but in fact inflict major harm on the legitimate polling process. Others argue that they are a way to get at the opinion of those people who hold firm opinions on an issue. This debate is especially relevant to the discussion in the previous article. For example, the statement by Rep. Klink that, "I just want accurate polls" juxtaposed next to the quote, "... no polls are fully accurate" give us a tension point, but the debate in this issue of *Chance* sharpens that tension even more.

(2) The book *The Super Pollsters*, by David W. Moore (Four Walls Eight Windows, 1992, NY) has lots of interesting history of the political polling process and is good supplementation for the texts Moore and FPP. (Note: This is a different David Moore.) In particular, chapter 2 is about George Gallup and is great reading, showing how entrepreneurial Gallup really was. Pages 64 - 66 gives a description of quota sampling as it really happened, by one of the early interviewers hired by Gallup. The accounts in this book can nicely flesh out the main texts.

(3) A fine resource for a unit on survey sampling is the *Gallup Poll Monthly*. Two features of it are worth pointing out. First, each issue reports the results of several interesting recent polls. For example, the January 1995 issue discusses the public perception of President Clinton's job performance as well as the public's ranking of the nation's most pressing problems. There is also a report on a "heaven and hell" poll that shows the public's belief in both hell and the devil is clearly on the rise. The articles (of which there are usually about 5 or 6) are concise and supplemented with effective graphs and tables. There is also a regular feature called "Gallup Short Subjects" that reports in concise tables the results of questions that have not been reported elsewhere in the press. (Sort of the cutting room floor of the Gallup organization.) The same January 1995 issue reports on 59 such questions, one of which estimates that the average adult American watches 3.6 hours of television per day.

The second feature of interest is the technical notes at the beginning and end of each issue. Each issue the *Monthly* replicates the technical description of how the sample is drawn, how estimates are made and weighted, and how the reader may use and interpret margins of error. Included there is an explanation of how to decide when a difference between two independent percentages is significant.

**Statistical topic: Experiments and observational studies**

Chance topics: Clinical trials, experiments, and other studies

Since many discussions in our Chance courses have centered on media reports of some new scientific study this material is central to much of the course. For this reason it might be good to include this reading early in the course. These readings and discussions typically do not require statistical inference.

***Text:***

Moore: The Introduction to Part I (pages 1 - 3) plus Chapter 2 on experimentation is the essential material. This is full of good examples and is highly non-technical.

FPP: Chapters 1 and 2 on controlled experiments and observational studies provide all the basic information. The Salk polio vaccine field trial of 1954 is the central example, but FPP have found many excellent examples and throughout them illustrate the efficacy of properly designed experiments. The examples in FPP would certainly be good background to discussion of the article below, given how counter-intuitive the need to proper clinical trials is. It is good to counter that intuition before launching into a discuss of ethics like the article below invites.

***Article to Discuss:***

Read the *New York Times* article "A new AIDS drug yielding optimism as well as caution" by Lawrence K. Altman

A New AIDS Drug Yielding Optimism As Well as Caution  
The New York Times, February 2, 1996  
By LAWRENCE K. ALTMAN

***The article begins:***

An experimental drug has nearly halved both the death rate from advanced AIDS and the number of serious complications of the disease in a large international study that lasted seven months, the drug's developer reported at a scientific meeting here today. Many of the 2,100 participants enthusiastically greeted the report that the drug, ritonavir, prolonged the lives of some patients in the study...

**Discussion questions**

1. Why were participants divided into groups by lot? What are placebos and why were they used?
2. Why were participants required to meet the criteria given?  
Why were they allowed to continue present medications?  
Would the study have been better if they were not?

3. The article reports that there were 543 patients in the ritonavir and 537 in the placebo group. Using the percentages given in the article, find the number of patients in each group who died or suffered further progression of AIDS. How many patients in each group died?
4. If retonavir makes no difference, how many of the 72 patients who died would you expect in each group? If you toss a coin 72 times do you think there is a reasonable chance that you would get as many as 46 heads? What could you do to help you answer this question? What does this have to do with the study?

### **Journal Assignment:**

Read the *New York Times* article "3-drug therapy shows promise against AIDS" by Lawrence K. Altman. This study was carried out very differently than the study in today's discussion article. Comment on these differences. Which study was more convincing?

### **Activity: Taste test**

Students are asked to design an experiment to determine if someone can really identify Coke or Pepsi. Have on hand some large bottles of Coke and Pepsi and several small cups. Even if you don't go into the significance testing, you can still have a good discussion (both before and after the taste test) on issues related to designing a good experiment. There is a nice chapter in Crossen's book on taste tests that can be assigned as a supplementary reading.

An alternative activity begins with a class discussion around the question "Which cola do people prefer?". This is a different question than can they tell Pepsi from Coke, but it may be more universally relevant since a vast majority of the population appear to be happy to drink an occasional cola but never claim to be able to tell one apart from another. An experiment can then be designed to answer this question.

Another good experiment used in a Chance course was to answer the question: Does peanut butter preference get influenced by brand recognition? Small groups devised experiments to answer this. Then they went out and did them in the dorms.

**Video:** Unit 15: Designing Experiments in the DTD series, describes the aspirin study.

## Statistical topic: Measurement

Chance Topics: Measuring cholesterol, Rescaling the SATS

**Text:** Moore begins Chapter 3 with a definition of measurement and a discussion of validity. Measurement accuracy is addressed in terms of bias and reliability. The four scales of measurement are described. The last section gives tips on how to look at data intelligently, suggesting questions to ask when reading about research results.

FPP's approach to measurement differs from Moore's in that it concentrates more on the measurement of physical objects than "social" quantities. Thus FPP does not formally define the concepts of validity and reliability. FPP does a nice job in chapter 6 of discussing measurement error and the inherent variability in measuring a physical object. This discussion is carried on in chapter 24, which is a more technical chapter on inference and that depends on the central probability concept in the book: a box model.

One example that is common to both Moore and FPP is the measurement of unemployment. This example recurs in Moore's chapter 3 and garners a separate chapter of its own in FPP -- chapter 22 on "Measuring employment and unemployment". This chapter tells us about the Current Population Survey, on which the U.S. unemployment rate is based. FPP concentrate their discussion on sampling issues: the method of sampling (cluster sampling) and how this method affects the calculation of standard errors. The concepts of sample weighting and half-sample estimates of standard error are described. They discuss sample bias briefly (mentioning the Census undercount problem). Certainly concepts of reliability and validity are lurking about FPP's discussion, even if these terms are not formally defined.

### *Articles to discuss*

Students can be given either copies of these articles or the following abstracts to read:

#### **How Shall We Measure Our Nation's Diversity?**

*Chance Magazine*, Winter 1995, pp. 7-14.  
Suzann Evinger

#### **Black, White, and Shades of Gray (and Brown and Yellow).**

*Chance Magazine*, Winter 1995, 15-18.  
Margo Anderson and Stephen E. Fienberg

#### Abstract:

The Office of Management and Budget (OMB) provides the racial and ethnic categories used by federal agencies. These categories are used, for example, in the census, in civil rights enforcement, and in demographic studies. The current categories, established in 1977, are **American Indian or Alaskan Native, Asian or Pacific Islander, Black, White, Hispanic**. The OMB is reviewing these categories and a wide range of changes have been recommended to them including:

- Change "Black" to "African American" and "American Indian or Alaskan Native" to "Native American".



- Include "Native Hawaiians" as a separate category or as part of "Native American" rather than as part of the "Asian or Pacific Islander" category.
- Add a "Multiracial" category to the list of racial designations.

In addition to these specific suggestions, more general suggestions have been made including eliminating racial and ethnic categories altogether since they appear to have no real genetic significance.

These two articles discuss the many issues facing the OMB. Needless to say a revision of racial categories raises a number of interesting statistical issues such as: "Should race and ethnicity be classified by self-identification or by third-party identification?" and "Would important information be lost by introducing a multiracial category?".

- Anderson and Fienberg make a number of recommendations which include:
- Do away with the labels 'race' and 'ethnicity' and substitute something like 'identified population groups'.
- Allow people to identify with more than one group.

Regarding this last recommendation they remark: "We can always construct statistical rules for taking multiple responses and producing aggregate information on the categories".

### **Discussion questions**

1. The article continually refers to both racial and ethnic categories. Is there a difference? If so, what is it? If a questionnaire asks for both race and ethnicity, how would you respond?
2. How would you respond to the questions raised in the article: "Should race and ethnicity be classified by self-identification or by third-party identification?" and "Would important information be lost by introducing a multiracial category?"
3. What do you think of the recommendations to do away with the labels "race" and "ethnicity" and substitute something like "identified population groups".
4. Should people be allowed to report their ethnicity or racial group in more than one category? If this were allowed, what effect would it have on reporting numbers in different groups?
5. Should race or ethnicity be reported at all on standard forms? What are reasons for including or not including this information?

### **Additional Articles:**

If one wishes to include the important example of measuring unemployment in a Chance course we would recommend the article "Federal Agencies Introduce Redesigned Current Population Survey", by Thomas J. Plewes (*Chance* magazine, Winter 1994, Vol. 7, No. 1). Plewes describes the recent overhaul of the CPS, emphasizing the survey design changes that had to be made, many of which were the result of recent research in cognitive psychology.

If you don't want to assign the full Chance article you might use the article "Pollsters Enlist Psychologists in Quest for Unbiased Results", by Daniel Goleman (NY Times 7 September 1993) Of course, this article goes beyond the application of cognitive psychology to the CPS, but that is one area it discusses. When reading an article like Goleman's it sometimes piques the students' interest by giving them an "ask-ahead" questionnaire. For example, asking these questions could make it clear to them, when their results are as predicted in the article, that the effects discussed by the article are real.

Sample Discussion questions for Goleman article:

1. If you are told that something will happen within a few years, how many years do you think they mean?
2. If someone tells you that he/she has been laid off from a job which of the following meanings are you most likely to infer?
  - a. He/she has been fired.
  - b. He/she has been temporarily suspended from a job that he/she expects to be called back to.
  - c. Something else. (Please explain.)

**Video:** Unit 14 DTD, Save the Bay. This video isn't explicitly on measurement but it deals with different types of measurements involved in determining the "health" of the Chesapeake Bay.

**Activity:** Surveying the Class

As an alternative to giving students an existing survey to complete at the beginning of class, some instructor prefer to have students design their own class survey. Ask the students to get into groups and think about good questions that they would really like to find out about students in the class. Each group needs to come up with four good questions, representing 3 different levels of measurement. Groups may also be asked to discuss ways to make the survey a valid and reliable instrument. All items are submitted to the teacher who chooses the best ones and adds a few, constructing a class survey of about 25 items. Students later fill out the survey and the data are used in computer labs and demos during the data analysis unit.

**Journal Entry:**

Read and comment on measurement issues raised on either of these two recent article, each relating to important but quite different measurement tasks: the measurement of intelligence a test for the AIDS virus.

Intelligence: knowns and unknowns.  
*American Psychologist*, 51(2),77-101  
By Ulric Neiser et. al.

*Review from Chance news:*

In the Fall of 1994, the book "The Bell Curve" by Herrnstein and Murray discussed the concept of intelligence, its measurement, and its relation to human behavior and political decisions. This book led to heated discussions in the press. Believing that when science and politics are mixed, scientific studies tend to be evaluated in relation to their political implications rather than to scientific merit, the "American Psychological Association" appointed a task force to provide an authoritative report on the present state of knowledge on intelligence.

While the report was inspired by "The Bell Curve", the authors make no attempt to analyze the arguments in this book. Rather they follow their charge: "to prepare a dispassionate survey of the state of the art: to make clear what has been scientifically established, what is presently in dispute, and what is still unknown." While less lively reading than the "The Bell Curve" and its critics, the report is well written with a minimum of jargon. It provides an admirably balanced view of the present state of knowledge on the issues that it addresses:

What are the significant conceptualizations of intelligence at this time?

What do intelligence test scores mean, what do they predict, and how well do they predict it?

Why do individuals differ in intelligence and especially in their scores on intelligence tests? In particular, what are the roles of genetic and environmental factors?

Do various ethnic groups display different patterns of performance on intelligence tests, and, if so, what might explain these differences?

What significant scientific issues are presently unresolved?

New Test Predicts Progress of AIDS Virus.  
The New York Times, 31 Jan. 1996  
By Lawrence K. Altman

*The article begins:*

Researchers from the University of Pittsburgh announced that a new test, called branched DNA, measures the amount of AIDS virus in blood and predicts the progression of infection to disease much sooner and more accurately than the standard test. In addition, the new test gives a better indication of a patient's chance of survival for five years and can be used to establish a system of stages of infection with H.I.V...

## Statistical topic: Distributions and Measures of Center

Chance topics: Grade inflation, SAT scores

### *Text:*

Moore Chapters 4, sections 1-3, gives many examples of ways to graphically display data and calculate measures of center.

FPP devotes chapter 3 to the histogram and includes a discussion of types of variables and what it means to control for a variable in this chapter. Warning: FPP defines histograms using the density scale, a level of sophistication unnecessary for many discussions, but the "right way" to talk about histograms if you want one definition of histogram to cover all uses we make of it. For data histograms of one variable the density scale is overkill and for superimposing two variables a relative frequency y-scale would suffice. But when fitting, say, a theoretical normal curve to some data or some simulation results the density scale makes it work. Once you get past this idiosyncrasy of FPP, chapter 3 contains many good intuitive discussions on interpreting histograms. Chapter 4, sections 1-3 talk about the mean and about its relationship to the histogram.

### *Articles to discuss:*

Mean statistics: when is the average best?  
Washington Post, 6 Dec. 1995, by John Schwartz

*From Chance News:*

Schwartz remarks that politicians and others often choose a definition of average that best suits their needs. He tells his readers what mean, median, and mode mean and gives examples of their use and misuse. He starts with the example of John Cannell, who notices that his state's school system claimed high scores on nationally standardized tests and requested test scores from all 50 states. Cannell found that every one claimed to be "above the national average" or the statistical "norm". He called this as the "Wobegan effect". A more detailed discussion of this example can be found in the article

Taking the tests.  
Dallas Morning News, 4 Oct. 1994, by Karel Holloway.

As another example, Schwartz remarks that if Bill Gates were to move to a town with 10,000 penniless people the average (mean) income would be more than a million and might suggest that the town is full of millionaires.

### **Discussion questions:**

(1) How could the answers Cannell received be correct?

(2) Someone once claimed that if any one person moved from state X to state Y the average intelligence in both states would be increased. How could this be? Can you think of an X and a Y that might make this statement true?

**Activity:** Counting raisins.

Bring bags of two different brands (e.g., Sun Maid and Dole) of little snack boxes of raisins. Hold up one box and ask students how many raisins they think are in one box. Then pass out the boxes of raisins so that some students get one brand and some get the other. Ask students to write down their estimates of how many raisins are in a box, then open the lid, look inside and revise their estimates. Then they are allowed to count (and then eat) their raisins. Students mark their results on plots on the board, (there will be a plot for their first estimates, revised estimates, and then actual number of raisins, separating the two brands). Students are asked to look at these plots, and describe the distributions and compare the number of raisins for the two different brands.

Another activity has students looking at graphs and descriptive statistics for class survey data and interpreting them.

**Article to read and critique for journals:**

Cheat sheets: Colleges inflate SAT's and graduation  
rates in popular guidebooks.  
*The Wall Street Journal*, 5 April 1995  
By Steve Stecklow

*Review from Chance News:*

This article documents how colleges cheat on reporting SAT scores and graduation rates. For college guides, they want the averages high. In Money magazine's 1994 college guide, New College of the University of South Florida was ranked number 1. New College had been cutting off the bottom-scoring 6% for years, increasing the average score about 40% for "marketing purposes". The most common cheat is to exclude certain groups of low scoring groups when reporting SAT scores.

Boston University excludes the verbal SAT scores, but not the math scores, of international students. Of course the colleges can give arguments for these practices, but such adjustments are specifically prohibited by the guide books. As a way to gauge the extent of this kind of cheating, the newspaper compared the numbers reported to debt-rating agencies and investors with those given to guides such as U.S. News guide, finding more than two dozen discrepancies in the enrollment data.

Similar games are played with acceptance rates and graduation rates. Acceptance rates can appear higher by not counting the waiting list. Schools want the graduation rates high for the College Guides and low for the NCAA so their student athletes will look good. This survey found that a number of schools were able to accomplish this.

Suggested Reading: The median isn't the message by Stephen J. Gould

**Video:** Histograms (DTD, Unit 3) or Measures of Center (DTD, Unit 4).

## Statistical Topic: Variability and the normal curve

**Text:** Moore Chapter 4, sections 4 and 5, covers the measures of variability and introduces the normal distribution and its characteristics. FPP devotes the last part of chapter 4 to the standard deviation and its interpretation as the likely size of a deviation from the average and also gives us the "68%-95% rule" for lists of numbers (before any mention of the normal curve, since FPP's point is that this rule is more robust than one might think). Then chapter 5 is all about the normal curve. Chapter 6, on measurement error, completes the discussion of standard deviation by giving the final interpretation of standard deviation as the "likely size of a chance error", if the measurement process is unbiased.

### *Article to discuss in class:*

Men at extreme ranges of IQ tests, study says.  
Sacramento Bee, 7 July, 1995

#### *Review from Chance News:*

The article reports that a new study, by Larry Hedges and Amy Nowell of the University of Chicago, has found that the average man and average woman share about the same level of intelligence, but men account for a higher proportion of both geniuses and the mentally deficient. The report analyzed six large national surveys of American male and female teenagers' performance on tests of mental ability, conducted over the past thirty years. Their results were presented in *Science* 7 July 1995.

Seven of every eight people in the top 1% of IQ tests are men, but men also represent an almost equally large percentage of the mentally disadvantaged. Neuroscientist Richard Haier of the University of California, Irvine, says that the findings of a higher percent of men in the top IQ levels is nothing new, but what is new is that "there were more males in the low end."

The article mentions that, while men and women differ in brain sizes and that male and female brains function differently, such physiological differences do not account for the differences in the abilities of the sexes. The study sheds little light on the origin of sex difference in aptitude.

NOTE: In the *Science* article the authors stress that it is important to analyze representative samples instead of samples selected from talent searches etc. While they analyze a number of studies, their main conclusions seem to be based on their analysis of the National Assessment of Educational Progress program which periodically tested large samples (70,000 to 100,000 students) in the areas of reading, mathematics, science, and writing. They found that in all four areas the men had higher variances than the women, typically of the order 3 to 15%. Men had higher average scores in mathematics and science and the women in reading and writing. They suggest that both the small number of women in the top 10% of math and science and the high number of men in the bottom 10% in writing and reading have policy implications. Hedges and Nowell suggest that intensive recruiting will be necessary to achieve a fair representation of women in science and that men will have difficulty finding employment in an increasingly information driven economy.

**Discussion questions:**

(1) In their article, the authors state that the difference in the means between men and women is relatively small while the difference in variances is quite large. They say that this may help explain the apparent contradiction between the high ratios of males to females in some highly talented samples and the generally small mean differences between the sexes in unselected samples. Could this help explain the fact that 60% of the National Merit Scholarships went to boys this year.

(2) For the sex difference in means, the authors used the standard mean difference-- the difference of the means divided by the standard deviation of the total population. Why did they not just use the difference in the means?

(3) The article states that men account for a higher proportion of both geniuses and the mentally deficient. The data in the *Science* article shows that for verbal skills there were a higher proportion of women in the top 10% and men in the bottom 10%. For mathematics skills there were a higher proportion of men in the top 10% and a higher proportion of women in the bottom 10%. Do you think the article's account describes this situation reasonably?

**Video:** DTD Unit 7: Normal Curves

**Activity:** Look at class survey data, try to determine if any variables have a normal distribution. Interpret measures of variability.

## Statistical Topic: Correlation and Causation

Chance Topics: Smoking

**Text:** Moore, Chapter 5, discusses relationships between two variables. Section 1 introduces cross-classified data in two-way tables. Section 2 describes scatterplots and correlation analysis. Section 3 distinguishes between association and causation, and the last section is on prediction using regression analysis. Moore has an excellent subsection to section 3 called "Establishing Causation", where he gives 3 guidelines for how to establish causation when a randomized experiment is impossible, such as in the smoking and lung cancer problem.

FPP discusses correlation in Chapters 8 and 9 and regression in Chapters 10, 11, and 12. The principal of "correlation is not causation" is firmly ingrained by Chapter 9. The regression chapters are descriptive treatments, but do get into technicalities to the point of interpreting the slope of the regression line as the ratio of the sd for  $y$  to the standard deviation for  $x$  multiplied by the correlation coefficient. The emphasis in these chapters is on understanding of what the concepts mean. This is consistent with the FPP spirit. FPP contains a fine exposition of the regression effect and the regression fallacy, including Galton's famous father-son height data and his discovery of "regression toward the mean".

Of course, the "correlation is not causation" mantra is first encountered in Chapter 2 of FPP, where the subject is observational studies and confounding and includes the famous example of sex bias and graduate school admissions at Berkeley.

**Articles to discuss:**

### **Embattled tobacco: one maker's struggle 1, 2, and 3.**

*The New York Times*, 16, 17, 18 June, 1994  
Philip J. Hilts

Based on documents from the archives of the Brown & Williamson Tobacco Corporation, this three-part series describes how the major American tobacco companies chose to publicly downplay or even deny the results of studies, many conducted by their own research departments, on the health risks and physiologically addictive properties of smoking cigarettes.

For example, the second article states that after an internal study from 1963 found evidence of significant physiological effects from smoking, considerable effort was put toward creating what the company called "a device for the controlled administration of nicotine"--- a "safer" cigarette with equal (or even greater) nicotine levels. A top researcher for Brown & Williamson's sister company, British-American Tobacco (Batco), described Batco in 1967 as being "in the nicotine rather than the tobacco industry." But when questioned in April of this year by the House Subcommittee on Health and the environment, each of the seven top executives in the American tobacco industry stated that nicotine is not addictive and that cigarettes may not cause cancer.

Meanwhile, on 17 June, 1994 a spokesman for Brown & Williamson is quoted as saying: "The tobacco industry was and is just as interested in research on smoking and health as those outside the industry. Our position continues to be that there are health risks statistically



associated with smoking, but there is no conclusive evidence of a causal link between tobacco use and disease."

The third article includes a "Chronology of Concern" that lists some of the results of major studies on tobacco and disease since the 1950's, along with actions taken by tobacco companies and government agencies.

### **Discussion questions:**

1. Referring to the statement by the industry spokesman, above, what do you think is meant by the phrases "statistically associated", "causal link", and "conclusive evidence"?
2. In testimony before congress, Andrew W. Tisch, chairman of the Lorillard Tobacco Co., stated that "We have looked at the data and the data we have been able to see has all been statistical data that has not convinced me that smoking causes death." What do you think of this remark?

**Video:** The Question of Causation, DTD Unit 16.

**Activity:** Cookie rating. This may be a one or two day activity, where one day students design the study and the next day, carry out the study.

### **Cookie Rating Activity**

Students are asked to devise an experiment to answer these question: Do more expensive cookies taste better? Is there a correlation between price and taste? Between taste and appearance? Students may discuss this problem in groups and come up with a plan that is carried out at the next class. One example of a cookie rating activity is on the Chance database .

### **Article to read and critique for journals:**

#### **Were you misled?**

*The Boston Globe*, 3 July 1994, p7.  
Paid advertisement, Philip Morris USA

This advertisement reprints an article from Forbes Media Critic entitled "Passive Reporting on Passive Smoke", by Jacob Sullum, the managing editor of Reason magazine. (This is clearly related, though not identical, to Sullum's recent National Review article; see CHANCE News, June 10, 1994) The central theme is that the media have been totally uncritical in their acceptance of the EPA's claims of evidence linking environmental tobacco smoke (ETS) to cancer. For example, the EPA examined 30 epidemiological studies looking for a link. While most of the studies found a positive association, this association was statistically significant in only six. Why, Sullum asks, has the issue of significance been ignored in press reports?

## Statistical topic: Time Series

Chance Topic: Sunscreen and Skin Cancer,  
Consumer Price Index and GDP

**Text:** Moore Chapter 6, begins by introducing the idea of an index, and then describing the Consumer Price Index. Different types of social and economic indicators are also described. The final section discusses time series and the interpretation of times series data.

FPP doesn't discuss time series or index numbers. The closest it comes in Chapter 24, where they discuss the importance of the box model assumptions in applying many classical inferential techniques, and they give examples of time series data showing trends or cycles, belying the independence between values in the list.

**Video :** Against All Odds: Time Series (Tape 6).

**Article to discuss:** Lies, Dammed Lies and Statistics, from *Chance Magazine* (by Shankar). This article shows how yearly data on SAT scores and money spent on education were misinterpreted.

### Mouse Study Raises Doubts on Sunscreens

by Gina Kolata  
*The New York Times*  
January 25, 1994

A new study using mice has raised questions about whether sunscreens can protect against melanoma, the deadliest of all skin cancers. At the same time, experts are asking what it is about sun exposure that increases the chances that people will get melanoma and are re-examining strategies for protection....The number of Americans diagnosed with melanoma, a cancer of the melanocytes, or pigmented cells of the skin, has increased steadily for decades...

#### Discussion questions:

1. For the three types of skin cancer discussed in this article -- basal, squamous, and melanoma -- discuss to what degree the case for causation meets each of the 3 guidelines on "establishing causation" in Moore's book (pages 276 - 277).
2. The article claims that easily tanned people are at lower risk, but fair-skinned people are not a higher risk. How can this be? Isn't this like Garrison Keillor's "all the children are above average?"

\*\*\*Note: This article may be supplemented with some time series graphics on melanoma from Cleveland's book, *Visualizing Data*.

***Journal Entry***

Read and comment on this article in relation to the relevant sections in chapter 6 of Moore's text.

**U.S. is Considering a Large Overhaul of Economic Data**

*The New York Times*, 16 Jan 1995, A1

Robert D. Hershey Jr.

Federal statisticians are fixing deficiencies in systems used to gauge nation's economic performance. The current system fails to capture the array of new technologies and structural changes in today's \$7 trillion economy.

Alan Greenspan, chairman of the Federal Reserve, has said that the Consumer Price Index overstates inflation by a maximum of 1.5 percentage points. With Social Security and other entitlements (as well as adjustments in tax brackets) pegged to the inflation rate, Republicans in Congress were quick to seize on the corollary that by changing the index's formula, they could free \$150 billion in Federal funds over the next five years without a single specific budget cut...

## **Statistical topic: Probability**

Chance Topics: Paradoxes, coincidences, gambling,  
DNA fingerprinting, streaks and runs (e.g., in sports),  
Card shuffling, Lotteries, Weather Forecasts.

**Text:** Moore has a very basic approach to probability in Chapter 7. This reading will provide the student with the concept of long-range relative frequency, the law of averages, equally likely models, simulation, and expected value. This is a spare tool kit, so the instructor should be prepared to supplement as needed.

FPP: Chapters 13 and 14 introduce the bare essentials needed to discuss probability topics. Chapter 13 includes the long-range frequency definition, the definition of conditional probability and the product rule for a sequence of 2 (possibly dependent) events, independence and the implications of independence, and the famous Collins case. Chapter 14 adds the concept of equally likely models and the addition rule for mutually exclusive events to compute probabilities. These two chapters are really sufficient to handle any of the examples topics listed above except for gambling, which requires the notion of expected value. To gain this concept one must read chapters 16 and 17 which tell the reader how to compute probabilities for sums (or averages) of draws from a box by using the normal approximation (i.e., without naming it, the Central Limit Theorem). Chapter 17 includes an excellent discussion of roulette.

**Video:** What is Probability, from Against All Odds (Tape 15): excerpt showing Diaconis tossing coins and getting heads each time.

**Activity 1:** Three approaches to estimating the probability of heads

What is the probability of getting heads when you toss a coin? When you spin a coin? When you stand coins on edge and hit the table?

Students divide into three groups, each group is given some coins and conducts 100 trials of one these three methods to estimate the probabilities of getting heads. Results are compared (and are quite different).

**For journal:** Pick one other article on DNA from the Chance data base to read and critique.

There is room for many articles and activities in this unit. Here are a few more:

**Activity 2:**

### **Let's Make a Deal**

Suppose you're on a game show (Let's Make a Deal) and you're given the choice of three doors to choose from. There's a big prize behind one (a car) and behind the other two: goats. You select a door, say Door 1, and the host (Monty Hall, who knows what's behind all the doors) opens a different door, say number 3, which reveals a goat.

Monty then asks you: "Do you want to open the door you picked (#1) or open door 2? What would you do?"

1. Discuss in your group whether you should switch or stay with your first door, or whether it makes a difference. What assumptions does your answer depend on? What are other assumptions and how would they affect your answer?
2. Carry out a simulation of the game show.
  - a. Monty Hall controls the three doors (now index cards). On the back of each is either a car or a goat (on two). Monty will lay out the three cards, blank side up making sure he/she knows which has the car on the other side.
  - b. You will begin to take turns playing the game, and as you do, keep a record sheet, listing your strategy (stay or switch) and the outcome (car or goat).
3. Which strategy is better? Count the number of times the strategy was to switch, and of those, how many times you won the car. Write this as a fraction and convert to a decimal:

#of wins using the switch strategy/ #of times you switched doors.

Do the same thing for the "stay" strategy. Fill in and compare these two decimals. They represent:

The probability of winning when you switch doors\_\_\_\_\_

The probability of winning when you keep the same door\_\_\_\_\_

Which strategy has the better chance of winning the car?

4. Pool data with the class to get better estimates of these probabilities.

Now which strategy seems to be the best to use?

Next, have students read and discuss this article:

**Behind Monty Hall's Doors: Puzzle, Debate and Answer?**

July 21, 1991, Sunday, Late Edition - Final

BYLINE: By JOHN TIERNEY, Special to *The New York Times*

*Article to read and discuss:*

### **In Shuffling Cards, 7 Is Winning Number**

By GINA KOLATA

*The article begins:*

It takes just seven ordinary, imperfect shuffles to mix a deck of cards thoroughly, researchers have found. Fewer are not enough and more do not significantly improve the mixing.

The mathematical proof, discovered after studies of results from elaborate computer calculations and careful observation of card games, confirms the intuition of many gamblers, bridge enthusiasts and casual players that most shuffling is inadequate...

The following is an excerpt from Chance News:

Ask Marilyn.  
Parade Magazine, 20 Nov 1994  
Marilyn vos Savant

When I play in card games I like to shuffle the cards several times, but I'm often told that if I shuffle them too much they'll be returned to their original order. What are the odds of this happening with five to 10 shuffles?

Teri Hitt, Irving, Tex.

In her answer Marilyn considers both the case where Hitt means a perfect shuffle and an imperfect shuffle. For a perfect shuffle she correctly states the following relevant result:

Some magicians are so deft with their hands that they can shuffle the cards "perfectly," meaning a shuffle in which the deck is exactly halved, and every single card is interwoven back and forth. If you do this eight times, the cards will be returned to their original position.

About imperfect shuffles she says:

A study shows that with ordinary imperfect shuffles, you need at least seven to make sure that the cards are randomly mixed. Six aren't quite enough, but eight aren't a significant improvement--although the mixing does improve with each shuffle.

If we interpret Hitt's question in terms of imperfect shuffles then it is natural to consider the model of a binomial cut followed by a riffle shuffle introduced and discussions of their work by Charles Grinstead and Brad Mann can be found in teaching aids on the chance data base.

This analysis provides a simple combinatorial expression for the probability that the deck will be back to the original order after  $k$  shuffles. Applying this result to 5 shuffles gives a probability of  $3.21097 \times 10^{-56}$  so this is probably not what Hitt had in mind. Marilyn assumes that if

she meant imperfect shuffles she was more interested in being sure that they are well mixed and suggests that seven shuffles suffices.

The shuffling article in the *New York Times* states that "...seven shuffles is a transition point, the first time that randomness is close. Additional shuffles do not appreciably alter things...." This suggests the question: Close enough for what? In other words, the deck is not fully randomized after seven shuffles, there should be ways to exploit this. The following activity explores this idea.

### **Activity 3:** New Age Solitaire

This fascinating game was introduced by Peter Doyle as a way of bringing home the fact that 7 ordinary riffle shuffles, followed by a cut, of a 52-card deck are not enough to make every permutation equally likely. Here is John Finn's description of the game.

We start with a brand new deck of cards, which in America are ordered so that if we put the deck face-down on the table, we have Ace through King of Hearts, Ace through King of Clubs, King through Ace of Diamonds, King through Ace of Spades. Hearts and Clubs are thus the High suits, and Diamonds and Spades the Low. Some would term these Yin and Yang, but not according to any scheme that I believe would satisfy Georges Osawa, who says that tomatoes and eggplants are both extremely yin because of their purple color.)

We shuffle the deck of cards 7 times, then cut it, and then start removing and revealing each card from the top of the deck, making a new pile of them face-up (so if this were all we did, we'd just have the deck unchanged after going through it once, except that the deck would be lying face-up on the table).

We start the pile for each suit when we discover its ace, and add cards of the same suit to each of these 4 piles, according to the rule that we must add the cards of each suit in order.

Thus a single pass through the deck is not going to accomplish much in the way of completing the 4 piles, so having made this pass, we turn the remaining deck back over, and make another pass. We continue this until we complete either the two high piles (hearts & clubs), or the two low piles (diamonds & spades). If the high piles get completed first, we call the game a win; it's a loss if the low piles get completed first.

If the deck has been thoroughly permuted (by having put the cards through a clothes dryer, say), then the lows and highs will be equally likely to be first to get completed. Thus our expected proportion of wins will be 1/2.

But it turns out that after 7 shuffles and a cut, we are significantly more likely to complete the highs before the lows, so our proportion of wins will be greater than 1/2. In fact, when you try this with your students you can expect that Yin will win over 70% of the time. You will find programs to simulate this game on the Chance web site.

## Statistical topic: Statistical Inference

Chance topics: Statistics in the law;  
Advanced articles on clinical trials, experiments and surveys

**Text:** Moore introduces Confidence Intervals in Chapter 1, his chapter on sampling. The exposition is non-technical and excellent. Section 3 introduces sampling variability and sampling distribution, bias, and precision. Section 4 introduces margin of error and confidence statements. No formulas are developed, but Moore provides a table of margins of error for Gallup-type polls indexed by sample size and  $p$ . Moore then revisits confidence intervals in chapter 8, a chapter on inference that follows the chapter on probability. Section 2 is more technical, giving formulas for confidence intervals of any desired level of confidence for a population proportion under simple random sampling.

Two sections give the essence of Significance Testing: sections 3 and 5 of Chapter 8. (Section 4 is an optional and more technical section.) It would be advantageous to have read Moore's probability material, but it is not absolutely necessary. Section 3 introduces tests, P-value, and statistical significance. Section 5 discusses "uses and abuses" of these concepts.

FPP: The essentials of confidence intervals are in Chapters 19, 20, 21, and 23. Chapter 19 is a stand-alone chapter introducing sampling, and although it does introduce sampling variability (i.e., chance error) it does not introduce the language of confidence or margin of error. To get this full story the other chapters need to be read and a full appreciation of them requires an understanding of chapters 16 and 17 on the box model.

The core material from FPP for significance testing is sections 1 -- 5 of Chapter 26 and all of Chapter 29. In Chapter 26 we learn what a statistical test is, about P-value, and about statistical significance. It really helps to know how sample means vary (Chapters 16 and 17) when reading Chapter 26. Chapter 29 includes the important "uses and abuses" discussion.

This unit is divided into two parts: Confidence Intervals and Hypothesis Tests



## Confidence Intervals

### *Article to discuss*

Study Finds More Drug Use  
but Less Concern About It  
The New York Times , July 21, 1994,  
By JOSEPH B. TREASTER

*The article begins:*

A 13-year decline in illicit drug use has halted as Americans are becoming less concerned about the hazards of drugs, Federal researchers said yesterday, suggesting to some experts that the country's drug problem might be on the brink of worsening.

In an annual survey of households across the country, researchers from the Department of Health and Human Services said there were already indications of increased drug use in two age groups: teen-agers and those 35 and older...

### **Discussion questions**

1. How did the authors arrive at the margin of error of less than plus or minus one percentage point?
2. Assuming the percentages given from the graph, is the increase in 1993 over 1992 significant?
3. Can you conclude that the increases for subgroups of people over 35 and for teenagers are significant even if the overall drug rate has not increased significantly?

The following letter from Mark Kernighan expressed his opinions on this study.

#### **Has drug-use study really told us anything?**

*The New York Times*, 5 Aug 1994, A24

Mark Kernighan (letter to the editor)

A Federal study has shown that the number of Americans who use illicit drugs rose from 1992 to 1993(news article, July 21). More precisely, the number of people in a sample of 26,000 who were willing to admit their drug habits to Federal researchers in door-to-door, face-to-face, non anonymous interviews grew by about 35, or 0.13 percent of the sample.

The Department of Health and Human Services researchers seemed to see no difference between drug users and drug users who are willing to chat about it.

The rise is not even close to statistically significant. Even if the drug rate were stable, 15 percent to 20 percent of all surveys would show such an increase just by chance.

The Government's offhand disclaimer that the number of heavy cocaine users "may be four times as high because the survey does not reach many of the heaviest drug-using groups" makes it clear that the vaunted 1 percent margin of error of the researchers is as empty as the wildest guess. The estimate is almost certainly too low, because the interviewers are overlooking the heaviest drug users, liars and subjects who don't answer the doorbell.

Whatever the survey measures, it's not drug use. "If you can't prove what you want to prove, demonstrate something else and pretend that they are the same thing," Darrell Huff wrote in his classic "How to Lie with Statistics."

### **Discussion questions:**

1. How would Mark justify his statement: "Even if the drug rate were stable, 15 percent to 20 percent of all surveys would show such an increase just by chance."
2. Do you agree with Mark that the survey is not really measuring drug use?
3. Do you agree with Mark that it is misleading to claim a margin of error of 1 percent in a survey of this kind?

**Video** : DTD video on confidence intervals (Unit 20)

**Activity:** Have students estimate the proportion of yellow m&m's (using 3 different populations: plain, peanut, and peanut butter m&m's). First students estimate the proportion of yellows in each of the three populations. Everyone will take a sample of 20 and calculate the proportion of yellows. We generate three different sampling distributions to demonstrate and describe the variability of the sample proportion. Each student will compute 95% confidence intervals for the population portion of yellow m&m's for each population based on their own samples. We will compare the confidence interval to the true % of yellow m&m's in each population.

### **Hypothesis Testing**

*Articles to read and discuss:*

Many find 'remote viewing' a far-fetched science.  
Washington Post, 2 December 1995  
Curt Supplee

*Review from Chance News:*

Federal intelligence agencies spent \$20 million dollars in the last two decades studying and trying to exploit a psychic ability called "remote viewing". These initiatives began in the early 1970's, because of a concern that the Soviet Union was making substantial progress in understanding extrasensory perception. People spoke of the "ESP" gap.

The CIA inherited this program and asked statistician Jessica Utts and Psychologist Raymond Hyman to evaluate the results of the program. Professor Utts has been looking at results in parapsychology for the past ten years as part of her research program. She is known to support the possibility of extra-sensory perception. Hyman is a psychologist and well-known skeptic who has consulted many times for government panels and others regarding the possibility of extra-sensory perception.

The program, called Stargate, had three parts: (1) keeping tabs on what other countries were doing, (2) actually using psychics to see if they could add anything to ordinary intelligence, and (3) carrying out research to see if there were psychics who were effective at remote viewing.

Most of the research was carried out by two California contractors, SRI International and Science Applications International Corp. They carried out thousands of tests, with hundreds of subjects resulting in the identification of a couple of dozen psychics. In a typical experiment an assistant in another room would hold up a randomly selected photo from a set of five different pictures, each in a sealed envelope. The subject would be asked to describe what was in the picture.

The subject would produce verbal impressions, hand drawings or both. A judge would rank each photo from 1 (best) to 5 (worst) according to how closely each corresponded to the psychic's description. Statistician Jessica Utts examined years of data and found that a substantial number of the tests turned up average ranks around 2.3, about 14 percent better than chance. She concluded that the results of these tests could not be accounted for by chance and that remote sensing was a real phenomenon. Hyman remains skeptical being concerned, about methodological problems that need further investigation and problems related to being able to replicate the results in these experiments.

The CIA has decided to drop the program in favor of watchful waiting. Utts and Hyman wrote individual reports and you can find them both on Professor Utt's homepage.

They are both very well written and describe many of the basic issues involved in using statistics to establish a scientific claim.

### **Discussion questions:**

- (1) How do you think Professor Utts got the 14% chance for getting an average ranking of 2.3 or lower?
- (2) Hyman claims that establishing statistically significant results related to ESP is a far cry from showing that there is such a thing. What does he mean by this?
- (3) What do you think are some of the methodological concerns one might have about experiments of the kind described above?

### ***For journals:***

Read and review Professor Hyman's report on psychic functioning and Professor Utts' reply to his report, both on the internet.

**Video:** on Tests of Significance DTD, Unit 21 (shows early tests on aspartame in Cola)

***Activity:***

Ask students if anyone thinks they have ESP. Take out a deck of cards, and discuss the probability of someone correctly guessing the suit of a card. Have students design experiments where pairs of students will carry out a test of ESP (where one looks at a card and the other tries to correctly guess its suit).

Discuss how large a number of correct guesses we would expect to get (out of 10 trials) in order to believe that the person has ESP. Discuss the experiment in terms of types of decisions and errors. Then carry out a test and have students calculate P-values.

# Appendix

## 8.1 Teaching Statistics Using Small-Group Cooperative Learning

Joan Garfield  
University of Minnesota

*Journal of Statistics Education* v.1, n.1 (1993)

Copyright (c) 1993 by Joan Garfield, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the author and advance notification of the editor.

KEY WORDS: Active learning; Small groups.

### Abstract

Current recommendations for reforming statistics education include the use of cooperative learning activities as a form of active learning to supplement or replace traditional lectures. This paper describes the use of cooperative learning activities in teaching and learning statistics. Different ways of using cooperative learning activities are described along with reasons for implementing this type of instructional method. Characteristics of good activities and guidelines for the use of groups and evaluation of group products are suggested.

### 1. WHAT IS COOPERATIVE LEARNING?

Cooperative learning is a topic frequently mentioned in conversations about improving education, regardless of the discipline or level of instruction. Some recent definitions of cooperative learning include:

An activity involving a small group of learners who work together as a team to solve a problem, complete a task, or accomplish a common goal (Artzt and Newman 1990).

The instructional use of small groups so that students work together to maximize their own and each other's learning (Johnson, Johnson, and Smith 1991).

A task for group discussion and resolution (if possible), requiring face-to-face interaction, an atmosphere of cooperation and mutual helpfulness, and individual accountability (Davidson 1990).

Cooperative learning also falls in the more general category of "collaborative learning," which is described as working in groups of two or more, mutually searching for understanding, solutions, or meanings, or creating a product (Goodsell, Maher, and Tinto 1992).

It is important to also consider what cooperative learning is not. According to Johnson et al. (1991), it is not having students sit side-by-side at the same table and talk with each other as they do their individual assignments, having students do a task individually with instructions that

those who finish first are to help the slower students, or assigning a report to a group where one student does all of the work and the others put their names on it.

## 2. WHY USE COOPERATIVE GROUPS?

Several recent reports urging reform of mathematics and science education in general (e.g., National Council of Teachers of Mathematics 1989, 1991; National Research Council 1989) and statistics education in particular (e.g., Cobb 1992), have described the need for specific changes in teaching. Instead of traditional lectures where teachers "tell" students information that they are to "remember," teachers are encouraged to introduce active-learning activities where students are able to construct knowledge. One way for teachers to incorporate active learning in their classes is to structure opportunities for students to learn together in small groups.

The suggestions made in these reports are supported by a growing set of research studies (over 375 studies, according to Johnson et al. 1991) documenting the effectiveness of cooperative learning activities in classrooms. A majority of the published research studies examine cooperative learning activities in elementary and secondary schools, and a subgroup of these studies focus on mathematics classes. The implication of these studies is that the use of small group learning activities leads to better group productivity, improved attitudes, and sometimes, increased achievement (Garfield, in press).

Only a few studies so far have examined the use of cooperative learning activities in college statistics courses. Shaughnessy (1977) found that the use of small groups appeared to help students overcome some misconceptions about probability and enhance student learning of statistics concepts. Dietz (1993) found that a cooperative learning activity on methods of selecting a sample allowed students to "invent" for themselves standard sampling methods, which resulted in better understanding of these methods. Jones (1991) introduced cooperative learning activities in several sections of a statistics course and observed dramatic increases in attendance, class participation, office visits, and student attitudes.

Another argument for using cooperative groups relates to the constructivist theory of learning, on which much of the current reform in mathematics and science education is based. This theory describes learning as actively constructing one's own knowledge. Constructivists view students as bringing to the classroom their own ideas, experiences, and beliefs, that affect how they understand and learn new material. Rather than "receiving" material in class as it is "delivered," students restructure the new information to fit into their own cognitive frameworks. In this manner, they actively and individually construct their own knowledge, rather than copying knowledge "transmitted" or "conveyed" to them. A related theory of teaching focuses on developing students' understanding, rather than on rote skill development.

Small-group learning activities may be designed to encourage students to construct knowledge as they learn new material, transforming the classroom into a community of learners, actively working together to understand statistics. The role of the teacher changes accordingly from that of "source of information" to "facilitator of learning." Part of this role is to be an ongoing assessor of student learning.

As part of the current reform of assessment of student performance, instructors are being encouraged to collect a variety of assessment information from sources other than individual student tests. Cooperative group activities may be structured to provide some rich information for teachers to use in assessing the nature of student learning. While walking around the class and observing students as they work in groups, the instructor is able to hear students express their understanding of what they have learned, which provides instructors with an ongoing, informal assessment of how well students are learning and understanding statistical ideas. Written reports on group activities may be used to assess students' ability to solve a particular

problem, apply a skill, demonstrate understanding of an important concept, or use higher-level reasoning skills.

A final argument for including cooperative group-learning activities in a statistics class is that businesses are increasingly looking for employees who are able to work collaboratively on projects and to solve problems as a team. Therefore, it is important to give students practice in developing these skills by working cooperatively on a variety of activities. This type of experience will not only build collaborative problem-solving skills, but will also help students learn to respect other viewpoints, other approaches to solving a problem, and other learning styles.

### 3. HOW COOPERATIVE LEARNING ACTIVITIES HELP STUDENTS LEARN

The use of small-group learning activities appears to benefit students in different ways. These activities often result in students teaching each other, especially when some understand the material better or learn more quickly than others. Those students who take on a "teaching" role often find that teaching someone else leads to their own improved understanding of the material. This result is reinforced by research on peer teaching that suggests that having students teach each other is an extremely effective way to increase student learning (McKeachie, Pintrich, Yi-Guang, and Smith 1986).

Just as "two heads are better than one," having students work together in a group activity often results in a higher level of learning and achievement than could be obtained individually. A necessary condition for this to occur is called "positive interdependence," the ability of group members to encourage and facilitate each other's efforts (Johnson et al. 1991). Positive interdependence can be promoted by careful design and monitoring of group activities.

Working together with peers encourages comparison of different solutions to statistical problems, problem solving strategies, and ways of understanding particular problems. This allows students to learn first-hand that there is not just one correct way to solve most statistics problems. Small group activities also provide students with opportunities to verbally express their understanding of what they have learned, as opposed to only interacting with material by listening and reading. By having frequent opportunities to practice communicating using the language of statistics they are better able to see where they have not yet mastered the material when they are unable to explain something adequately or communicate effectively with group members. Small-group discussions also allow students to ask and answer more questions than they would be able to in large-group discussions where typically a few students dominate the discussion.

Finally, students' achievement motivation is often higher in small-group activities because students feel more positive about being able to complete a task with others than by working individually (Johnson et al. 1991). Working together towards a mutual goal also results in emotional bonding where group members develop positive feelings towards the group and commitment towards working together. This increase in motivation may also lead to improved student attitudes towards the subject and the course.

### 4. TYPES OF COOPERATIVE GROUPS

There is not only one correct way in which to use groups, although there are guidelines for the effective use of groups in different types of course settings. The instructor may allow students to self-select groups or groups may be formed by the instructor to be either homogeneous or heterogeneous on particular characteristics (e.g., grouping together all students who received

A's on the last quiz, or mixing students with different majors). Johnson et al. (1991) describe several different types of groups, including informal and formal groups.

Informal groups are often used to supplement lectures in large classes, and may change everyday. These might consist of "turn to your neighbor" discussions to summarize the answer to a question being discussed, give a reaction to a discussion, or relate new information to past learning. In formal groups, students work with the same students for a longer period of time, sometimes for an entire semester. In these groups students may divide up work to be done on a particular in-class activity, work together to solve a problem or apply a statistical method, or work on long-term projects. Students may also use these groups to review material, complete homework assignments, teach each other information, encourage and support each other, and inform each other about information if a class has been missed.

## 5. IMPLEMENTING GROUPS

When first introducing a group activity it is useful to establish some rules for students. They should be informed that they are always responsible for their own work but must also be willing to help any group member who asks for help. If they have questions on an activity they should first ask each other, and may ask the instructor only if no one in the group can answer their question. They need to listen carefully to each other and share the leadership, making sure everyone participates and no one dominates.

It is also important to establish respect and consideration for all members of the group. One way is to point out that all people learn at different rates and that there are many different ways in which people best learn. It is important that students recognize and accept these differences and be respectful of each other. Students may be told that they will learn statistics more effectively by asking questions, answering questions, helping each other, and analyzing each other's mistakes. This is quite a contrast from the role they may be used to in most college classes where they passively listen to lectures.

Finally, students need to be aware that the problems they will be solving may often be solved in different, correct ways. They should be encouraged to try to learn from each other by comparing and explaining different solutions.

## 6. GROUP ROLES

In order to encourage positive interdependence among group members, students may be assigned to specific roles, which can be rotated each day (Johnson et al. 1991). These roles may help students get started on the activity and also prevent one person from doing all of the work. A "moderator/organizer" is in charge of assigning tasks to the group members, moderating group discussions, overseeing that the assigned task is being carried out, and helping to keep the group on course. A "summarizer's" job is to summarize discussions or group solutions to a problem, so that the "recorder" may write down what the summarizer says. Sometimes it is useful to have a "strategy suggester" or "seeker of alternative methods," who challenges the group to try other methods or explore other ways to solve a problem. A "mistake manager" may ask the group what went wrong and what can be learned from mistakes made. Finally, an "encourager" can be designated to encourage participation from all group members by using probes such as: "What do you think," "Can you add to that?" or by giving positive reinforcement to group members as they contribute to the discussion.

## 7. HOW TO USE COOPERATIVE GROUPS IN A STATISTICS CLASS

It is recommended that the instructor carefully read one of the excellent resources on cooperative or collaborative learning in higher education (e.g., Goodsell et al. 1992; Johnson et al. 1991)



before incorporating cooperative groups in a statistics class. These resources provide complete information on structuring and monitoring groups, and developing and evaluating group activities. Based on the models of cooperative learning in these references, cooperative group activities for a particular statistics class can be developed. These activities might include:

1. Having groups individually solve a problem and then compare their solutions (e.g., home work problems or problems from the textbook requiring particular skills).
2. Having groups discuss a concept or procedure, or compare different concepts or procedures (e.g., discuss the steps involved in testing an hypothesis, or compare the advantages and disadvantages of using the mean, median, and mode to summarize a data set).
3. Giving groups a data set to analyze and to discuss, followed by a written report of what they have learned about the data (e.g., data sets from the Quantitative Literacy materials or data generated in class such as estimating the distribution of the number of raisins in a small box).
4. Having each group collaborate on a large project involving collecting, analyzing, and interpreting data. Groups may meet in and/or outside of class to work on these projects, and may present the results in a written report and/or an oral in-class presentation.
5. Using groups as a way to learn new material. The jigsaw method can be used, where students are assigned temporarily to new groups, and each new group learns something new, such as a different type of plot. Then, students return to their original groups and teach each other the material they just learned.
6. Having groups compare their ideas about chance phenomena, and then generate or simulate data to test their beliefs (e.g., distributions of heads and tails when coins are tossed, or the best strategy to choose in the Monty Hall problem).

The three articles mentioned previously on using cooperative groups in college statistics courses (Dietz 1993; Jones 1991; Shaughnessy 1977) provide more detailed descriptions of particular cooperative group activities. Samples of other group activities are included in the appendix.

## 8. CHARACTERISTICS OF GOOD GROUP ACTIVITIES

Although cooperative group activities can be used in many different ways, it is important to consider characteristics of good activities in designing activities for a statistics class. Activities should require that all members of the group be involved, and not allow just one or two students to do the work while the others observe them. The instructions for the activity should be made very clear so that students do not spend time trying to figure out what it is they are supposed to do, or take a wrong path because of misunderstanding the activity.

Students should know in advance that they are accountable for the end product, both individually and as a group. This may result in individual write-ups or contributions to a group product, where students may be asked to evaluate the extent of their individual contributions. There should also be some assessment of the results of a group activity so that students receive feedback and learn from any mistakes made.

## 9. EVALUATING STUDENT LEARNING

It is important that group activities conclude with some type of summary of what students have learned. Students may be asked to turn in their individual work or to write one group summary of the results of the activity. Grades or points may be assigned in different ways. If students work together but turn in separate reports, these may be rated individually and then a group score

based on the average assigned as well. If only a group score is assigned to a group product, students may be asked to volunteer the percentage of their contribution, and that may be used to determine their share of the group points. Or, a group score may be assigned and everyone receives that score. Methods will vary based on the types of projects and students.

In addition to having the instructor evaluate a group product, students should be encouraged to assess their own group product, as well as how effectively their group worked together. Sometimes products may be exchanged between groups so that students may critique each other's work. Students may also participate in a group quiz, where they work together to solve problems on the quiz. This method is particularly effective if it takes place after students have individually taken the same quiz.

## 10. CONCERNS ABOUT USING SMALL GROUPS

Despite the recommendations and encouragement for using groups, there are still concerns expressed by statistics instructors who are either contemplating using groups or have tried it and have had some negative reactions.

Some instructors feel uncomfortable losing their role of being on center stage, performing in front of appreciative students. In using groups, the teacher's role in class is more in the background, where they may observe, listen, and assist students only as needed. Instead of elegantly demonstrating the solution of problems or proofs in front of the students, instructors will often step back and watch the students struggle through these same problems. They also take on the role of questioners, asking members of the groups about their conclusions or solutions to problems, asking them to justify what they did and why.

Instructors may be discouraged by students who resist an activity that appears challenging and difficult, forces them to think, and does not allow them to be passive learners, because students are used to sitting in lectures where they are not required to talk, solve problems, or struggle with learning new material. Students may want the teacher to do more explaining, and telling them the right answers, rather than struggle with a problem themselves. Some students may prefer to work alone, and resist being forced to work in a group. Sometimes this concern is related to the issue of grading fairness; students may feel it is unfair to give one grade to the entire group rather than separate grades to individual students, especially if students do not contribute equally. By adopting a grading policy where the amount of individual student contributions are used to assign points may alleviate this concern.

Sometimes group activities are an instant success, and the instructor is encouraged to continue using groups in class. Other times concerns such as the ones listed above arise. Goodsell et al. (1992) have several helpful suggestions for overcoming problems with students in using groups. As teachers develop more experience in designing and managing group activities, and as students become more accustomed to learning and working together, these problems usually disappear.

## 11. CONCLUSION

Cooperative group learning includes a wide variety of activities that may be implemented in several different ways in a college statistics class. These activities offer ways for students to become more involved in learning and to develop improved skills in working with others. The strong support of research and the recommendations from recent reports urging educational reform should encourage more instructors to introduce cooperative group activities in their classes. Perhaps as more statistics faculty begin to experiment with the use of small-groups and to evaluate their effectiveness in improving student learning, we will be able to develop a core of research to inform us as to the best types of activities to use in helping students learn particular statistical concepts.

## APPENDIX: THREE SAMPLE GROUP ACTIVITIES

### A. Measures of Center

Each group of students is given some different data sets (e.g., prices of running shoes, fat content of fast foods, Olympic medals, temperatures for a month). These may either be plots of data or sets of numbers that students can use to first construct simple plots. They are given the following instructions:

1. In your group, discuss each of the four measures of center you read about in Chapter 4. Make sure that everyone understands what each measure is and how it is calculated.
2. Discuss the advantages and disadvantages of using each of the four measures to summarize a data set.
3. For each of the distributed data sets, determine which measure of center would be most appropriate as a single number summary and why.
4. Turn in one written summary of your discussion. Be sure to include a description of each measure and how it is calculated, advantages and disadvantages of each measure, and a discussion of which measure of center is most appropriate to use in representing each data set and why.

### A. Coke vs. Pepsi

This activity was developed by the NSF-funded CHANCE project. For this activity you will need some large bottles of Coke and Pepsi and many small paper cups. Students are asked to select one person in their group who they think can distinguish between Coke and Pepsi. They are asked to design and conduct an experiment to determine how well this person can actually distinguish between the two types of drinks. After the experiment is completed, groups share their methods and results in a class discussion. Characteristics of good experiments are highlighted and students are asked to consider which results they most trust.

### B. Monty Hall

This activity is designed to help students empirically test the wisdom of their intuitive ideas about chance events. They are given the following scenario:

Suppose you're on the game show, Let's Make a Deal, and you're given the choice of three doors. Behind one door is a car; behind the other two doors are goats. You pick a door, say Door Number 1, and the host, who knows what's behind the doors, opens another door, say Door Number 3, which has a goat. He then says to you, "Do you want to pick Door Number 2?" Is it to your advantage to switch your choice?

1. Discuss in your group whether you should switch or stay with your original choice, or whether it makes no difference. What assumptions does your answer depend on? What other assumptions might you make and how would they affect your answer? Summarize your discussion in a paragraph.

2. Carry out an experiment to test your decision. For example, one person can serve as the host, Monty Hall, and the other group members as contestants. The host takes three cards, each representing a door. On the back of each is a goat (on cards) or a car. The host lays out the three cards, blank side up, making sure he/she knows which one has the car on the other side. After each game, the cards are rearranged and the game is played again, keeping track of the strategy used (switch or stay) and the result.

3. Which strategy is better? Summarize the data used to support your decision. Determine the probabilities of winning if you switch doors and if you stay with the original choice of doors.

## REFERENCES

Artzt, A. and Newman, C. (1990), *How to Use Cooperative Learning in the Mathematics Class*, Reston, VA: National Council of Teachers of Mathematics.

Cobb, George (1992), "Teaching Statistics," in *Heeding the Call for Change: Suggestions for Curricular Action*, ed. L. Steen, MAA Notes, No. 22.

Davidson, N. (ed.) (1990), *Cooperative Learning in Mathematics: A Handbook for Teachers*, Menlo Park: Addison Wesley.

Dietz, E. J. (1993), "A Cooperative Learning Activity on Methods of Selecting a Sample," *The American Statistician*, 47, 104-108.

Garfield, J. (in press), "How Students Learn Statistics," *International Statistical Review*.

Goodsell, A., Maher, M., and Tinto, V. (1992), *Collaborative Learning: A Source book for Higher Education*, University Park, PA: National Center on Post secondary Teaching, Learning and Assessment.

Johnson, D., Johnson, R., and Smith, K. (1991), *Cooperative Learning: Increasing College Faculty Instructional Productivity*, ASHE-ERIC Higher Education Report No. 4, Washington, DC: The George Washington University.

Jones, L. (1991), "Using Cooperative Learning to Teach Statistics," Research Report Number 91-2, The L.L. Thurstone Psychometric Laboratory, University of North Carolina.

McKeachie, W., Pintrich, P., Yi-Guang, L., and Smith, D. (1986), *Teaching and Learning in the College Classroom: A Review of the Research Literature*, Ann Arbor: Regents of the University of Michigan.

National Council of Teachers of Mathematics (1989), *Curriculum and Evaluation Standards for School Mathematics*, Reston, VA: Author.

National Council of Teachers of Mathematics (1991), *Professional Standards for Teaching Mathematics*, Reston, VA: Author.

National Research Council (1989), *Everybody Counts: A Report to the Nation on the Future of Mathematics Education*, Washington, DC: National Academy Press.

Shaughnessy, J. M. (1977), "Misconceptions of Probability: An Experiment With a Small-Group Activity-Based Model Building Approach to Introductory Probability at the College Level," *Educational Studies in Mathematics*, 8, 285-316.

## 8.2 Beyond Testing and Grading

by Joan Garfield

Journal of Statistics Education v.2, n.1 (1994)

### 1. An Evolving View of Assessment

1 The term "assessment" is often used in different contexts and means different things to different people. Most statistics faculty think of assessment in terms of testing and grading: scoring quizzes and exams and assigning course grades to students. We typically use assessment as a way to inform students about how well they are doing or how well they did in the courses we teach. An emerging vision of assessment is that of a dynamic process that continuously yields information about student progress toward the achievement of learning goals (NCTM 1993). This vision of assessment acknowledges that when the information gathered is consistent with learning goals and is used appropriately to inform instruction, it can enhance student learning as well as document it (NCTM 1993). Rather than being an activity separate from instruction, assessment is now being viewed as an integral part of teaching and learning, and not just the culmination of instruction (MSEB 1993).

2 Because learning statistics has often been viewed as mastering a set of skills, procedures, and vocabulary, student assessment has focused on whether these have been mastered, by testing students' computational skills or their ability to retrieve information from memory (Hawkins, Jolliffe, and Glickman 1992). Statistics items that appear on traditional tests typically test skills in isolation of a problem context and do not test whether or not students understand statistical concepts, are able to integrate statistical knowledge to solve a problem, or are able to communicate effectively using the language of statistics. Research has shown that some students who produce a correct "solution" on a test item may not even understand this solution or the underlying question behind it (Jolliffe 1991).

3 As goals for statistics education change to broader and more ambitious objectives, such as developing statistical thinkers who can apply their knowledge to solving real problems, a mismatch is revealed between traditional assessment and the desired student outcomes. It is no longer appropriate to assess student knowledge by having students compute answers and apply formulas, because these methods do not reveal the current goals of solving real problems and using statistical reasoning.

4 The current reform movement in educational assessment encourages teachers to think about assessment more broadly than "testing" and using test results to assign grades and rank students (e.g., Romberg 1992, Lesh and Lamon 1992). The recent report on assessment, *Measuring What Counts* (MSEB 1993), offers some basic principles of mathematics assessment. Two of these principles, rephrased to focus on statistics instead of mathematics, are:

- **The Content Principle:** Assessment should reflect the statistical content that is most important for students to learn.
- **The Learning Principle:** Assessment should enhance learning of statistics and support good instructional practice.

5 These principles directly lead to the use of alternative forms of assessment to provide more complete information about what students have learned and are able to do with their knowledge, and to provide more detailed and timely feedback to students about the quality of their learning.

Assessment approaches now being used in mathematics better capture how students think, reason, and apply their learning, rather than merely having students "tell" the teacher what they have remembered or show that they can perform calculations or carry out procedures correctly (e.g., EQUALS 1989). Some of these alternative methods -- portfolio assessment, authentic assessment, and performance assessment -- are described below.

- **Portfolio assessment:** The collection and evaluation of a carefully chosen selection of students' work. The number and types of selections included in a portfolio may vary, but are typically agreed upon by the teacher and student for the purpose of representing what that student has learned (Pandey 1991).
- **Authentic assessment:** A method of obtaining information about students' understanding in a context that reflects realistic situations, and that challenges students to use what they have learned in class in an authentic context (Archbald and Newmann 1988).
- **Performance assessment:** Presenting students with a task, project, or investigation, then evaluating the products to assess what students actually know and can do (Stenmark 1991).

6 Before selecting these or other alternatives to traditional testing, it is important to consider criteria for their appropriate use. In reviewing the National Council of Teachers of Mathematics (NCTM) standards for assessment of mathematics learning, Webb and Romberg (1992) provide criteria for assessment instruments and procedures that are relevant to the development or selection of statistical assessment materials as well. These criteria specify that good assessment should:

- Provide information that will contribute to decisions regarding the improvement of instruction.
- Be aligned with instructional goals.
- Provide information on what students know.
- Supplement other assessment results to provide a global description of what students know.

7 In considering these criteria, a broader view of assessment emerges, beyond that of testing and grading. In this view, assessment becomes an integral part of instruction, consists of multiple methods yielding complementary sources of information about student learning, and provides both the student and instructor with a more complete analysis of what has happened in a particular course.

## 2. Purposes of Assessment

8 Why should a statistics instructor consider implementing assessment methods other than traditional tests and quizzes in a college statistics course? I feel the most compelling reason is because traditional forms of assessment rarely lead to improved teaching and learning and offer us limited understanding of our students: what attitudes and beliefs they bring to class, how they think about and understand statistics, and how well they are able to apply their knowledge.

Without this knowledge it is difficult to determine how to make changes or design instruction to improve student learning.

9 The primary purpose of any student assessment should be to improve student learning (NCTM 1993). Some secondary purposes for gathering assessment information include:

- To provide individual information to students about how well they have learned a particular topic and where they are having difficulty.
- To provide information to the instructor about how well the class seems to understand a particular topic and what additional activities might need to be introduced, or whether it is time to move on to another topic.
- To provide diagnostic information to instructors about individual students' understanding or difficulties in understanding new material.
- To provide information to teachers about students' perceptions and reactions to the class, the material, the subject matter, or particular activities.
- To provide an overall indicator of students' success in achieving course goals.
- To help students determine their overall strengths and weaknesses in learning the course material.

10 Selection of appropriate assessment methods and instruments depends on the purpose of assessment: why the information is being gathered and how it will be used. If the purpose of a particular assessment activity is to determine how well students in the class have learned some important concepts or skills, this may result in a different instrument or approach than if the purpose is to provide quick feedback to students so that they may review material on a particular topic.

11 Regardless of the specific purpose of an assessment procedure, incorporating an assessment program in our classes offers us a way to reflect about what we are doing and to find out what is really happening in our classes. It provides us with a systematic way to gather and evaluate information to use to improve our knowledge, not only of students in a particular course, but our general knowledge of teaching statistics. By using assessment to identify what is not working, as well as what is working, we can help our students become more aware of their own success in learning statistics, as well as become better at assessing their own skills and knowledge.

### **3. What Should Be Assessed?**

12 Because assessment is often viewed as driving the curriculum, and students learn to value what they know they will be tested on, we should assess what we value. First we need to determine what students should know and be able to do as a result of taking a statistics course. This information should be translated into clearly articulated goals and objectives (both broad and narrow) in order to determine what types of assessment are appropriate for evaluating attainment of these goals. One way to begin thinking about the main goals for a course is to consider what students will need to know and do to succeed in future courses or jobs. Wiggins (1992) suggests that we think of students as apprentices who are required to produce quality work, and are therefore assessed on their real performance and use of knowledge. Another way to determine

important course goals is to decide what ideas you really want students to retain six months after completing your statistics class.

13 I believe that the main goals of an introductory statistics course are to develop an understanding of important concepts such as mean, variability, and correlation. We also want students to understand ideas such as the variability of sample statistics, the usefulness of the normal distribution as a model for data, and the importance of considering how a sample was selected in evaluating inferences based on that sample. We would like our students to be able to intelligently collect, analyze, and interpret data; to use statistical thinking and reasoning; and to communicate effectively using the language of statistics.

14 In addition to concepts, skills, and types of thinking, most instructors have general attitude goals for how we would like students to view statistics as a result of our courses. Such attitude goals include understanding how the discipline of statistics differs from mathematics, realizing that you do not have to be a mathematician to learn and understand statistics, believing that there are often different ways to solve a statistical problem, and recognizing that people may come to different conclusions based on the same data if they have different assumptions and use different methods of analysis (Garfield, in press).

15 Once we have articulated goals for students in our statistics classes, we are better able to specify what to focus on to determine what is really happening to students as they experience our courses. Are they learning to use statistical thinking and reasoning, to collect and analyze data, to write up and communicate the results of solving real statistical problems? Some goals may not be easy to assess individually, and may be more appropriately evaluated in the context of clusters of concepts and skills. For example, in order to evaluate whether students use statistical reasoning in drawing conclusions about a data set, students may need to be given the context of a report of a research study that requires them to evaluate several related pieces of information (e.g., distributions of variables, summary statistics, and inferences based on that data set). Determining if students have achieved the goal of understanding how to best represent a data set with a single number may require that students examine and evaluate several distributions of data.

#### **4. How To Assess Student Learning**

16 There are several ways to gather assessment information, and it is often recommended that multiple methods be used to provide a richer and more complete representation of student learning (e.g., NCTM 1989). What all types of assessment have in common is that they consist of a situation, task, or questions; a student response; an interpretation (by the teacher or one who reviews the assessment information); an assignment of meaning to the interpretation; and reporting and recording of results (Webb 1993).

17 Different assessment methods to use in a statistics class include:

- quizzes (including calculations and/or essay questions)
- minute papers (e.g., on what students have learned in a particular session, or what they found to be most confusing)
- essay questions/journal entries (explaining their understanding of concepts and principles)
- projects/reports (individual or group)
- portfolios (including a selection of different materials)



- exams (covering a broad range of material)
- attitude surveys (administered at different times, about the course, content, or view of statistics)
- written reports (of in-class activities or computer labs)
- open-ended questions or problems to solve
- enhanced multiple-choice questions where responses are designed to characterize students' reasoning (e.g., Garfield 1991).

18 How are these different types of assessment evaluated? Quizzes and essay questions may be graded and assigned a single grade or score. More complex assessments such as projects and written reports may be evaluated using alternative scoring procedures. Although these procedures may be used to assign a grade, they may also be used to help students learn how to improve their performance, either on this task or future ones. Evaluation procedures for projects and reports may consist of:

- Giving a grade of either A or "needs work," allowing the student to revise and improve the product until it meets the instructor's standard.
- Using a scoring rubric to assign points (such as 0, 1, 2) to different components of the assessment, providing more detailed feedback to students on different aspects of their performance. For example, the following categories may be used to evaluate a student project:
  - o Understands the problem
  - o Describes an effective solution
  - o Discusses limitations of the solution
  - o Communicates effectively

19 For a list of other guidelines to use in scoring student projects or reports, see Hawkins et al. (1992).

20 For any type of assessment used to assign student grades, it is recommended that the scoring rubrics to be used, some model papers, and exemplars of good performance be shared with students in advance. These samples help provide students with insights into what is expected as good performance, allowing them to acquire standards comparable to the instructor's standards of performance (Wiggins 1992). Other assessment information such as minute papers or attitude surveys need not be given a score or grade, but can be used to inform the teacher about student understanding and feelings, as input for modifying instruction.

## 5. An Assessment Framework

21 An assessment framework emerges from the different aspects of assessment: what we want to have happen to students in a statistics course, different methods and purposes for assessment, along with some additional dimensions. (This framework is based on an earlier version developed in collaboration with Iddo Gal.)

22 The first dimension of this framework is **WHAT** to assess, which may be broken down into: concepts, skills, applications, attitudes, and beliefs.

23 The second dimension of the framework is the **PURPOSE** of assessment: why the information is being gathered and how the information will be used (e.g., to inform students about strengths and weaknesses of learning, or to inform the teacher about how to modify instruction).

24 The third dimension is **WHO** will do the assessment: the student, peers (such as members of the student's work group), or the teacher. It is important to point out that engaging the student in self-assessment is a critical and early part of the assessment process, and that no major piece of work should be turned in without self-criticism (Wiggins 1992). Students need to learn how to take a critical look at their own knowledge, skills, and applications of their knowledge and skills. They need to be given opportunities to step back from their work and think about what they did and what they learned (Kenney and Silver 1993). This does not imply that a grade from a self-rating given by a student is to be recorded and used by the teacher in computing a course grade, but rather that students should have opportunities to apply scoring criteria to their own work and to other students so that they may learn how their ratings compare to those of their teacher.

25 The fourth dimension of the framework is the **METHOD** to be used (e.g., quiz, report, group project, individual project, writing, or portfolio).

26 The fifth dimension is the **ACTION** that is taken and the nature of the **FEEDBACK** given to students. This is a crucial component of the assessment process that provides the link between assessment and improved student learning.

27 This framework is not intended to imply that an intersection of categories for each of the four dimensions will yield a meaningful assessment technique. For example, measuring students' understanding of the concept of variability (**WHAT** to assess) for the **PURPOSE** of finding out if students understand this concept, using students in the group as assessors (**WHO**), with the **METHOD** being a quiz, and the **ACTION/FEEDBACK** being a numerical score, may not yield particularly meaningful and useful results. It also doesn't make sense to assess student attitudes towards computer labs (**WHAT**) by having peers (**WHO**) read and evaluate student essays (**METHOD**). Obviously, some categories of dimensions are more appropriately linked than others.

28 Another important point in applying this framework is that it is often difficult to assess a single concept in isolation of other concepts and skills. It may not be possible to assess understanding of standard deviation without understanding the concepts of mean, variability, and distribution. When given the task last fall, a group of statistics educators were unable to design an appropriate assessment for understanding the concept of "average" without bringing in several other concepts and skills.

29 Here are four examples of assessment activities illustrating the dimensions of the framework.

### 30 Example 1:

**WHAT:** Students' understanding of the Central Limit Theorem.

**PURPOSE:** To find out if students need to review text material or if the teacher needs to introduce additional activities designed to illustrate the concept (e.g., computer simulations of sampling distribution).

**METHOD:** An essay question written in class as a quiz, asking students to explain the theorem and illustrate it using an original example.

**WHO:** The instructor will evaluate the written responses.

**ACTIONS/FEEDBACK:** The instructor reads the essay responses and assigns a score of 0 (shows no understanding) to 3 (shows very clear understanding). Students with scores of 0 and 1 are assigned additional materials to read or activities to complete. Students with scores of 2 are given feedback on where their responses could be strengthened.

### 31 Example 2:

**WHAT:** Students' ability to apply basic exploratory data analysis skills.

**PURPOSE:** To determine if students are able to apply their skills to the collection, analysis, and interpretation of data.

**METHOD:** A student project, where instructions are given as to the sample size, format of report, etc. (e.g., Garfield 1993).

**WHO:** First the student completes a self-assessment using a copy of the rating sheet the instructor will use, which has been distributed prior to completing the project. Then, the instructor evaluates the project using an analytic scoring method (adapted from the holistic scoring method for evaluating student solutions to mathematical problems offered by Charles, Lester, and O'Daffer (1987)). A score of 0 to 2 points for each of six categories is assigned, where 2 points indicates correct use, 1 point indicates partially correct use, and 0 points indicates incorrect use. Explanations of each category are given below:

- **Communication:** The appropriate use of statistical language and symbols.
- **Visual Representation:** The appropriate construction and display of tables and graphs.
- **Statistical Calculations:** Statistical measures are calculated correctly and appear reasonable.
- **Decision Making:** The tables or graphs selected are appropriate for representing the data, and the appropriate summary measures are calculated.
- **Interpretation of Results:** The ability to use information from

representations and summary measures to describe a data set.

- **Drawing Conclusions:** The ability to draw conclusions about the data, point out missing information, or relate this study to other information.

**ACTIONS/FEEDBACK:** Scores are assigned to each category and given back to students along with written comments, early enough in the course so that they may learn from this feedback in working on their next project.

### 32 Example 3:

**WHAT:** Students' perceptions of the use of cooperative group activities in learning statistics.

**PURPOSE:** To find out how well groups are working and to determine if groups or group procedures need to change.

**METHOD:** A "minute paper" is assigned during the last five minutes of class, where students are asked to write anonymously about their perceptions of what they like best and like least about their experience with group activities.

**WHO:** The teacher reads the minute papers.

**ACTIONS/FEEDBACK:** The teacher summarizes the responses and shares them with the class, and makes changes in groups or group methods as necessary.

### 33 Example 4:

**WHAT:** Students' understanding of statistical inference.

**PURPOSE:** To evaluate students' understanding of statistical inference and give a grade to students for a major portion of course work on this topic.

**METHOD:** A portfolio. Students are asked to select samples of their work from a three-week unit on inference to put in a portfolio folder. They select examples of written assignments, computer lab write-ups, group activities, and writing assignments, making sure that particular topics are represented (such as constructing and interpreting confidence intervals). Students select samples of their work, write a brief summary describing why they selected each piece, and give their own rating of the overall quality of their work in this unit.

**WHO:** The teacher reviews the portfolios and completes a rating sheet for each one. A scoring rubric is used, including categories such as the following:

- Demonstrates understanding of hypothesis testing.
- Correctly determines and interprets type 1 and type 2 error.
- Correctly constructs and interprets confidence intervals.
- Correctly uses and interprets p-values.

- Selects the appropriate procedures in testing hypotheses involving one or two groups.
- Demonstrates understanding of statistical significance.

ACTIONS/FEEDBACK: Portfolios are returned to students with completed rating sheets. The students are asked to review areas of weakness or errors made and may submit a follow-up paper demonstrating their understanding of these topics. The teacher may address some common mistakes/weaknesses in class before going on to the next topic. (For more information on portfolios, see Crowley 1993.)

## 6. Implications for Statistics Instructors

34 Given the calls for reform of statistical education and the new goals envisioned for students, it is crucial that we look carefully at what is happening to students in our classes. Without scrutinizing what is really happening to our students and using that information to make changes, it is unlikely that instruction will improve and we will be able to achieve our new vision of statistical education. I would like to offer some suggestions for instructors contemplating alternative assessment procedures for their classes:

- Try to look at every assessment activity as a way to provide students with feedback on how to improve their learning and not just as an activity used to assign a grade.
- Don't try to do it all at once. Pick one method, try it for a while, and then gradually introduce and experiment with other techniques.
- Don't try to do it alone. Plan, review, and discuss with other colleagues or teaching assistants what you are doing and what you are learning from assessment information.
- Be open with the students about why and how they are being assessed (Stenmark 1991).
- Make sure that you have opportunities to reflect on the assessment information you obtain, and monitor the impact of these results on your perceptions of the class and your teaching.
- Consult resources for ideas of different approaches to use and ways to evaluate assessment information (e.g., Angelo and Cross 1993; Kulm 1990).

35 Finally, remember that assessment drives instruction, so be careful to assess what you believe is really important for students to learn. Use assessment to confirm, reinforce, and support your ideas of what students should be learning. Never lose track of the main purpose of assessment: to improve learning.

-----

### References

Angelo, T., and Cross, K. (1993), *A Handbook of Classroom Assessment Techniques for College Teachers*, San Francisco: Jossey-Bass.

Archbald, D., and Newmann, F. (1988), *Beyond standardized testing: Assessing authentic academic achievement in the secondary school*, Reston, VA: National Association of Secondary School Principals.

Charles, R., Lester, F., and O'Daffer, P. (1987), *How to Evaluate Progress in Problem Solving*, Reston, VA: National Council of Teachers of Mathematics.

Cobb, G. (1992), "Teaching Statistics," in *Heeding the Call for Change: Suggestions for Curricular Action*, ed. L. Steen, MAA Notes, No. 22.

Crowley, M. L. (1993), "Student Mathematics Portfolio: More Than a Display Case," *The Mathematics Teacher*, 87, 544-547.

EQUALS Staff, (1989), *Assessment Alternatives in Mathematics*, Berkeley, CA: Lawrence Hall of Science, University of California.

Garfield, J. (in press), "How Students Learn Statistics," *International Statistical Review*.

Garfield, J. (1993), "An Authentic Assessment of Students' Statistical Knowledge," in *National Council of Teachers of Mathematics 1993 Yearbook: Assessment in the Mathematics Classroom*, ed. N. Webb, Reston, VA: NCTM, pp. 187-196.

Garfield, J. (1991), "Evaluating Students' Understanding of Statistics: Development of the Statistical Reasoning Assessment," in *Proceedings of the Thirteenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Volume 2*, ed. R. Underhill, Blacksburg, VA, pp. 1-7.

Hawkins, A., Jolliffe, F., and Glickman, L. (1992), *Teaching Statistical Concepts*, Harlow, Essex, England: Longman Group UK Limited.

Jolliffe, F. (1991), "Assessment of the Understanding of Statistical Concepts," in *Proceedings of the Third International Conference on Teaching Statistics, Vol. 1*, ed. D. Vere-Jones, Otago, NZ: Otago University Press, pp. 461-466.

Kenney, P., and Silver, E. (1993), "Student Self-Assessment in Mathematics," in *National Council of Teachers of Mathematics 1993 Yearbook: Assessment in the Mathematics Classroom*, ed. N. Webb, Reston, VA: NCTM, pp. 229-238.

Kulm, G., ed. (1990), *Assessing Higher Order Thinking in Mathematics*, Washington, DC: AAAS.

Lesh, R., and Lamon, S. (1992), *Assessment of Authentic Performance in School Mathematics*, Washington, DC: AAAS.

Mathematical Sciences Education Board (1993), *Measuring What Counts: A Conceptual Guide for Mathematical Assessment*, Washington, DC: National Academy Press.

National Council of Teachers of Mathematics (1993), *Assessment Standards for School Mathematics: Working Draft*, Reston, VA: NCTM.

National Council of Teachers of Mathematics (1989), *Curriculum and Evaluation Standards for School Mathematics*, Reston, VA: NCTM.

Pandey, T. (1991), *A Sampler of Mathematics Assessment*, Sacramento, CA: California Department of Education.

Romberg, T., ed. (1992), *Mathematics Assessment and Evaluation: Imperatives for Mathematics Education*, Albany: State University of New York Press.

Stenmark, J. (1991), *Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions*, Reston, VA: NCTM.

Webb, N. (1993), "Assessment for the Mathematics Classroom," in *National Council of Teachers of Mathematics 1993 Yearbook: Assessment in the Mathematics Classroom*, ed. N. Webb, Reston, VA: NCTM, pp. 1-6.

Webb, N., and Romberg, T. (1992), "Implications of the NCTM Standards for Mathematics Assessment," in *Mathematics Assessment and Evaluation: Imperatives for Mathematics Education*, ed. T. Romberg, Albany: State University of New York Press, pp. 37-60.

Wiggins, G. (1990), "The Truth May Make You Free, but the Test May Keep You Imprisoned," *AAHE Assessment Forum*, 17-31.

Earlier versions of this paper were presented at meetings of the American Statistical Association (August, 1993) and the American Educational Research Association (April, 1994).

### 8.3 Experiences with Authentic Assessment Techniques in an Introductory Statistics Course

Beth L. Chance  
University of the Pacific

*Journal of Statistics Education v.5, n.3 (1997)*

Copyright (c) 1997 by Beth L. Chance, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the author and advance notification of the editor.

-----

Key Words: Computer labs; Journals; Projects; Take-home exams.

#### Abstract

In an effort to align evaluation with new instructional goals, authentic assessment techniques (see, e.g., Archbald and Newmann 1988, Crowley 1993, and Garfield 1994) have recently been introduced in introductory statistics courses at the University of the Pacific. Such techniques include computer lab exercises, term projects with presentations and peer reviews, take-home final exam questions, and student journals. In this article, I discuss the University of the Pacific's goals and experiences with these techniques, along with strategies for more effective implementation.

#### 1. Introduction

As instructional goals in statistics courses change, so must the assessment techniques used to evaluate progress towards these goals. Many statistics courses are shifting focus (see, e.g., Cobb 1992 and NCTM 1989), emphasizing skills such as the ability to interpret, evaluate, and apply statistical ideas over procedural calculations. Many of these outcomes are not adequately assessed using traditional tests, which too often emphasize the final numerical answer over the reasoning process (Garfield 1993, NCTM 1993). Thus, instructors need to accompany these new instructional aims with more authentic assessment techniques that address students' ability to evaluate and utilize statistical knowledge in new domains, communicate and justify statistical results, and produce and interpret computer output. Further, students need to receive feedback not only on their exam performance, but also constructive indications of their strengths and weaknesses, guidelines for improving their understanding, and challenges to extend their knowledge. The instructor needs to know not only how the students are mastering the material, but also how to improve instructional techniques and enhance statistical understanding. Since students' attitudes towards statistics can affect their learning (see, e.g., Gal and Ginsburg 1994), an assessment program should also include a way of judging how students are reacting to the material, including their impression of the relevance and fairness of the assessment process. Above all, assessment should mirror the skills students will need in order to be effective



consumers and evaluators of statistical information.

Towards these goals, I have incorporated several techniques into my courses that seem well suited to statistics instruction. These techniques include a computer laboratory component, a term project with peer reviews and oral presentations, a take-home component to the final exam, minute papers, and student journals. In this paper, I first discuss my goals for these assessment procedures and how my implementation of these techniques has evolved. Based on what I find works and doesn't work in my courses, I describe what I feel are the essential features of effective assessment techniques in the introductory statistics course.

## **2. Background**

The University of the Pacific is a small (4,000 students), private, comprehensive university. One of the university's goals is to maintain small class sizes, allowing more personalized instruction. The main statistics course, Math 37, fulfills a General Education requirement and is taken by students in numerous disciplines, such as business, science (computer, natural, social, sports), physical therapy, humanities, and pharmacy. Business is the most common major (Fall 1995: 36%, Spring 1996: 61%), and students come from all years (Fall 1995: 62% juniors, Spring 1996: 77% freshman and sophomores). The prerequisite is intermediate algebra, and the course is worth four credits. A different course is generally taken by mathematics and engineering majors. Typically, Math 37 is the only statistics course these students will see, and there are various levels of mathematics and computer anxiety and competency among the students. From surveys distributed the first day of class, most students indicate they are taking the course to fulfill a requirement for their major or to fulfill the General Education requirement (Fall 1995: 80%, Spring 1996: 92%). When asked their initial interest level in statistics, on a scale of one to five with five being high, 79% in the Fall and 88% in the Spring indicated an interest level of three or below. I believe authentic assessment techniques should work well in this environment due to the focus on student-centered learning at my university, and the flexibility I am given by my department in structuring the course.

## **3. Goals of Incorporating Nontraditional Techniques**

When designing my course, I identified several ways I wanted to enhance students' statistical learning experience and my assessment of that experience. In particular I wanted to better gauge the students' understanding of statistical ways of thinking, evaluate their communication and collaboration skills, measure their statistical and computer literacy, and monitor their interest in statistics. To see if these goals were met, I felt I needed to move beyond traditional assessment techniques and incorporate several nontraditional techniques in conjunction with traditional exams, homework exercises, and quizzes. Below I discuss how I combined several different assessment components together to address these goals. The Appendix gives an overview of how these different techniques are implemented in my introductory course.

### **3.1. Understanding of the Statistical Process**

I want my students to think of statistics as not just plugging numbers into formulas, but as a process for gaining information. Thus, I feel it is important to evaluate student understanding of this process by requiring them to complete such a process from beginning (conception of the question of interest) to end (presentation of results). One way to accomplish this is through semester-long projects. Use of projects in statistics courses has received considerable discussion in the literature (for example, Fillebrown 1994; Garfield 1993, 1994; Mackisack 1994; and Roberts 1992). My goals include seeing how students apply their knowledge to real (and possibly messy) data that they collect, while integrating ideas from many parts of the course (data collection, description, and analysis) into one complete package. By using the same project throughout the semester, students are able to rework their plan as new ideas occur, illustrating the dynamic nature of the process. Students also have an application to constantly refer to as new topics are covered in lecture. While the students' final projects may not incorporate all of the statistical techniques covered during the course, they need to evaluate each idea to judge its suitability for incorporation. Thus the students are determining the necessary components for their analysis. I feel this leads to more creativity in the projects, and better mastery of the statistical ideas, than when I tell them to apply one specific idea.

Students also see this cycle in their labs and final exam. Five of their 14 labs require students to collect the data themselves and then complete a technical report detailing their data collection procedures (e.g., precautions, operational definitions), analysis, diagnostic checks, and final conclusions. Students must discuss pitfalls in the data collection and analysis, and suggest potential improvements. At the end of the course, to further assess their ability to apply what they have learned, I now incorporate a take-home component to the final exam. Students are given a series of questions pertaining to a new data set. Students identify an appropriate test procedure based on the question asked, perform the computer analysis (choosing informative graphs and relevant diagnostic procedures), and interpret the results. Students need to integrate different portions of the course and review several potential approaches to the problem. Students are graded on their ability to identify and justify appropriate techniques, perform the analysis, and interpret their results to formulate a conclusion.

These assessment techniques provide the instructor with an authentic performance assessment that cannot be obtained solely from traditional timed, in-class exams. In particular, these approaches judge whether students can move beyond the textbook and apply the statistical concepts to new situations.

### **3.2. Statistical and Computer Literacy**

Another assessment goal in my course is to judge whether students possess sufficient knowledge to interpret statistical arguments and computer output. Many current newspaper articles are discussed in the course, and I often ask students to evaluate the merits of the numerical arguments and generalizations. For example, many of the homework assignments and test questions involve recent articles, following the model given in the CHANCE course (Snell and Finn 1991). Students are also asked to critique classmates' project reports. These reviews are completed in 10 to 20 minutes

at the end of one class, reviewed by the instructor, and then returned to the groups the next class period. Each project group receives two to four reviews. It would also be possible to grade the peer reviewer on the quality of the review to see if he or she can evaluate the proposal critically. In both these cases, I can see whether the students understand the ideas well enough to distinguish between good and bad uses of statistics.

While many questions can ask students to evaluate computer output, it is also important for students to work with computers directly, producing statistical output and interpreting that output in their own words. This was my initial justification for incorporating weekly computer lab sessions into the course. In most of these labs, students work with Minitab to analyze existing or self-collected data. The rest of the labs are directly aimed at enhancing student understanding of statistical concepts. These labs work with programs developed by delMas (delMas 1996) to develop visual understanding of concepts such as the standard normal curve, sampling distributions, and probability. Thus, students use the computer as a tool for analysis and for obtaining deeper conceptual understanding. In all of these labs, students' explanations of what they learned are emphasized.

### **3.3. Communication and Collaboration Skills**

Writing and being able to explain ideas to others have become important components of my course. In the lab write-ups, students either explain an important statistical concept at length, or identify, execute, and justify the relevant statistical procedure. I feel some of the students' most significant learning gains come from debating ideas with each other, so I encourage students to work in pairs on the lab write-ups to further increase this practice. Requiring students to explain ideas in their own words also greatly augments my ability to judge their level of understanding, while giving the students repetitive practice with the language of statistics and further internalization of the concepts. (For research on how giving explanations helps, see, for example, Rosenshine and Meister 1994.)

My method for allocating points on assignments reinforces the importance I give to the different components. When I grade the full lab write-ups, 50% of the points are given for discussion. Thus, students learn that interpretations and explanations are important and highly valued components of the analysis. Students find the writing component quite time consuming, often indicating that their least favorite lab is "any full lab write-up." They can have tremendous difficulty explaining their reasoning, reinforcing my belief in its necessity. The most common feedback I give is "Why?" or "How do you know this is true?" as I try to teach the students to build a logical argument based only on the statistical evidence. In grading the labs it is important to reward thoughtfulness and creativity, and especially to allow multiple interpretations of the same idea as long as the interpretations have been justified.

Another way I sometimes address the need for writing is through the use of student journals. I use many in-class activities (primarily based on those found in Scheaffer, Gnanadesikan, Watkins, and Witmer 1996), but I notice that while students find the in-class activities entertaining, they often miss the important statistical concept involved unless they are asked to further reflect on the ideas. Thus, when I use journals, I require

students to include some explanation of the activity, such as a simple paragraph explaining what the activity was about or answers to a series of directed questions. I also require chapter summaries to encourage students to continually review the material, blending lectures with text explanations and tying topics together. Due dates of project reports roughly correspond with midterms to help students assimilate the ideas they are reviewing. Hearing presentations of other project groups enhances review for the final exam. These additional reflections greatly aid students' understanding of the underlying ideas.

Peer reviews and oral presentations require the students to express their understanding to each other. I find that if only I present examples in class, students often don't participate fully and then later cannot extend the knowledge. Now each week, three to four students are responsible for presenting homework solutions on the board, with no prior notice. Also, at the end of the semester, each project group is required to give a presentation of their results to the class. The rest of the class and I evaluate these presentations. From homework and project presentations, I learn much more about the student's or group's ability to communicate results, as well as their ability to summarize the important points and to effectively convey results visually. In each of these tasks, students must learn to utilize the statistical terminology in a coherent manner and to explain the ideas to themselves, to other students, and to me. This requires them to develop a deeper understanding of statistical concepts.

Learning to function in groups should also be an important component of statistics education. Thus, the term projects are completed by teams of 3 to 5 students. In extreme situations (e.g., commuting from another city), students have been allowed to work individually. However, I encourage the group experience to foster debate and discussion among the group members, as well as to model their future working environment. The first semester I tried project groups, I learned that students at the University of the Pacific can have trouble working together in groups. To address these difficulties, I have given them more opportunities to work with other students (e.g., in-class activities) prior to selecting groups. Also, I now split the grades into an individual component and a group component. A student's final grade is based on a group grade (85%) and an individual grade (15%) that varies among the team members if there is a severe discrepancy in effort. This discrepancy is identified by confidential input from the group members at the end of the semester. The discrepancy has to be clear before points are adjusted. Knowing from the beginning that their teammates will provide this information has eliminated many of the potential problems. Contracts for projects can also be included, detailing which parts of the project each person will work on. Some instructors have also recommended including a quiz or test question requiring the individual students to demonstrate knowledge of the group project. Overall, I believe learning to work in groups should be a requirement, but the instructor needs to incorporate some flexibility and individual responsibility into the grading.

### **3.4. Establishing a Dynamic Assessment Process**

I also want to utilize assessment practices that allow students to reevaluate their work, in order to establish a dialogue with the instructor

and improve students' own self-evaluation skills. Through reevaluation of their work, students can develop their knowledge over time. For example, I grade the lab reports "progressively," requiring slightly higher levels of explanation as they proceed through the labs and incorporate the earlier feedback. Pairing students also allows me to provide more immediate feedback. Another process for continual review is the use of periodic project reports. Four times during the semester each group is asked to submit a report on their current progress for the instructor to review. This allows me to give guidance, prompt the students with questions, and answer their questions at several points during the semester while they are thinking about different stages of the project. I find this has greatly enhanced my ability to monitor and challenge the teams. Students see that statistics is an evolving process, and they also learn to justify their decisions.

Similarly, with the journals, the only grade that counts is their final grade, and I encourage students to resubmit earlier work to increase their credit. My goal is for students to continually reflect on a concept until they understand it fully. This approach has been most effective with the lab reports as I do see a tremendous improvement in the quality of the reports over the semester. It is important for students to be able to revisit an idea, so they can clear up misconceptions and enhance their understanding.

It is also important for students to learn to assess their own work. In the journals, I require self-evaluations so students will think critically about their own participation in the course and knowledge of the material. I have found that many of my students do not ask questions in class or office hours, and so I hope that by having them submit questions through the journal I can establish a dialogue with them and provide personally directed feedback in a nonthreatening environment.

### **3.5. Increasing Student Interest in Statistics**

Since my students do not enter the course with much interest in the material, I want my assessment program to monitor and enhance their interest levels. Thus, I choose techniques that are effective not only as instructional and assessment tools, but also as motivational tools. This is important because I believe increasing their interest will also increase their learning and retention of the material. Students often agree, as shown in the following comments.

"Doing the project did help strengthen my knowledge of proportions. When I actually had to sit down and do the calculations I found that the results made much more sense than before."

"When you apply what we've learned to the real world we see everyday it makes more sense and becomes more interesting."

"Putting the ideas to use further implants them into the brain cells."

To maximize interest, I allow students to pick their own project topics. I want to be sure the topic is of interest to the students, and I encourage

them to expand on activities from other courses. This fosters student creativity and ownership in the topic. Students appreciate seeing a "real world application."

"It certainly made it more interesting to see that we could actually use what we learned in class."

"It was good to analyze something we were interested in."

"It was good to be able to use what we learned in a study we conducted ourselves."

"The project helped me to see the practical use of statistics. This type of project was more interesting than just computing data for homework."

By allowing some flexibility in what topics the students examine, they will better see the relevance of statistics and want to learn the ideas. When asked if they enjoyed the project, 79% of Spring 1996 students (n = 63) said yes; 86% said the project helped them understand the material; 62% said the project increased their interest in the material.

Similarly, I find increasing the number of labs where students collect data about themselves and their classmates increases their interest in the answers. I also incorporate an "Other" category in the journal assignment. Students must contribute experiences they have with statistics outside the course, for example, an interesting article or news reference. My aim is to give students the opportunity to expand their knowledge and find interesting uses of statistics in the world around them.

To track student interest levels I ask the class to complete minute papers at various points during the semester. These papers, which are completed at the end of a class period and can be turned in anonymously, are designed to inform me how I can improve the course or identify concepts that need further explanation. Students are also asked again at the end of the semester to rate their interest level in statistics, revealing how their views have changed. For example, typically I've found 10 to 20% of the students have an initial interest level of four or higher, but over 50% do at the end of the semester. I can also compare these numbers across semesters. Thus, I have very quick and efficient ways of tracking student opinions during the course, allowing immediate adjustments.

#### **4. Essential Features of Effective Assessment Techniques**

After implementing the above strategies in various stages, I have found the following guidelines to be essential to the success of the evaluation process.

##### **4.1. Provide Students with Timely, Constructive, Regenerative Feedback**

Assessment techniques should provide a dialogue between the instructor and the student. The feedback needs to identify the problem to the student and provide guidance for how he or she may proceed. On homeworks, instead of simply telling students the correct answers, I try to show them how their

solutions are insufficient and guide them towards an improved solution. When I use student graders, I also distribute homework solutions for further review. One student explained, "The solutions handed out were more helpful to determine how to fix what I'd done wrong."

Students also need to be given the opportunity to reevaluate their work and receive additional feedback. The journals are an excellent way to maintain ongoing discussions with the students. However, I find that many students may not naturally take the opportunity to resubmit earlier work. Conversely, the project reports have worked well to continually monitor students' efforts and provide hints at different stages of the process, and students have found them valuable. The lab reports also allow me to provide feedback on data collection issues before the group project has been completed. The "full" write-ups provide the students with lots of practice before they produce the final project report. These approaches reinforce to students that statistics is an evolving process, while allowing me to monitor their methods as they are being developed.

I find peer reviews of student projects quite helpful. These reviews show students what others in the course are doing, and provide additional feedback from their peers. This feedback is sometimes more valuable to the students than the instructor's input. Overall, the student feedback is quite constructive, providing input that is prompt and rich with several new perspectives. These techniques provide the students with information they can use in their next assignment, instead of only giving them a final interpretation of their work at the end of the term.

#### **4.2. Promote Self-Reflection and Higher-Order Thinking**

It is important to develop students' ability to evaluate their own work and challenge themselves. Decreasing the amount of specific instructions in the labs as the semester progresses allows me to measure students' level of self-guidance and independence. The labs have proven more and more successful as I incorporate students' suggestions on how to clarify the instructions. For example, at the students' request, the lab instructions now include an index, glossary, and frequently asked questions section. By minimizing the computer learning time, students are able to concentrate longer on the conceptual points of the lab. I also now hire a temporary student assistant each semester to provide feedback on the lab exercises and the clarity of the instructions while I am developing them. I find this has greatly enhanced my ability to present the material at the appropriate level for the students, so they can focus on the concepts being addressed.

The index, glossary, and frequently asked questions sections also help me to assess the students' ability to help themselves when they encounter a problem. Including these sections has certainly helped to provide students with guidance, but I am still trying to find the proper balance between leading and telling. I could also include more grading information in the lab instructions, but I think it is important for the students to learn to identify the components of an effective statistical analysis or explanation for themselves. I want students to do much of the discovery for themselves, because what students construct for themselves they will understand better and remember longer (see, e.g., Resnick 1987).

Often it is difficult to get students to become more independent learners. For example, with the journals, many students didn't appreciate the opportunity to outline the chapters and summarize the relevant points for later review. I thought students would like the opportunity to receive feedback and credit for review techniques they were already utilizing, but the study habits of my students are not at this level. Still, I think it is advantageous for students to generate their own summaries, and I continue to ask them to develop their own review sheets (which I look over) instead of relying on the instructor. (An overview of research on how students actively construct knowledge is presented in Anderson 1983).

Students also had mixed reactions about the "Other" category of the journals. A few students took the opportunity to expand their knowledge and find interesting examples in the world around them, but most did not. Thus, instead of increasing their opportunity to see the relevance of statistics, the requirement instead tended to add to their anxiety about the course load.

It is important to put students in situations that require the skills you are trying to assess. For example, I use the take-home component of the final exam to allow students to show me what they can produce independently. By putting them in the role of a statistician, I challenge them to develop a solution to a new problem on their own. I encourage them to ask questions about the computer commands, as this is not my main focus, and I want them to be able to move on with the problem. However, if they ask for statistical knowledge, they are "charged" by losing points if I tell them the answer. I indicate that this is similar to having to pay an external consultant when they need statistical expertise. Usually, if I tell them a question will cost them, students accept the challenge to think more about the problem. I hope by offering myself as a resource I also reduce unpermitted collaboration. Students are given the freedom to demonstrate to me how they can use the knowledge they have gained in the course.

#### **4.3. Provide Students with Guidelines of What is Expected**

Almost every time I have implemented a new technique, I have erred in not providing the students with enough guidance as to what I expect. This is primarily because I have not fully developed the ideas ahead of time. Students have commented that they "did not always know what to write about the different activities" or "what the point was." To address these difficulties, I now give quizzes to preview exam questions, include project checklists in the syllabus, supply a detailed lab manual, and make model papers available for review.

For example, initially my in-class exams were slightly different from what students were expecting because the exams focused as much on the conceptual topics in the course as on the calculations. Now I give biweekly quizzes, partly for evaluation, but also to introduce students to the types of questions that may appear on an exam. The quiz questions are quite conceptual in nature, typically requiring more thought than the homework problems. I alleviate some test anxiety with this approach, and indicate to students the topics I find most important. I have found this to be sufficient exposure to the exams, and do not provide exams from previous semesters for study.



For the projects, I attempt to incorporate the suggestions made by Sevin (1995): give clear, detailed instructions; provide feedback and guidance. For example, I give students a checklist in their syllabus explaining what needs to be discussed in the final project report. A clear timeline for the periodic project reports is also included. Students are also given access to a web page of previous project topics to review at their convenience. This gives the students more of a framework for the project, and encourages them to extend themselves beyond what was done previously.

The largest initial problem when developing lab exercises was the students' computer anxiety and lack of experience. I now produce a lab manual that is given to the students at the beginning of the semester (as part of a lab fee paid with the course). The manual, modeled after those by Egge, Foley, Haskins, and Johnson (1995) and Spurrier, Edwards, and Thombs (1995), includes instructions for the 14 lab assignments, sections on how to use the computers, Microsoft Word, and Minitab, as well as the index, glossary, and frequently asked questions sections. The manual provides detailed computer steps, including pictures of the menus they will see, for all of the labs throughout the semester. Guidelines for the full lab write-ups are also included (similar to Spurrier et al. 1995). Thus, the manual provides students with a constant reference and knowledge of what is expected of them.

Each week, a lab write-up that received one of the highest grades is posted in the lab and on the course web page. The goal is for students to review these papers, feel pride if their paper is selected, see the work of their peers (and thus what they are also capable of), and notice that there can be multiple interpretations and conclusions for the same data. I purposely post these materials after the lab is due, so that students can utilize the feedback for their next lab but cannot use the papers as templates for each lab. Unfortunately, less than 20% of students indicated that they reviewed these model papers. In fact, three students indicated that they were not aware of their existence. Perhaps this is one reason students still felt there wasn't enough guidance on what was expected in the lab write-ups. In the future, I need to increase student awareness of this resource.

Lack of guidance also appeared to be a problem with the journals. A definitive set of guidelines for the journals, similar to those for the projects and labs, given at the beginning of the semester, would probably help significantly. In the future, I will be more clear about which topics will be included in the journals and what I expect from students. For example, the journal questions are now distributed as handouts, with more focused questions to be addressed, rather than requiring merely a simple summary. These handouts also include graphs of data collected in class for students to reflect on and utilize to support their conclusions. I now also include an initial practice example and distribute a model explanation of the activity.

In short, students need to be given enough direction to start the problem, but not so much that their approach becomes rote. Evaluation techniques should give students the ability to continue to learn and develop, while being flexible enough to allow students to develop in their

own ways.

#### **4.4. Make Sure the Assessment is Consistent and Fair**

Students always know if they are getting different grades for the same work and may discount feedback if they don't feel it is valid. This is especially a concern to me because grading the labs and projects can be time consuming (approximately 15 to 20 minutes for full lab write-ups, 30 minutes for projects). Development of point systems (how much is taken off for each type of mistake) and scoring rubrics has proven quite helpful for consistency, and can be reused in subsequent semesters. The point system can also be revealed to students if they desire additional information on the grading process.

For example, I use a breakdown sheet for the projects each semester: 20 points for oral presentation, 20 points for the quality of the write-up, 20 points for the data collection design, 20 points for choice and correctness of the analysis, and 20 points for interpretation of the results. During the project presentations, students rate the following features on a five point scale: introduction, organization, clarification, visual aids, and conclusions. Comments and suggestions from the class are summarized into a final feedback report for the group. Such sheets have helped my consistency from semester to semester and have clarified to the students what I am looking for.

For the journals, I grade the explanations of in-class activities on a three point scale with 0 = incomplete, 1 = severe misunderstanding, 2 = partially correct, 3 = correct (Garfield 1993). Half points are often assigned. This point system allows me to assign grades more quickly, but such totals often need to be adjusted to assign letter grades. Furthermore, a three was seldom given initially, as I hope to encourage further reflection. However, if students don't resubmit their work, the grading can seem harsh to them.

Most of the student complaints about lack of fairness came with the journals. While several students expressed appreciation in the course evaluations that their questions were answered personally and in detail, others felt their grade shouldn't be lowered because they did not have questions. Evaluating the self-evaluations also proved problematic. I tried to make it clear that these self-evaluations are different from the in-class minute papers, in that students should reflect on their efforts (instead of mine). I gave full credit if I felt the student was being constructive and insightful. However, students didn't see how I could grade their opinions. Again, more clarity is needed regarding my expectations.

#### **4.5. Make Sure Students Understand the Goals of the Evaluation**

For the assessment technique to be informative, students need to understand why it is being used and why it is important. Otherwise the evaluation will probably not be an accurate reflection of their abilities, and their anxiety about the course may increase.

Initially, students saw the lab activities and journals as time consuming and not adding to their knowledge. Of the students working with journals,

66% found them too time consuming, with 41% pinpointing the chapter summaries as the most time consuming component. One of the main problems with my implementation of the journal technique was not explaining to students the concept of the journal. Students didn't really understand what I meant by "journal," thinking of a diary that had to be completed daily. Furthermore, students did not understand how the journals assessed a component of their learning that is separate from the traditional homework assignments. One approach I may adopt is to have the journals substitute for the homework assignments for one week or at the end of a large block of material, but feasibility of this may depend on the pace of the course. I will also encourage students to keep their journal entries on the computer so that later revision will be less troublesome. Journal entries can also be submitted through e-mail.

Some students don't see the data collection and writing components of the course as "worth the effort." I need to further emphasize why these tasks are being asked of them and what I want them to get out of the experience. This will help the students concentrate on expressing what they have learned instead of dwelling on how long the task is taking. Some students have expressed appreciation, though perhaps reluctantly, for the gain in understanding of the material, despite the additional work:

"I hated doing them (journals), but I realize now that they really helped in learning the concepts."

"(I didn't like) writing about the chapter summaries, but it did help me get some of the concepts down."

Furthermore, students need to realize how each assessment tool is different. This was evident when journals were first introduced:

"You already provided most of the service the journal was meant to be for."

"If I have a problem, I would go straight and ask you and you can judge my understanding by looking at my homework and lab assignment."

Comments like this illustrate why students did not easily accept the journal component. Conversely, students can more easily see the purposes of the take-home component of the final exam and the lab activities.

I believe additional clarification of the distinct goals of the different assessment activities will increase student participation. Students will better appreciate what they are learning from the activities, and the assessment will more accurately reflect students' abilities.

#### **4.6. Make Sure the Assessment is Well Integrated Into the Course**

For evaluation procedures to be successful, they need to be well organized and complementary to the lecture material. All assessment components should be described at the beginning of the course. Consideration also needs to be given to the amount of time required by the students and the instructor for each component. If too many tasks are assigned, neither

the students nor the instructor will be able to devote sufficient energy to each task, and the speed and consistency of the assessment will decline.

Initially, labs and journals were not well received because of my lack of organization and the need for better integration of the journal and lab assignments with traditional homework assignments. I find I have more success when I make of point of briefly mentioning the labs and the projects during the lecture time. These reminders are helpful to illustrate to the students when and where the ideas fit in. I have also been able to better time the lab exercises to be concurrent with lecture topics, although this can lead to a rather rigid syllabus. It is important for students to see that all these ideas are related, and not just several disjoint tasks.

Students will also resent an assessment activity if they feel the time required is excessive.

"The labs didn't really help because I was too busy trying to get it done that I didn't learn anything."

Students claim that a lab requiring 15 minutes of observation outside of class is unreasonable. They prefer smaller tasks or data collected in class. The lab activities became more successful after I introduced the separate lab hour into the course and incorporated more of the data collection into the lecture time. During the lab, half of the class meets at a time. Students work in pairs on the lab assignment, with the instructor available for immediate assistance and clarification. Students appreciate having access to the instructor and seem able to approach the computer with higher efficiency and less anxiety. When designing lab assignments, I plan them so the computer work can be completed during this lab hour, with the students' explanations and interpretations being incorporated on their own during the week. I also reduced the number of full lab write-ups to five of the 14 labs. These changes allow students time to absorb the ideas and to incorporate ideas from lecture into their lab reports. Reducing the number of full lab write-ups and encouraging students to work in pairs also reduces the grading load.

Furthermore, the assessment strategies need to complement each other. I think this is accomplished with the activities I am currently using, providing me with a more complete picture of student understanding and performance. Each technique gives me different sources of information, whereas, as Garfield (1994) states, a single indicator would limit the amount of information I receive. Still, too many competing assignments can overwhelm the students or seem too repetitive. I must be careful not to use so many distinct indicators that they overload the students and are no longer effective. Currently, I do not require a separate journal assignment, but I do integrate some of the questions into the homeworks. While undergraduates grade most of the homework problems, I grade these more open-ended questions. This approach achieves many of the journal goals. I also include exam questions that directly relate to the main ideas I want the students to learn from the labs and in-class activities. (For example, Scheaffer et al., 1996, include assessment questions after each activity.) This reminds students to reflect on the learning goals of these activities.

I concur with Garfield's (1994) suggestion to only attempt to incorporate

one new assessment technique at a time. These techniques can be quite time consuming and demanding on the instructor, and need to be well organized and thought out ahead of time. Both the instructor and the administration need to be cognizant of the additional demands required by these techniques. Successful implementation will not be instantaneous, but will require continual review and redevelopment. My current system of weekly homework, biweekly quizzes, weekly labs, and a semester-long project seems to be at the upper limit of what I can ask of my students and of myself.

## 5. Summary

I have found that traditional assessment techniques often do not sufficiently inform me as to whether my students possess the skills I desire. For example, at what level do they understand the material? Will they be able to read and critique articles in their field? Can they read and debate statistical evidence? Can they make reasoned arguments with data and explain their ideas to others? Can they read and interpret computer output? Can they use what they have learned? I have found the techniques discussed here to be quite effective in expanding my knowledge base about the students, including where they are struggling and what topics need to be reviewed. Still, such techniques should be implemented gradually and with careful planning. For example, I did not find the journals added significantly to the techniques I had already implemented, but instead created an excessive workload. However, I think journals could successfully be substituted for more traditional homework assignments or in situations where more individualized assessment is not otherwise possible. At the university level, we often have tremendous flexibility in the types of assessment we can incorporate into a course. Furthermore, as students become accustomed to such approaches, we will be able to more easily expand their use. I find the above techniques to be excellent tools for informing me about what students know and can do. At the same time, these techniques tell me whether students appreciate statistics and computers. They also allow me to provide prompt feedback that is meaningful to the students. Such information is crucial to both the instructor and the students, and is much richer than that provided by traditional techniques.

---

## Appendix - Current Assessment Components

Math 37 is a four unit course with three 80-minute lectures per week and one 50-minute lab per week for 15 weeks. Sections are limited to 45 students. The text is Moore and McCabe (1993). The following assessment components are incorporated in Math 37:

- \* Two in-class midterms (15% each)
- \* Final exam (15%)  
85% from in-class exam, 15% from take-home exam distributed one week before traditional in-class final. Students work individually on distinct questions. Sources of data: ASA datasets, Journal of Statistics Education, casebooks (e.g., Chatterjee, Handcock, and Simonoff 1995).

- \* Weekly homework assignments (15%)  
Due on Friday. Randomly selected students present "practice problem" solutions to the class on Wednesday.
- \* Weekly computer labs (15%)  
Students meet for one hour per week in computer lab with instructor and have one week to complete lab write-up. Sections are limited to 23 students. Lab manuals are handed out the first week. Students complete nine 25-point labs and five 50-point labs ("full labs"). The full lab write-ups are graded on presentation (10%), computer output (40%), and interpretation of results (50%).
- \* Quizzes (10%)  
Eight to ten quizzes given, lowest two grades dropped.
- \* Term Project (15%)  
Groups of 3 to 5 work together on one project over the semester. Projects are graded on data collection design (20%), analysis (20%), interpretation (20%), written report (20%), and final oral presentation to the class (20%). Grade given is 85% group grade and 15% individual grade. Previous project topics can be viewed at <http://www.uop.edu/cop/math/chance/projects.html>

Groups submit four periodic project reports:

- o Report 1 (Week 4): Topic, population, type of study, sampling frame.
- o Report 2 (Week 6): Data collection with copy of design or survey. Submitted for peer review.
- o Report 3 (Week 11): Preliminary descriptive statistics, goals for analysis.
- o Report 4 (Week 12): Final selection of statistical tests.
- o Rough Draft (Week 14): Optional.
- o Final Report and Presentations (Week 15)

In the peer reviews students are asked to review the proposed study, indicating where the study is lacking in clarity or has potential biases. The instructor looks over the reviews and returns them to the groups.

- \* If journals are used, they are worth one-half of the homework grade. A 3-point scoring rubric is used to assign points for write-ups on ten in-class activities, and a 2-point rubric is used for ten chapter summaries. There are 10 points total for self-evaluations, 10 points total for questions on the material, and 10 points total for finding "other" applications. Journals are collected every 2 to 3 weeks.

## References

Anderson, J. R. (1983), *The Architecture of Cognition*, Cambridge, MA: Harvard University Press.

Archbald, D. and Newmann, F. (1988), *Beyond Standardized Testing: Assessing Authentic Academic Achievement in the Secondary School*, Reston, VA: National Association of Secondary School Principals.

Chatterjee, S., Handcock, M., and Simonoff, J. (1996), *A Casebook for a First Course in Statistics and Data Analysis*, New York: John Wiley and Sons.

Cobb, G. (1992), "Teaching Statistics," in *Heeding the Call for Change: Suggestions for Curricular Action*, ed. L. Steen. MAA Notes, No. 22, Washington: Mathematical Association of America, pp. 3-34.

Crowley, M. L. (1993), "Student Mathematics Portfolio: More Than a Display Case," *The Mathematics Teacher*, 87, 544-547.

delMas, R. (1996), "A Framework for the Development of Software for Teaching Statistical Concepts," *Proceedings of the 1996 International Association of Statistics Education (IASE) Roundtable*, Granada, Spain.

Egge, E., Foley, S., Haskins, L., and Johnson, R. (1995), "Statistics Lab Manual," Carleton University, Mathematics and Computer Science Department, 3rd edition.

Fillebrown, S. (1994), "Using Projects in an Elementary Statistics Course for Non-Science Majors," *Journal of Statistics Education [Online]*, 2(2). Available by e-mail: [archive@jse.stat.ncsu.edu](mailto:archive@jse.stat.ncsu.edu)  
Message: send jse/v2n2/fillebrown

Gal, I. and Ginsburg, L. (1994), "The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework," *Journal of Statistics Education [Online]*, 2(2). Available by e-mail: [archive@jse.stat.ncsu.edu](mailto:archive@jse.stat.ncsu.edu)  
Message: send jse/v2n2/gal

Garfield, J. (1993), "An Authentic Assessment of Students' Statistical Knowledge," in *National Council of Teachers of Mathematics 1993 Yearbook: Assessment in the Mathematics Classroom*, ed. N. Webb, Reston, VA: NCTM, pp. 187-196.

---- (1994), "Beyond Testing and Grading: Using Assessment to Improve Student Learning," *Journal of Statistics Education [Online]*, 2(1). Available by e-mail: [archive@jse.stat.ncsu.edu](mailto:archive@jse.stat.ncsu.edu)  
Message: send jse/v2n1/garfield

Mackisack, M. (1994), "What is the Use of Experiments Conducted by Statistics Students?" *Journal of Statistics Education [Online]*, 2(1). Available by e-mail: [archive@jse.stat.ncsu.edu](mailto:archive@jse.stat.ncsu.edu)  
Message: send jse/v2n1/mackisack

Moore, D. S. and McCabe, G. P. (1993), *Introduction to the Practice of*

Statistics (2nd ed.), New York: W. H. Freeman.

National Council of Teachers of Mathematics (1989), Curriculum and Evaluation Standards for School Mathematics, Reston, VA: NCTM.

----- (1993), Assessment Standards for School Mathematics, Reston, VA: NCTM.

Resnick, L. (1987), Education and Learning to Think, Washington, DC: National Research Council.

Roberts, H. V. (1992), "Student-Conducted Projects in Introductory Statistics Courses," in Statistics for the Twenty-First Century, eds. F. Gordon and S. Gordon, MAA Notes, No. 26, Washington: Mathematical Association of America, pp. 109-121.

Rosenshine, B., and Meister, C. (1994), "Reciprocal Teaching: A Review of the Research," Review of Educational Research, 64(4), 479-530.

Scheaffer, R., Gnanadesikan, M., Watkins, A., and Witmer, J. (1996), Activity-Based Statistics, New York: Springer-Verlag.

Sevin, A. (1995), "Some Tips for Helping Students in Introductory Statistics Classes Carry Out Successful Data Analysis Projects," presented at the Annual Meeting of the American Statistical Association, Orlando, FL.

Snell, J. L., and Finn, J. (1992), "A Course called `Chance'," Chance, 5(3-4), 12-16.

Spurrier, J. D., Edwards, D., and Thombs, L. A. (1995), Elementary Statistics Lab Manual, Belmont: Wadsworth Publishing Co.

-----  
Beth L. Chance  
Department of Mathematics  
University of the Pacific  
Stockton, CA 95211  
bchance@uop.edu



## 8.4 Other Instructional Resources on the Web:

### THE INTERNET: A NEW DIMENSION IN TEACHING STATISTICS

J. Laurie Snell, Dartmouth College (U.S.)

From: Garfield and Burrill (editors) *Research on the Role of Technology in Teaching and Learning Statistics*, Voorburg, The Netherlands: International Statistical Institute, 1997.

#### INTRODUCTION

The Internet was developed to permit researchers at universities and colleges to freely communicate their ideas and results with others doing similar research. This was accomplished by connecting the universities by an electronic network called the **Internet** and providing a method for sending messages called **Electronic Mail** or **E-mail**.

E-mail was fine for simple text messages. However, transmitting results of research often requires the capability of transmitting formulas, graphics, and pictures and occasionally even sound and video. Tools to accomplish this were developed and the Internet, with these capabilities, was called the **World Wide Web** or, more simply, the **Web**. Instead of directly transmitting this more complex information between two researchers, say John and Mary, it was decided to allow John to deposit his results on a machine at his institution and let Mary obtain them from this machine. This made John's results not only accessible to Mary but also to anyone else in the world who had access to the Web. This resulted in a remarkable new way to share information. Common usage now uses the terms Web and Internet interchangeably and we shall use the term Internet.

The enriched Internet was such a success that it was extended to allow the same kind of transmission of information by the general public and industry.

While the Internet has grown to have all of the best and worst elements of our society, it is still a wonderful way to achieve its original goal: to allow academics to freely share information.

The original e-mail still works very much like it did in the beginning and continues to be a natural and useful way to communicate. When we write a letter, we imagine this letter may be kept as a permanent record of our thoughts. For this reason, most of us take some care in the way we express our thoughts in a letter. E-mail is much more informal-- it is not a sin to misspell a word or make a grammatical error. You usually are just writing to ask someone a technical question, help a student, give a colleague a reference to a paper, etc. Most of the time, when you receive an e-mail message, you reply and never again look at the message again.

Somewhat the same philosophy has been applied to putting materials on the Internet. People often put on their Web site their first thoughts on an issue somewhat like a first draft of an article or a book. However, unlike e-mail, this material stays where it is put and can be viewed by anyone in the world. Thus if you just start looking around randomly on the Internet, much of what you see is outdated and you may get pretty discouraged by the quality of what we find. Thus it is important to find ways to help identify interesting material. In this paper we hope to do this in the area of statistical education. We shall see sources where useful information is shared on the Internet in the form:

- Course descriptions and materials for teaching a course

- Data sets
- Articles and background information on current statistical issues in the news from newspapers such as *The New York Times* and *The Washington Post*, radio programs such as National *Public Radio* and popular science journals such as *Science*, *Nature*, and *The New England Journal of Medicine*.
- Interactive modules and texts that illustrate basic concepts of statistics and can be run from the Internet.
- Electronic journals such as *The Journal of Statistics Education*

Methods for putting materials on the Internet are constantly being improved, and the materials you can find there are constantly changing to take advantage of new technology. As this is written the rate of transmission is not sufficient to permit the user to see more than a minute or two of video material but soon this will change.

In this period of development not all will be smooth sailing when you attempt to use the Internet. You may find, just as you are about to use the Internet in a class, your network is down or the speed of transmission is too slow. With new applications you will have to learn how to configure your software to accommodate them. Also materials that are on the Internet today may have been moved or removed altogether by the time you want to use them. Thus we cannot guarantee that everything to tell you at the time of our 1996 conference will be available when you read this. However, we can assure you that what you will find will be even more exciting than what we will tell you about here.

## **THE CHANCE COURSE**

Chance is a course designed to make students better able to read and critically assess articles they read in the newspapers that depend upon concepts from probability or statistics. Chance was developed cooperatively by several colleges and universities in the United States: Dartmouth, Middlebury, Spelman, Grinnell, University of California San Diego, and the University of Minnesota. In the Chance course we discuss basic concepts of probability and statistics in the context of issues that occur in the news: clinical trials, DNA fingerprinting, medical testing, economic indicators, statistics of sports, etc. We use the Internet to provide resources for teaching a Chance course.

To assist in teaching a Chance course we provide an electronic newsletter called *Chance News* that abstracts current issues in the news that use probability or statistical concepts. This newsletter is sent out about once every three weeks. Anyone interested in receiving it can send a request to [jlsnell@dartmouth.edu](mailto:jlsnell@dartmouth.edu).

In addition, we maintain a Chance Database on the Internet<sup>1</sup>. We keep, in this database, syllabi of previous courses we have taught and links to courses others have taught. We also keep description of activities, data sets and other materials we have found useful in teaching a Chance course as well as a teacher's guide for a Chance course. You will also find here the current and previous issues of *Chance News* with links to the full text of most of the newspaper articles.

## **USING THE INTERNET TO TEACH CHANCE**

In this section we will illustrate how we use the Internet in teaching the Chance course. We will discuss the following uses of the Internet:

---

<sup>1</sup><http://www.geom.umn.edu/locate/chance>

- e-mail communication between students and instructors
- posting daily handouts on a Internet site
- posting class data on a Internet site
- finding articles from the popular media
- finding additional information on articles
- gathering of information and data for student projects

We illustrate these uses in the context of our most recent Chance course taught at Princeton in the Winter term 1996. The Princeton students all use e-mail, and we used it throughout the course to help the students with questions on homework, to arrange appointments, and to send other information about the course to the students.

Our classes follow the following format: we choose an article that has appeared in the news, usually a very recent article, and prepare some questions relating to this article. The students divide into groups of three or four, read the article, and discuss our questions for about twenty minutes. They then report their conclusions and we use these as the basis of additional discussion of the article and the statistics and probability issues suggested by it. Occasionally, instead of discussing a current article, we ask the students to carry out, in their groups, an activity designed to help them understand statistical concepts that came up in their reading. For example, when the notion of hypotheses testing came up, we asked them to identify one member of their group who believes that he or she can tell the difference between Pepsi and Coke and to design and carry out an experiment to test this claim.

We put these class handouts on our Internet site. This means that, if a student missed a class or lost a handout, there was no problem getting another copy. This also allows teachers at other schools teaching or interested in teaching a Chance course to see exactly how we do it and makes it easy for us to use some of the materials in a later offering of the course. We hope that others teaching a Chance course will share their materials on the Internet. One who has already done this is Professor Nancy Reid at the University of Toronto. She keeps a complete account of every class, including articles discussed and activities carried out on her Internet site<sup>2</sup>. She uses articles from the Canadian newspapers so this gave us another source of interesting articles for our course.

We started our Princeton course with an activity. We asked the students to help us design a questionnaire to give them statistical information about the class: intended major, number of hours they watch television, etc. We then administered the questionnaire, tallied the data, and sent it to the students by e-mail. We asked them to get acquainted with the statistical package we were using (JMP) by exploring this dataset. We discovered that they had a hard time moving the data from their e-mail to machines on the lab where JMP resided, so we put the data on our Internet site and the students had no trouble downloading the data from there. We coordinated our efforts with Tom Moore at Grinnell who was teaching an elementary statistics course and gave the same questionnaire to his class. We put the results of both surveys on our Internet site allowing students at either school to make comparisons between the students at the two colleges.

Of course we asked the students to read Chance News on the Internet and we used Chance News to help us pick articles for class discussion. We illustrate this in terms of the March issue of Chance News: Here are the articles abstracted in this issue.

## **Chance News 28 February to 28 March 1996.**

### **Contents**

---

<sup>2</sup><http://utstat.toronto.edu/reid/>

- 1. In a first, 2000 census is to use sampling.
- 2. Are all those tests really necessary?
- 3. The use of IQ tests in special education.
- 4. The expected value of Windows Vegas Solitaire.
- 5. A treatment for cancer risks another.
- 6. Is Monopoly a fair game?
- 7. Hawking fires a brief tirade against the lottery.
- 8. On Isle Royale wolves recover from disaster.
- 9. Silent sperm.
- 10. Evaluation of the military's ESP program.
- 11. How safe are Tylenol and Advil?
- 12. Neyer's stats class back in session.
- 13. Fetal heart monitor called not helpful.
- 14. Intelligence: knowns and unknowns.
- 15. HMO prescription limit results in more doctor visits.
- 16. Radioactive implants effective for prostate cancer.
- 17. Unconventional wisdom: Love, marriage, and the IRS.
- 18. Unconventional wisdom: majoring in money.
- 19. Why does toast always land butter-side down?
- 20. Ask Marilyn: Which tire?

We discussed several of these articles in our course. For example, here are the discussion questions we used for the article "Silent Sperm."

### **Class 19: Sperm Count**

#### **Discussion**

Read the article "What's wrong with our sperm?" by Bruce Conley et. al., *Time Magazine* 18 March 1996, p. 78 and the "Sperm counts: some experts see a fall, others see bad data" by Gina Kolata, *The New York Times*, 19 March, 1996, C10.

- (1) What are some of the differences in the way the two articles address this topic? Which do you think gives the better description of the problem?

- (2) What are some of the problems of meta-analysis (combining data from past and present studies) in order to decide whether sperm count is declining? What factors should you control for?
- (3) How would you design a study to test the hypothesis that sperm counts are declining?
- (4) *The New York Times* article cites Dr. Sherins as saying that there is no evidence that infertility is on the rise in the United States. If this is so, why worry about sperm count?
- (5) If infertility is on the rise, what might be the reasons?

For the next class, we asked the students to read the original research articles that were the basis for the newspaper articles. To discuss these papers it would have been a great help to have the raw data for the studies. For example, it was obvious from the results given in the paper that sperm counts are not normally distributed. The authors suggested that the logarithms of the sperm counts are. We would have liked the students to be able to check this and further explore the data after reading the article. We tried to contact the authors, but the relevant person was on vacation. We hope that researchers will begin to make their data available on the Internet.

We also discussed the article on the census 2000. This article was about the decision of the Census Bureau to use some sampling in this survey rather than just enumeration. Here we were helped by being able to query researchers at the Census Bureau by e-mail about their plans. We also found, on the Internet, an article by David Freedman relating to his research on some of the difficulties in implementing the methods under consideration by the Census Bureau for the 2000 census.

The last article in the March Chance News also shows how the Internet can enrich a discussion of a topic in the news. This story starts with the Marilyn vos Savant column in *Parade Magazine*, 3 March 1996 where we find:

A reader writes:

My dad heard this story on the radio. At Duke University, two students had received A's in chemistry all semester. But on the night before the final exam, they were partying in another state and didn't get back to Duke until it was over. Their excuse to the professor was that they had a flat tire, and they asked if they could take a make-up test. The professor agreed, wrote out a test and sent the two to separate rooms to take it. The first question (on one side of the paper) was worth 5 points, and they answered it easily. Then they flipped the paper over and found the second question, worth 95 points: 'Which tire was it?' What was the probability that both students would say the same thing? My dad and I think its 1 in 16. Is that right?

Marilyn answers that the correct probability is 1/4 and explains why. We found, on the Internet, an earlier account of this incident indicating that the professor was a chemistry professor at Duke University named Bonk. A check on the Duke homepage revealed that there was, indeed, a chemistry professor at Duke named Bonk. We sent an e-mail message to Professor Bonk and got the following reply.

Laurie,

The story is part truth and part urban legend. It is based on a real incident and I am the person who was involved. However, it happened so long ago that I do not

remember the exact details anymore. I am sure that it has been embellished to make it more interesting.

J. Bonk

Professor Bonk included an e-mail message he had received from Roger Koppl, an economist at Fairleigh Dickinson University who writes:

When I read the story of Professor Bonk, I thought immediately of the right front tire. I was then reminded of something economists call a "Schelling point," after the Harvard economist Thomas Schelling. Schelling had the insight that certain places, numbers, ratios, and so on are more prominent in our minds than others. He asked people to say where they would go to meet someone if they were told (and knew the other was told) only the time and that it would be somewhere in New York. Most chose Grand Central Station. How to divide a prize? 50-50. And so on. The existence of these prominent places and numbers and such permit us to coordinate our actions in contexts where a more "pure" and "formal" rationality would fail. These prominent things are called "Schelling points."

Professor Koppl goes on to describe a survey he carried out that verified that the right front tire would be the most popular answer to the question: If I told you that I had flat tire and asked you to guess which tire it was, what would you say? Another e-mail writer stated that he had consulted a tire expert and was told that, in fact, the most likely place to get a flat tire is the rear right tire.

Thus, thanks to the wonders of e-mail, a routine probability problem brought out the complexities of applying probability theory in the real world. It also provided an introduction to the interplay of probability and psychology and led naturally to our discussion of the work of Kahneman and Tversky.

As this example shows, e-mail provides a good way for the instructor and the students to obtain additional information about a topic that might be rather briefly described in the newspaper. Another source is research articles posted on the Internet. For example, an article in *The New York Times* discussed a debate on the reliability of an estimate for the number of Internet users obtained by Nielsen using a telephone survey. Two market researchers, who had helped plan the Nielsen study, disagreed with the way Nielsen handled the data and made available on the Internet a paper<sup>3</sup> where they explained how they arrived at quite different numbers from the same data.

The students in our Chance course carry out a significant project to be presented in poster form at the Chance Fair at the end of the course. The Internet was a great help to the Princeton students working on their final projects. Inspired by the "upset" of Princeton over UCLA in the 1996 NCAA College Basketball, two students were interested in seeing how often such upsets occur. To do this they needed the seedings of the teams in the previous tournaments. They easily found this information on one of the basketball Internet sites.

Another student wanted to analyze lottery data consisting of numbers that people actually chose for a lottery. Calls to the state lottery offices led nowhere. Officials gave all kinds of reasons why they could not release data of this kind. However, a few e-mail messages at addresses found on the lottery Internet pages led to a lottery official who was interested in having such data analyzed and was happy to give the student the data he needed.

The success of this project, and the data obtained, led us to write a module on the use of lotteries in teaching probability in a Chance course. You can find this module on the Chance Database under "teaching aids."

---

<sup>3</sup><http://www2000.ogsm.vanderbilt.edu/baseline/1995.Internet.estimate.html>

Another interesting project that made good use of the Internet dealt with weather prediction. The students wanted to know, for example: How is the probability of rain determined and what does it mean? Are weather predictors rewarded according to the quality of their predictions? To help answer such questions they made a small survey and sent it, by e-mail, to a number of weather forecasters. They got good responses and learned a great deal from the answers to their questions.

We next describe some other sources on the Internet that are of interest to statistics courses generally.

## THE JOURNAL OF STATISTICS EDUCATION

The *Journal of Statistics Education* (JSE) is a refereed electronic journal that deals with post secondary statistics education. The managing editor is Tim Arnold and the editor is E. Jacquelin Dietz. The first issue was published on July 1, 1993.

The current issue (Vol. 4, No. 1 1996) has an article by Maxine Pfannkuch and Constance M. Brown entitled "Building on and Challenging Students' Intuitions About Probability: Can We Improve Undergraduate Learning?" It is well known that students intuitive ideas of determinism, variability, and probability are often not in agreement with formal probability. The authors observe that the proper understanding of the role of probability in statistical reasoning about real world problems requires a resolution of these differences. They report on a pilot study to identify some of the differences in intuitions and formal concepts that students have and to determine how these differences can be resolved.

The JSE has two regular departments. The first, "Data Sets and Stories," is edited by Robin Lock and Tim Arnold. Readers are invited to submit an interesting data set along with a "story" describing the data. This story includes the source of the data, a description of the variables and some indication of the statistical concepts that are best illustrated by the data set. The data is put in a form that makes it very easy to download to any statistical package. Each article of the JSE features a description of one or more of these data sets, but the entire collection of data sets can be considered a part of the journal. The data set in the current issue of JSE was supplied by Roger W. Johnson. This data set allows the student, using multiple regression, to estimate body fat for men using only a scale and a measuring tape. Johnson describes his experiences using this data set in his classes.

The second department, "Teaching Bits" is edited by Joan Garfield and William Peterson. Garfield provides abstracts for articles in other journals of interest to teachers of statistics including abstracts for the articles in the British journal "Teaching Statistics." The current issue features abstracts edited by Carmen Batanero that appeared in the January 1996 *Newsletter of the International Study of Group for Research on Learning Probability and Statistics*. Peterson provides abstracts of articles in current journals and newspapers that use statistical concepts to serve as the basis for class discussions similar to those we have already seen in *Chance News*.

One important feature of an electronic journal such as the JSE is that it is possible to search through all previous issues for articles on a given topic. For example, if you are interested in assessment you need only put "assessment" into the JSE search mechanism. You will find, for example, a paper by Iddo Gal and Lynda Ginsburg that "reviews the role of affect and attitudes in the learning of statistics, critiques current instruments for assessing attitudes and beliefs of students, and explores assessment methods teachers can use to gauge students' disposition regarding statistics.

The *Journal of Statistics Education* is part of the "The Journal of Statistics Information Service." This service provides a number of other interesting resources for statistics teachers. These include archives of discussion groups, free statistical software, and links to other statistical resources.

## ELECTRONIC DISCUSSION GROUPS

There are several discussion groups for statistics. The most relevant discussion group for teaching statistics is Edstat-L (sci.stat.edu) maintained by the JSE Information Service. You can be an active member of this discussion group by e-mail or by using one of several Internet news group readers. The archives of this group are kept on the JSE Information Service. These archives can be searched to bring together the discussion on a specific topic. For example, if you plan to discuss Simpson's paradox and want some fresh ideas, a search of the archives will lead to references for real life examples, connections with regression paradoxes, and a suggestion to use the video "Against all Odds." If you are trying to decide which statistical package to use, a search will produce discussions of the merits of any particular package and compare it to others.

## FINDING YOUR WAY AROUND THE INTERNET

If you are looking for a fairly specific kind of information on the Internet the best way to find it is to use one of the several different search engines that will search on the entire Internet. Our student found the required basketball data by just searching for Internet sites using the word "basketball." It is sometimes not quite this simple. For example, the lottery data was found by first searching on "lottery" to find a variety of homepages that deal with lottery questions and then sending e-mail messages to the "Webmaster" at three of the most relevant sites asking if they knew how to obtain the required data. One of those recommended turned out to be the friendly lottery expert who sent the student the data. The point is that people who maintain Internet pages on a subject tend to know people who are willing to share resources -- after all that is what the Internet is all about.

In most areas, there are Internet sites that try to provide, in addition to their own resources, a guide to other related Internet sites. To get a more general picture of what is available on the Internet for teaching statistics it is useful to go to one of these sites. A good choice is the Internet site of John Behrens at Arizona State University which is called "The Statistical Instruction Internet Palette<sup>4</sup> (SIIP)."

## THE STATISTICAL INSTRUCTION INTERNET PALETTE

The home page of the SIIP Internet site displays a palette with several items to choose from. When you click on a particular palette you are taken to a page that provides links to resources related to the topic of the palette. The palettes represent the different kinds of statistical information that would be useful to a student or teacher of statistics. We describe briefly these palettes.

**Data Gallery:** Clicking on this palette leads you to a menu where you can choose histograms, boxplots, summary statistics and other summary information about data from a large study relating to young people. Links are provided to statistics courses at Arizona State illustrating how these graphics are used in their courses.

**Graphing Studio:** Choosing this palette leads you to an interactive site where you can put in your own data and request graphic representations. For example, you can give data from two variables and ask for a scatter plot.

**Computing Studio:** Here students can learn to compute statistical quantities like mean, median, standard deviation etc. by putting in their own data. They see a step by step calculations with appropriate comments along the way.

---

<sup>4</sup><http://seamonkey.ed.asu.edu/~behrens/siip/>



**Equation Gallery:** It is fashionable to put formulas at the end of chapters or somewhere else out of the way "for those who like formulas." Here the formulas are right up front, and the student need only click on "standard-deviation of a population" to get the formula to compute it.

**Classroom Gallery:** At this palette you are invited to obtain information about statistics classes by clicking on "Classroom" or teaching resources by clicking on the "Teacher's Lounge." If you choose "Classroom" you will find a list of courses that have materials on the Internet. The first one is "Introduction to quantitative methods" taught by Gene Glass at Arizona State. Here you find essentially a text for this introductory statistics course provided in a series of lessons.

In the lessons you find questions for the students to answer. The student can then call up the answers and additional comments. The students are often asked to download a data set to carry out an analysis of some aspect of the data. In the first lesson Glass provides a paragraph about himself and a form for the students to do likewise. He then makes the students responses available to the class on the Internet site.

The second course is "Basic statistical analysis in education" taught by John T. Behrens. Here you will find a discussion of each week's work organized in terms of the questions: "What did we do last week, What are we going to do this week? What did we do last week? Were are we going in the future? How will we get there? On each weeks' page, links are made to other resources such as the data gallery or perhaps to material from another course such as that of Gene Glass.

If you choose "Teachers Lounge" from the palette you are provided links to resources for teaching statistics. These include links to courses at other institutions that use the Internet to keep materials for their classes on the Internet, sites with data sets appropriate for teachings statistics, the Chance database etc.

**Wide World of Internet Data:** This palette describes a large number of sites where data is available. These sources are classified by field and you will find brief comments about what you can expect to find at each site. Some data sites make special efforts to present data in a form that is easy to use in a classroom setting.

We have already mentioned the data sets of the *Journal of Statistical Education*, where each data set is accompanied by a discussion by the person who submitted the data explaining how they used the data set in their teaching. Another example of this is the "The Data and Story Library"<sup>5</sup> (DASL) maintained on Statlib<sup>6</sup> at Carnegie Mellon. These data sets are accompanied by stories telling how the data arose and how they can be used. The data sets are classified by methods as well as subjects. Thus if you are looking for a data set to illustrate the chi-squared test you can easily get a list of datasets appropriate for this. Similarly, if you are interested in examples from economics, you can find these using the subject classification. StatLib itself is a wonderful source of data sets useful for teaching purposes.

You will find at this palette, links to sites that indicate how interactive data will be available in the future. For example, you can find a book on AIDS that has a mathematical model built in that allows a reader to provide data relevant for a particular area or country and the model will compute future estimations for way that AIDS will spread in this location. In another example, you can find the current rate of traffic at any

---

<sup>5</sup><http://lib.stat.cmu.edu/DASL/>

<sup>6</sup><http://lib.stat.cmu.edu/>

point on any of the major highways is Los Angeles at the very time you are looking at this site. You will also find data available by video to show, for example, how a glacier changes over several years.

**Collaboration Gallery.** This palette provides students with their own listserve "Run by and for students learning statistics at all levels and from all fields to share questions, concerns resources and reflections with other students ." It also has a link to a student forum which uses special software to allow a more interactive form of communication.

### WHAT ABOUT THE FUTURE?

What can we expect for the future of the Internet? Just looking at what is already happening we can expect:

- Increased use of the Internet to share teaching and research materials including traditional and new forms of interactive textbooks
- Increased " bandwidth" that makes video materials and full multi-media documents accessible.
- Increased development of computer programs and statistical packages that are run by the Internet browsers.
- Improved methods for paying for materials used on the Internet with resulting commercialization of the Internet.
- Increased interest in developing text material and computer programs that are in the public domain to allow users to freely use them and to contribute to making them better.

We are already beginning to see course notes turning into text books on the Internet. A place to see what the textbook of the future on the Internet might look like is the "Electronic Textbook" provided by Jan Deleeuw on UCLA statistics department Internet site<sup>7</sup>. This book is far from complete, but it has a number of examples to show how standard text materials can be enhanced by special features of the Internet. For example a student reading about correlation can click at the appropriate point and a scatterplot and regression line will appear. The student is then invited to vary the regression coefficient to see, for example, what happens to a scatterplot when the correlation is .7 as compared to .2. Another such demo allows the student to move points in a scatterplot and see what effect this has on the regression line.

These interactive graphics are produced in two different ways. One method uses programs written in Xlist-Stat. This is a statistical package developed by Luke Tierney at the University of Minnesota. It is free and available on the standard platforms: Mac, PC, and Unix. For interactive graphics produced this way, you must first download the Xlist-Stat package onto your machine. This is easy to do. A good description of how to do it can be found on the home page for Statistics 1001 at Minnesota<sup>8</sup>. Of course, many other kinds of computations are possible using the Xlist-Stat language and you can, if you wish, write your own.

Two other interesting interactive projects that use Xlist-Stat are the "Teach Modules"<sup>9</sup> provided by the statistics department at Iowa State University and the "Visual Statistics System"

---

<sup>7</sup><http://www.stat.ucla.edu/>

<sup>8</sup>[http://www.cee.umn.edu/dis/courses/STAT1001\\_7271\\_01.www/](http://www.cee.umn.edu/dis/courses/STAT1001_7271_01.www/)

<sup>9</sup> <http://www.public.iastate.edu/~sts/lesson/head/head.html>

(ViSta)<sup>10</sup> developed by Professor Forrest W. Young at the Department of Psychology University of North Carolina.

Teach Modules provide modules on several important statistical concepts including the central limit theorem, confidence intervals, sample means, and regression. The modules include written descriptions of these basic concepts and suggestions for ways for the student to explore these concepts with the interactive Xlist-Stat programs. They also provide exercises for the students.

ViSta is a much more ambitious project that provides a statistical package which can serve both as a research tool and a learning tool. In its learning mode the student is provided guided tours on how to use the package to explore data sets supplied by ViSta or by the student. ViSta is designed to allow the user to choose the appropriate level of expertise and includes extensive graphical tools. Documentation that will one day be a published book is provided in the friendly Acrobat PDF format.

A second method for producing interactive pictures is by using the language JAVA. The language JAVA permits programs to be written that the Internet browser itself, in effect, runs. JAVA has the advantage that you don't need any additional software, other than the Internet browser, on your machine. Such JAVA programs are called "applets". You will find in the UCLA textbook an applet that illustrates the meaning of confidence intervals. It does this by first asking the student to put in the relevant information: population mean, desired confidence level, sample size and number of confidence intervals. The applet then computes the confidence intervals and plots them as lines so the student can see how these lines vary and verify that approximately the proportion corresponding to the confidence level will include the true population mean.

At the moment, JAVA seems to be gaining in popularity. You can find links to a wide variety of applets on the Chance Database under "teaching aids." Running applets on the Internet is much more chancy than running programs on your machine using Xlist-Stat. The developers could solve this by providing the sources for the applets that would allow users to run them independent of the Internet using an Applet Viewer. However, the concept of freely sharing the language and programs that we find with the Xlist-Stat developers seems not to have developed within the JAVA community.

A third method of computing on the Internet is illustrated by a "power calculator" found at the UCLA site. The power calculator calculates the power of a statistical test when you input the information needed to determine the power. For this computation, the power calculator sends the information it receives back to the UCLA computer which calculates the power and sends the answer back to your machine.

The demos described above were constructed by people working at different Universities who shared them with Jan Deleeuw for his project. The Internet was started as a way to freely distribute information worldwide. The UCLA text book is in this spirit where developers of statistical materials are freely sharing them with the statistical community.

Another project along these lines is an introductory probability book<sup>11</sup> that Charles Grinstead and I have made available on the Internet. This is a traditional book in its present form, but we are working on making it interactive with the use of JAVA. We hope that, by putting our book on the Internet, others will want to make links to parts of our book to assist them in teaching a probability course. For example, we have a treatment of the recent results by Diaconis and his colleagues that seven shuffles are needed to reasonably mix up a deck of cards. This is a quite self-contained unit and would be useful for someone teaching a probability course who would like to include this new and interesting topic. We hope also that they will contribute to improving the book as it appears on the Internet.

The next big improvement in the Internet over standard textbook materials will come soon when it is possible to transmit information fast enough to make audio and video materials

---

<sup>10</sup> <http://forrest.psych.unc.edu/research/ViSta.html>

<sup>11</sup> [http://www.geom.umn.edu/docs/education/chance/teaching\\_aids/probability\\_book/book.html](http://www.geom.umn.edu/docs/education/chance/teaching_aids/probability_book/book.html)

routinely available. Actually, for audio this is already the case. In particular, National Public Radio keeps most of its programs, current and previous, on their Web site<sup>12</sup>. This includes interesting discussions with the researchers who are the authors of studies reported in the news as well as with other experts in the field. It is quite effective to use these in a class to enhance the discussion of the news.

The well known video series "Against all Odds" is currently used in the classroom to supplement text material by showing statistics as it is carried out in the real world. The use of such materials will be greatly improved when they can be integrated into text material on the Internet.

I hope I have convinced you that there are terrific resources on the Internet to enhance the teaching of statistics. Of course, much of this is still in the experimental stage so not everything works as it should, but by the time you read this much of what I have talked about will be working smoothly and new and better things will be in the experimental stage.

**1998 Update:** Not surprisingly, our contribution to the ISAE roundtable discussion was a paper on statistical resources on the internet. Also, not surprisingly, this paper is already out of date, and so we will add here some interesting new internet resources we have found since writing this paper. These are all linked to the Chance Database.

---

WebStat 1.0 Beta

Webster West, Department of Statistics, University of South Carolina

Webster West wrote the elegant applets for illustrating statistical concepts that we have referred to many times. His latest product, Webstat, is a statistical package designed to allow the user to analyze data on the web with the usual graphic tools and statistical tests. It works on any platform and is free.

---

Statlets. Java Applets for statistical analysis and graphics.  
NWP Associates, Inc., Princeton NJ

Like WebStat, Statlets allows you to analyze data on the internet. It provides the standard types of graphical output and statistical tests. You can also download Statlets to your machine and run it locally. The academic version is free and permits data with 100 rows and 10 columns. The commercial version (\$195) permits up to 20,000 rows and 100 columns.

---

HyperStat Online.

David Lane, Departments of Statistics and Psychology, Rice University

This is an introductory-level hypertext statistics book that can be read on the web. Alongside each chapter there are links to related material on the web including demos and related excerpts from other on-line text materials.

---

Introductory Statistics: Concepts, Models, and Applications.

---

<sup>12</sup> <http://www.npr.org/>

David W. Stockburger, Psychology Department, Southwest Missouri State University

This is an introductory text by David Stockburger that is freely available on the web and can be downloaded as a zip file. This book is written for psychology and other behavioral science students with an emphasis on understanding the relation between statistics and models and measurement as a part of modeling.

---

Journal of Statistical Software.  
UCLA Department of Statistics.

This journal publishes software and descriptions of software useful for statisticians. The Journal is peer-reviewed, electronic, and free. Articles are mostly appropriate for statistical research and advanced courses but it is where we learned WebStats.

---

Virtual Laboratories in Probability and Statistics.  
Kyle Siegrist, Department of Mathematical Sciences, University of Alabama in Huntsville

This is an NSF project to develop interactive, web-based modules in probability and statistics. Each module explores a topic by means of expository text, exercises graphics, and interactive applets written in Java. This is an excellent site for those who still want to have their students understand, in a painless way, the probability theory behind some of the basic statistical concepts and tests. Kyle Siegrist, is also the author of "Interactive Probability" Wadsworth 1997 (See Chance News 6.06).

---

Lies, Damn Lies, and Psychology.  
David Howell, Department of Psychology University of Vermont

This is the homepage for a course modeled after the Chance course but adapted for psychology students. This course was taught in the Fall Term 1997.

---

Seeing Statistics.  
Gary H. McClelland, Department of Psychology, University of Colorado at Boulder.

This is an overview of a project to develop an interactive elementary statistics book using Java Applets. This book is under development in conjunction with Duxbury Press. You will find here a discussion of the design of the book and a sample chapter.

---

StatVillage.  
Carl James Schwarz, Department Mathematics and Statistics, Simon Fraser University

StatVillage is a hypothetical city on the web consisting of 128 blocks with 8 houses in each block and designed to permit students to carry out real-life surveys. Students decide on questions they want to ask, choose their sample, click on the houses in their sample, and get the results of their survey. The questions they can ask are ones that can be answered by census information and the

answers they get are based on real census data. We tried it in a class and it was a great success (See Chance News 6.09).

---

GARY C. RAMSEYER'S FIRST INTERNET GALLERY OF STATISTICS JOKES.  
Gary C. Ramseyer, Department of Psychology, Illinois State University

A self-explanatory site!

---

Robin Lock's WWW RESOURCES FOR TEACHING STATISTICS  
Robin H. Lock, Mathematics Department, St. Lawrence University  
Canton, NY 13617 USA  
<http://it.stlawu.edu/~rlock/tise98/onepage.html>

## 8.5 Student Projects

### Notes from Tom Moore:

Student projects are a long-term assignment where students ask a question, devise a plan to collect data to answer the question, collect and analyze the data, and tell me and the class about the results in written and oral reports.

### Why one might assign student projects.

Student projects are not an idea I invented, and although I have used them for the past 15 years I am by no means an expert on them. Partly this stems from my basic naiveté about matters of educational theory. So I'll break this section into two parts:

- My reasons
- Support from the experts

A personal experience sparked my enthusiasm for doing student projects. In 1980 I took a regression course from Professor Ledolter at the University of Iowa. He had us do a project and I analyzed data from that year's University of Iowa men's basketball team, the last year they went to the Final Four. I had a great deal of *fun* doing it and felt it was the best learning experience in an otherwise excellent course. I felt like I learned how to *do* regression when I tackled this project. I recall Professor Ledolter asking me if I thought that the coach would fully appreciate the log transformation. As imperfect as my project was, it was a great learning experience. I think two principles were in operation:

(1) The Fun Principle: Bob Hogg has for years, in very public ways -- including an editorial in the *Des Moines Register*--advocated that our courses need to be fun, and projects inject this feature into the course.

(2) The Do Principle: Projects make students experience more fully what it means to do statistics. They get to see a problem from its inception to the reporting of a data analysis that in part at least answers the question.

The reading I've done of authors more knowledgeable than myself tends to support these principles. The Fun principle is given credence in a paper in the Journal of Statistics Education:

"The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework", by Gal and Ginsburg, *JSE*, v2, n2 (1994)

where the authors delineate the importance of a positive attitude in the student if he or she is to learn effectively.

The Do Principle is at the heart of what is usually called authentic assessment. The paper:

"Beyond Testing and Grading: Using Assessment to Improve Student Learning", by Joan Garfield, *JSE*, v2,n1 (1994)

summarizes the theoretical justification for authentic assessment and also gives many excellent, practical guidelines for implementing various forms of assessment. I highly recommend it. In this paper Garfield gives two guiding principles for good assessment:

- The Content Principle -- assessment should reflect the statistical content that is most important for students to learn, and
- The Learning Principle -- Assessment should enhance the learning of statistics and support good instructional practice.

Projects, it seems to me, adhere to both principles: (1) projects help teach an important content goal of the course -- the overall problem solving process, and (2) Projects do more than help the teacher assess what the student has learned, but they also teach the student statistical principles.

Extended example: In the spring of 1992 (and this date is important), two students, Meredith Goulet and Jennifer Wolfson, wanted to find out if male and female students differ with respect to political knowledge, interest, participation, and philosophy. They proposed this idea to me, we had a short consultation about it, and they decided to do a phone survey using a short questionnaire including a small set of political awareness questions. I suggested they discuss the latter with one of their political science teachers, which they did. They carried out their survey with a simple random sample of 60 Grinnell students, using a very persistent call-back scheme so that they got 59 respondents.

The first question on their 3-question quiz was: "Can you tell me who H. Ross Perot is?" The choices were: (a) Prime minister of France, (b) Director of the National Right to Life Committee, (c) Possible US presidential candidate, or (d) don't know. Figure 4 gives the results. (The P-value is 0.2%.)

### Who is H. Ross Perot?

	Wrong answer	Right answer
Male	11	20
Female	21	7

Source: Grinnell student project,  
by Meredith Goulet & Jennifer Wolfson,  
May, 1992.

They asked the question: "How informed do you consider yourself to be about national/international politics, on a scale of 1 to 5?", where 1 was "generally uninformed" and 5 was "very well informed". Figure 5 shows that generally males thought they were more informed (P-value of 4.4%).

**How informed are you?**

	Low	Average	High
Male	4	17	10
Female	11	13	4

Source: Goulet and Wolfson

To get another read on being informed they asked the question, "How often do you read newspapers or news magazines other than the comics or sports sections?". Figure 6 suggests that male students engaged in these activities more frequently (P- value of 0.1%).

**Read newspaper or news magazine?**

	Times per Week	
	1 or fewer	2 or more
Males	5	26
Females	16	12

Source: Goulet and Wolfson

To get an idea about the student's philosophy they asked, "Do you believe that national and international politics ought to be ruled by moral and ethical considerations or by practical power considerations? On a scale of 1 to 5 how would you rate your beliefs?", where 1 was "purely moral/ethical" and 5 was "purely by practical power". The results, given in Figure 7, show the men to be the pragmatists and the women to be the idealists (P-value is 0.01%).

**What philosophy guides politics?**

	Moral <----->Practical				
	2	3	4	5	
Male	7	11	10	3	(31)
Female	19	8	1	0	(28)

Source: Goulet and Wolfson

Then the students asked the question on participation, asking the subject if he/she was eligible to vote, was registered to vote, and had ever voted. Figure 8 summarizes the surprising results: Male students at Grinnell, despite some evidence that they may be better informed about political matters, are less likely to participate in the electoral process than their more idealistic female classmates. (A P-value of 5.5%)



### Participation in the electoral process

	Eligible only	Registered	Has voted	
Male	11	3	14	(28)
Female	3	5	19	(27)

Source: Goulet and Wolfson

This was not an easy project for these students and the survey was not of professional quality, but I still feel they gained much from the assignment. Consider the steps required of the students in solving their problem:

- They had to identify the questions of interest.
- They had to develop operational definitions of variables.
- They had to ask for subject matter expertise, i.e., the political scientist.
- They had to choose a survey design.
- They had to consider and obtain informed consent.
- They had to decide on a proper data analysis.
- They had to present and interpret quantitative information, orally and in a written report.

These are issues not brought out by standard textbook exercises and suggest the kinds of lessons I hope students gain from the project experience. As in anything we do, some gain more than others.

#### Nuts and Bolts.

Here are my suggestions as to what will make for a successful implementation:

- (1) Have the students work in teams. Self-chosen teams of 2 or 3 is my preference. My main rationale for this is that by working together students do better projects and learn more. They experience more of the problem solving process in working together than they would if they came to me every time they got stuck.
- (2) Give importance to the projects. Make it known on day one that projects are an integral part of the course. Give the assignments as early as possible. Use former projects for course examples.
- (3) Structure the assignment. I suggest intermediate due dates for the various stages of the project. For example:
  - a. project description -- brief, conversation starter.
  - b. formal project proposal -- describing details of the population or process under study, variables to be collected, method of producing the data, plans for analysis, expectations about results, etc..
  - c. data and codebook
  - d. rough draft (I make this optional)
  - e. oral report
  - f. written report

This structure helps ensure progress and improves team relations.

(4) Use a scoring rubric to assign points to different components of the project. I actually didn't try this until last year and it is a wonderful tool. Joan Garfield discusses rubrics in the aforementioned paper. For example, on my written reports, I used the categories :

- a. description of the data and variables (10 points)
- b. statistical correctness (20 points)
- c. quality of graphics (15 points)
- d. organization of report (10 points)
- e. overall quality of writing (15 points)

(5) Assign two reports. My students convinced me last fall that we should have two projects instead of one and it worked beautifully. The first project asked them to find a data set with several variables from an almanac, statistical abstract, or some such source, and to describe some interesting relationships. I used the same scoring rubric, but the assignment counted for far less of the final grade than the second project. This provided the students with a lower stress first opportunity to write a statistical report and to learn about my expectations. The point here is to give the students more than once Chance to do something you deem important. Other ways to do this would be to give other kinds of data analysis or data presentation assignments.

## **Conclusion.**

I think projects are a fine addition to a statistics course. I use them at the upper level as well as the introductory level and others have written about them in other settings (some are included in my bibliography). I enjoy reading projects; they contain a lot more variety than an hourly exam, and there are always some pleasant surprises.

## **A Partial Bibliography for Student Projects and Related Issues**

**These papers provide guidelines and motivation for conducting projects in a variety of settings:**

Burrill, G., J.C. Burrill, P. Coffield, G. Davis, J. de Lange, D. Resnick, and M. Siegel (1992), *Data Analysis Across the Curriculum*, Reston, VA: The National Council of Teachers of Mathematics.

Includes a 10 page chapter on student projects and a list of 35 sample projects.

Fillebrown, Sandra (1994), "Using Projects in an Elementary Statistics Course for Non-Science Majors," *Journal of Statistics Education*, v.2, n.2.

The author describes the use of projects in an introductory course for non-science students. She gives practical advice on managing the projects and includes sample student projects.

Halvorsen, Katherine T. & Moore, Thomas L. (1991), "Motivating, Monitoring, and Evaluating Student Projects," *Proceedings of the Section on Statistical Education. of the American Statistical Association – 1991*, pp. 20-25.

Provides very practical guidelines for doing student projects and gives some sample projects as well.

Hunter, W.G. (1977), "Some Ideas About Teaching Design of Experiments with 2<sup>5</sup> Examples of Experiments Conducted by Students," *American Statistician*, v.31, p.12.

This may be the seminal paper on the use of student projects for teaching statistics.

Ledolter, Johannes (1995), "Projects in Introductory Statistics Courses," *The American Statistician*, v.39, pp. 364-367.

Describes using student projects in a large introductory course for business students.

Mackisack, Margaret (1994), "What is the Use of Experiments Conducted by Statistics Students?" *Journal of Statistics Education*, v.2, n.1.

Here the course is one for science and mathematics students. Projects are experiments. The author beautifully examines the educational motivations for assigning projects. Even if you aren't teaching lots of experimental design, this paper is a "must read" if you are considering student projects in your teaching.

Moore, Thomas L. (1996), "Using Student Projects in an Introductory Course for Liberal Arts Students," *Communications in Statistics*, v.25, pp.2647-2661.

These "notes on projects" above are essentially a draft of this paper.

Roberts, Harry V. (1992), "Student-Conducted Projects in Introductory Statistics Courses," in *Statistics for the Twenty-First Century*, MAA Notes, Number 26, p. 109, Gordon, F. and Gordon, S., editors.

To the query, "How can experienced teachers get started on projects?" Roberts answers, "Just jump in and do them, and learn as you go." Fortunately this article gives you some solid advice backed by Roberts's many years teaching introductory statistics, primarily to MBA students. Roberts figures course grades solely on student projects.

Sevin, Anne (1995), "Some Tips for Helping Students in Introductory Statistics Classes Carry Out Successful Data Analysis Projects," *Proceedings of the Section on Statistical Education. of the American Statistical Association – 1995*, pp.159-164.

Sevin gives excellent advice for managing student projects, including the use of interim reports, frequent feedback, and peer review. She also gives samples of actual assignments.

Short, Thomas H. and Joseph G. Pigeon (1998), "Protocols and Pilot Studies: Taking Data Collection Projects Seriously," *Journal of Statistics Education*, v.6, n.1.

Short and Pigeon describe ways to engage students in the often-overlooked planning stages of a study. They describe both short assignments and student projects more in the mold of other works in this bibliography.

Zahn, Douglas A. (1992), "Student Projects in a Large Lecture Introductory Business Course." *American Statistical Association, Proceedings of the Section on Statistical Education. of the American Statistical Association – 1992*, pp.147-154.

If you write Zahn he will send you (for copying costs) an annotated bibliography of resources and his extensive instructions, including a 33 page set of instructions for term projects. Dept. of Statistics, Florida State University, Tallahassee, FL 32306-3033.

**These papers discuss important teaching issues that relate to using projects:**

Gal, Iddo & Ginsburg, Lynda (1993), "The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework," *Journal of Statistics Education*, v.1, n.1.

Garfield, Joan (1993), "Teaching Statistics Using Small-Group Cooperative Learning," *Journal of Statistics Education*, v.1, n.1.

Garfield, Joan (1994), "Beyond Testing and Grading: Using Assessment to Improve Student Learning," *Journal of Statistics Education*, v.2, n.1.

## 8.6 Evaluation Instruments

### Probability and Statistics Pre-course Survey

**Purpose** The purpose of this survey is to indicate what you already know and think about probability and statistics.

**Take your time** The questions require you to read and think carefully about various situations. If you are unsure of what you are being asked to do, please raise your hand for assistance.

#### Part I

On the first few pages are a series of statements concerning beliefs or attitudes about probability, statistics and mathematics. Following each statement is an "agreement" scale which ranges from 1 to 5, as shown below.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree, nor disagree	Agree	Strongly Agree

If you strongly agree with a particular statement, circle the number 5 on the scale. If you strongly disagree with the statement, circling the number 1.

---

1. I often use statistical information in forming my opinions or making decisions.

1    2    3    4    5

2. To be an intelligent consumer, it is necessary to know something about statistics.

1    2    3    4    5

3. Because it is easy to lie with statistics, I don't trust them at all.

1    2    3    4    5

4. Understanding probability and statistics is becoming increasingly important in our society, and may become as essential as being able to add and subtract.

1    2    3    4    5

5. Given the chance, I would like to learn more about probability and statistics.

1    2    3    4    5

6. You must be good at mathematics to understand basic statistical concepts.

1    2    3    4    5

7. When buying a new car, asking a few friends about problems they have had with their cars is preferable to consulting an owner satisfaction survey in a consumer magazine.

1      2      3      4      5

8. Statements about probability (such as what the odds are of winning a lottery) seem very clear to me.

1      2      3      4      5

9. I can understand almost all of the statistical terms that I encounter in newspapers or on television.

1      2      3      4      5

10. I could easily explain how an opinion poll works.

1      2      3      4      5

1. A small object was weighed on the same scale separately by nine students in a science class. The weights (in grams) recorded by each student are shown below.

6.2    6.0    6.0    15.3    6.1    6.3    6.2    6.15    6.2

The students want to determine as accurately as they can the actual weight of this object. Of the following methods, which would you recommend they use?

- a. Use the most common number, which is 6.2.
- b. Use the 6.15 since it is the most accurate weighing.
- c. Add up the 9 numbers and divide by 9.
- d. Throw out the 15.3, add up the other 8 numbers and divide by 8.

2. A marketing research company was asked to determine how much money teenagers (ages 13 - 19) spend on recorded music (cassette tapes, CDs and records). The company randomly selected 80 malls located around the country. A field researcher stood in a central location in the mall and asked passers-by who appeared to be the appropriate age to fill out a questionnaire. A total of 2,050 questionnaires were completed by teenagers. On the basis of this survey, the research company reported that the average teenager in this country spends \$155 each year on recorded music.

Listed below are several statements concerning this survey. Place a check by every statement that you agree with.

- a. The average is based on teenagers' estimates of what they spend and therefore could be quite different from what teenagers actually spend.
- b. They should have done the survey at more than 80 malls if they wanted an average based on teenagers throughout the country.
- c. The sample of 2,050 teenagers is too small to permit drawing conclusions about the entire country.
- d. They should have asked teenagers coming out of music stores.
- e. The average could be a poor estimate of the spending of all teenagers given that teenagers were not randomly chosen to fill out the questionnaire.
- f. The average could be a poor estimate of the spending of all teenagers given that only teenagers in malls were sampled.
- g. Calculating an average in this case is inappropriate since there is a lot of variation in how much teenagers spend.
- h. I don't agree with any of these statements.

3. Which of the following sequences is most likely to result from flipping a fair coin 5 times?

- a. H H H T T
- b. T H H T H
- c. T H T T T
- d. H T H T H
- e. All four sequences are equally likely

4. Select the alternative below that is the best explanation for the answer you gave for the item above.

- a. Since the coin is fair, you ought to get roughly equal numbers of heads and tails.
- b. Since coin flipping is random, the coin ought to alternate frequently between landing heads and tails.
- c. Any of the sequences could occur.
- d. If you repeatedly flipped a coin five times, each of these sequences would occur about as often as any other sequence.
- e. If you get a couple of heads in a row, the probability of a tails on the next flip increases.
- f. Every sequence of five flips has exactly the same probability of occurring.

5. Listed below are the same sequences of Hs and Ts that were listed in Item 3. Which of the sequences is least likely to result from flipping a fair coin 5 times?

- a. H H H T T
- b. T H H T H
- c. T H T T T
- d. H T H T H
- e. All four sequences are equally unlikely

6. The Caldwells want to buy a new car, and they have narrowed their choices to a Buick or a Oldsmobile. They first consulted an issue of Consumer Reports, which compared rates of repairs for various cars. Records of repairs done on 400 cars of each type showed somewhat fewer mechanical problems with the Buick than with the Oldsmobile.

The Caldwells then talked to three friends, two Oldsmobile owners, and one former Buick owner. Both Oldsmobile owners reported having a few mechanical problems, but nothing major. The Buick owner, however, exploded when asked how he like his car: First, the fuel injection went out - \$250 bucks. Next, I started having trouble with the rear end and had to replace it. I finally decided to sell it after the transmission went. I'd never buy another Buick.

The Caldwells want to buy the car that is less likely to require major repair work. Given what they currently know, which car would you recommend that they buy?



- a. I would recommend that they buy the Oldsmobile, primarily because of all the trouble their friend had with his Buick. Since they haven't heard similar horror stories about the Oldsmobile, they should go with it.
- b. I would recommend that they buy the Buick in spite of their friend's bad experience. That is just one case, while the information reported in Consumer Reports is based on many cases. And according to that data, the Buick is somewhat less likely to require repairs.
- c. I would tell them that it didn't matter which car they bought. Even though one of the models might be more likely than the other to require repairs, they could still, just by chance, get stuck with a particular car that would need a lot of repairs. They may as well toss a coin to decide.

7. Half of all newborns are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. On a particular day, which hospital is more likely to record 80% or more female births?

- a. Hospital A (with 50 births a day)
- b. Hospital B (with 10 births a day)
- c. The two hospitals are equally likely to record such an event.

8. "Megabucks" is a weekly lottery played in many states. The numbers 1 through 36 are placed into a container. Six numbers are randomly drawn out, without replacement. In order to win, a player must correctly predict all 6 numbers. The drawing is conducted once a week, each time beginning with the numbers 1 through 36.

The following question about the lottery appeared in The New York Times (May 22, 1990).

Are your odds of winning the lottery better if you play the same numbers week after week or if you change the numbers every week?

What do you think?

- a. I think the odds are better if you play the same numbers week after week.
- b. I think the odds are better if you change the numbers every week.
- c. I think the odds are the same for each strategy.

9. For one month, 500 elementary students kept a daily record of the hours they spent watching television. The average number of hours per week spent watching television was 28. The researchers conducting the study also obtained report cards for each of the students. They found that the students who did well in school spent less time watching television than those students who did poorly. Listed below are several possible statements concerning the results of this research. Place a check by every statement that you agree with.

- \_\_\_\_\_ a. The sample of 500 is too small to permit drawing conclusions.
- \_\_\_\_\_ b. If a student decreased the amount of time spent watching television, his or her performance in school would improve.
- \_\_\_\_\_ c. Even though students who did well watched less television, this doesn't necessarily mean that watching television hurts school performance.
- \_\_\_\_\_ d. One month is not a long enough period of time to estimate how many hours the students really spend watching television.
- \_\_\_\_\_ e. The research demonstrates that watching television causes poorer performance in school.
- \_\_\_\_\_ f. I don't agree with any of these statements.

10. An experiment is conducted to test the efficacy of a new drug on curing a disease. The experiment is designed so that the number of patients who are cured using the new drug is compared to the number of patients who are cured using the current treatment. The percentage of patients who are cured using the current treatment is 50% and 65% are cured who have used the new drug. A P-value of 5% (.05) is given as an indication of the statistical significance of these results. The P-value tells you:

- \_\_\_\_\_ a. There is a 5% chance that the new drug is more effective than the current treatment.
- \_\_\_\_\_ b. If the current treatment and the new drug were equally effective, then 5% of the times we conducted the experiment we would observe a difference as big or bigger than the 15% we observed here.
- \_\_\_\_\_ c. There is a 5% chance that the new drug is at least better than the current treatment by at least 15%.

11. Gallup reports the results of a poll that shows that 58% of a random sample of adult Americans approve of President Clinton's performance as president. The report says that the margin of error is 3%. What does this margin of error mean?

- \_\_\_\_\_ a. One can be 95% "confident" that between 55% and 61% of all adult Americans approve of the President's performance.
- \_\_\_\_\_ b. One can be sure that between 55% and 61% of all adult Americans approve of the President's performance.
- \_\_\_\_\_ c. The sample percentage of 58% could be off by 3% in either direction due to inaccuracies in the survey process.
- \_\_\_\_\_ d. There is a 3% chance that the percentage of 58 is an inaccurate estimate of the population of all Americans who approve of President Clinton's performance as president.

## Evaluation of Newspaper Critiques: Scoring Rubric

Student paper: \_\_\_\_\_

Assign 0-3 points for each of the following categories:

\_\_\_1. Clearly states the purpose of the research study and the basic research question.

- 0 pt: doesn't attempt to state purpose of question
- 1 pt: purpose or question stated are incorrect
- 2 pt: purpose or question are moderately well stated
- 3 pt: purpose and question are appropriately and correctly stated

\_\_\_2. Clearly defines methods used to answer the research question.

- 0 pt: doesn't define methods
- 1 pt: incorrectly describes methods
- 2 pt: does a fair job of describing methods
- 3 pt: accurately described research methods

\_\_\_3. Poses reasonable questions to ask the investigators to better understand the study.

- 0 pt: doesn't post any questions to ask investigators
- 1 pt: asks irrelevant or inappropriate questions
- 2 pt: asks questions moderately clear and appropriate
- 3 pt: asks perceptive and appropriate questions about the study

\_\_\_4. Raises appropriate questions about the validity of the conclusions.

- 0 pt: doesn't address this question
- 1 pt: accepts conclusions without any reservation
- 2 pt: raises issues (somewhat relevant or not clearly defined) that question validity of conclusions
- 3 pt. raises relevant issues that question the validity of conclusions

Overall score: \_\_\_\_\_

Comments:

## Assessment of Student Projects: Scoring Rubric

Student's name \_\_\_\_\_

Title of Paper \_\_\_\_\_

Assign 0-3 points for each of the following categories:

- \_\_\_ Understands and clearly states the problem
- \_\_\_ Describes the process of investigating the problem
- \_\_\_ Discussions strengths and weaknesses of the literature
- \_\_\_ Draws reasonable conclusions based on supporting evidence
- \_\_\_ Communicates effectively

Comments:

## Student Evaluation of Chance Course

Note to instructors: please modify this survey as needed so that it reflects the components of your version of the CHANCE course.

Please use the following scale to rate each component of the course, and add any comments you might have that will help us understand the reason for your rating.

U = unsatisfactory  
M = marginal  
FG = fairly good  
VG = very good  
E = excellent

1. The selection of topics presented in the course (e.g. list some here....)

U            M            FG            VG            E

Topic you liked best: \_\_\_\_\_ Why?

Topic you liked least: \_\_\_\_\_ Why?

2. The course format: (problems, discussions, small groups, projects)

U            M            FG            VG            E

3. The use of the computer

U            M            FG            VG            E

4. The use of journals.

U            M            FG            VG            E

5. The textbook

U            M            FG            VG            E

6. The newspaper and journal articles used in the course

U            M            FG            VG            E

7. The student projects

U            M            FG            VG            E

8. The class discussions

U            M            FG            VG            E

9. The small-group discussions

U            M            FG            VG            E

10. How much have you learned in this course?

An exceptional amount    Very Much    Much    Some    Little

11. All things considered, how would you rate this course?

Exceptionally Good    Excellent    Very Good    Good    Fair    Poor    very poor

12. What did you like best about this class?

13. What did you like least about this class?

14. Do you plan to enroll in another statistics course in the future?

15. How confident do you feel about assessing statistical information reported in a scientific study?

Very confident    Somewhat Confident    Unsure    Not Confident at all

16. How has this course affected your view of how research is conducted?

17. How has this course affected your opinion of how research is presented in the media?

18. Any final suggestions for improving the course?

## 8.7 Book Reviews

### **Tainted Truth" by Cynthia Crossen**

Tainted Truth: the Manipulation of Fact in America  
Cynthia Crossen.  
Simon&Schuster, N.Y., 1994

From Chance News

This book might well be required reading for students in a CHANCE course.

Cynthia Crossen argues that a lot can go wrong with statistics that cannot be blamed on the whims of chance. Her many insightful observations include:

For very good reasons having little to do with statistics, Coca Cola taste studies show that people prefer Coke and Pepsi Cola studies show that they prefer Pepsi.

Polls are a politician's weapon and are frequently designed and interpreted accordingly.

A study paid for by the tobacco company is not likely to conclude that second hand smoke is dangerous and medical researchers funded by drug companies may not be completely immune to bias.

Parameters in risk models can and often are chosen judiciously to make the outcome agreeable to the developer of the model.

Expert witnesses in the courts are paid a lot of money to try to reach different conclusions from the same statistical data.

Each chapter has numerous examples. Here are chapter titles and examples from each chapter that I particular enjoyed.

#### Chapter 1: The Study Game

Example: The author asked Gallup five questions about credibility and information. Gallup responded by making the questions into a survey which they carried out for a modest fee (\$4,500).

#### Chapter 2: The Truth about Food.

Example: the oat bran mania of the 80's and the stampede against Alar started by the 60 Minutes report.

#### Chapter 3: Numerical Lies of Advertising:

Example: An account of the taste tests carried out by Coca Cola and Pepsi companies.

#### Chapter 4: False Barometers of Opinion:

Example: Polls taken at the time of the Clarence Thomas hearings indicated that people did not believe Anita Hill. It is likely that these polls influenced the outcome of the Thomas appointment.

However, similar polls a year later indicated that Anita Hill was at least as believable, if not more so, than Clarence Thomas.

#### Chapter 5: False Truth and the Future of the World

Example: The environmental battle between disposable and cloth diaper industries. A 1988 study by the cloth diaper industry served as ammunition for opponents of disposable until this study was neutralized by a 1990 study produced by Proctor & Gamble. This in turn was followed in 1991 by another study sponsored by the cloth diaper industry showing that cloth diapers were environmentally superior.

#### Chapter 6: Drugs and Money

Example: Two conflicting studies on the effectiveness of a drug based on the same data were submitted simultaneously to the New England Journal of Medicine. The author who found the drug effective had large grants from the pharmaceutical companies and had his paper accepted. The other with no grants had his paper rejected by the NEJM but then accepted by the Journal of the American Medical Association but this did not lead to a happy ending.

#### Chapter 7: Research in the courtroom

Example: The account of the Dalkon Shield case is interesting but I would have preferred DNA finger-printing.

#### Chapter 8: Solutions

Example: I liked the recommendation that high schools and colleges teach critical assessment of such news. In other words, Take a chance on CHANCE.

### **A Mathematician Reads the Newspaper" by John Paulos**

John Paulos' books, *Innumercy* and *Beyond Innumercy* are useful but most useful is his recent book *A Mathematician Reads the Newspaper*. Here is a review of this book.

*Philadelphia Inquirer Book Review*, 30 April, 1995

*A Mathematician Reads the Newspaper*

John Allen Paulos

Basic Books. 212 pp. \$18

Reviewed by Charles Seife

"We are all going to starve," cries one newspaper. "The Earth's population, laid end to end, would stretch to the moon and back eight times. There are simply too many people on the planet for us to survive for long!"

"Don't panic," exclaims another. "If every man, woman, and child on Earth were given a cubical room measuring 20 feet per side, all of the apartments would fit inside the Grand Canyon. There's plenty of space for everyone!"

Which one is correct? Both are.

In a similar vein, imagine that someone accidentally dumps a radioactive chemical (something like tritium hydroxide) into the Atlantic. Over time, the chemical spreads evenly in the world's oceans, and several years later, every pint of water in every ocean in the world contains about 6,000 molecules of that chemical. Newspapers would carry furious headlines like "Tremendous Spill Causes Cataclysmic Contamination." How big was the spill? One pint.

Americans are scared and awed by mathematics, so most people don't have an intuitive grasp about what numbers represent. This allows spin doctors to manipulate stories to suit their interests. John Allen Paulos, author of the best-selling *Innumeracy*, exposes their tricks in *A Mathematician Reads the Newspaper*.

But Paulos, a professor of mathematics at Temple University, does not stop there. The book is a mathematician's perspective on the news, and Paulos writes about DNA testing, NAFTA, affirmative action, the SATs, tennis, the gulf war, AIDS, and even aliens from outer space.

Why would anyone care about Paulos' opinions on subjects that are out of a mathematician's realm of experience? After all, scientists, economists, politicians, educators, sportscasters, historians, doctors, and Shirley MacLaine are the authorities on those news stories. Paulos demonstrates that mathematical knowledge is crucial to understanding newspaper articles.

Paulos shows the "experts" little mercy. For instance, though economists theorize about when distant recessions will begin and end, their predictions are seldom correct. With cutting logic, Paulos uses chaos theory to explain their inevitable inaccuracy.

Complex systems such as the weather and the economy share an extreme long-term sensitivity to slight changes in conditions. The oft-repeated example that a butterfly flapping its wings in China on Monday can cause a hurricane in Philadelphia on Friday is more than mere hyperbole. It is the reason that meteorologists rarely hazard a prediction beyond the five-day forecast. Just as the kingdom was lost for want of a nail, small events have big consequences as their effects propagate through time.

The same problem haunts economists and market analysts. Imagine that a banker's indigestion causes him to refuse a loan. This causes a company to have its credit rating downgraded, shaking faith in a sector of an industry, sending ripples through the economy. An emperor cannot account for every nail in the kingdom; a market analyst cannot predict the actions of every banker and trader. As a result, we must take long-range economic or political predictions with a very large grain of salt.

*A Mathematician Reads the Newspaper* is an unusual book, and doing justice to it in a short review is impossible. Its format mimics that of a newspaper. Organized into roughly 50 short articles, it deals with topics in politics, business, social issues, soft news, science, the environment, - even the obituaries. Each section provides new insight and new methods for interpreting a news story. Trade negotiations are linked to game theory. Conditional probability sheds light upon health scares. Paulos tilts at subjects from the "appropriately named" Laffer Curve of Reaganomics to the overly precise calorie counts in food sections. He lays bare a number of tricks of the spin doctors and offers some way to see through the fog of advertisements. Even better, Paulos' wit and humor - admirably displayed in *Innumeracy* - are in top form. His irreverent and pointed comments entertain as well as educate. Though Paulos writes about a bewildering number of topics, he has something fresh and interesting to say about each.

The book provides a toolbox full of mathematical ideas perfectly suited for extracting the truth from a newspaper article. Paulos forces the reader to rethink positions on political and social issues without paying much attention to the issues themselves. Instead, he attacks the logical



framework on which these positions rest - with amazing effect. He draws the reader into a mathematician's world; in the process, he introduces abstract concepts from information theory to conditional probability. His adept explanations are simple and accessible.

Like *Innumeracy*, *A Mathematician Reads the Newspaper* is an excellent place to pick up some math - math that is interesting and vital for understanding the news. After reading *A Mathematician Reads the Newspaper*, it will be impossible to look at a newspaper in the same way.

Charles Seife is a graduate student in mathematics at Yale.

### **"The Power of Logical Thinking" by Marilyn vos Savant**

The Power of Logical Thinking  
Marilyn vos Savant  
Saint Martin's Press, 1996, New York

From Chance News 5.05

Marilyn vos Savant's latest book organizes material from her recent columns according to three themes: Part 1, how our own minds can work against us; Part 2, how numbers and statistics can mislead; and Part 3, how politicians exploit our innocence.

Part 1 features a discussion of the now famous Monte Hall problem and the other related paradoxes that came her way inspired by the interest in the Monte Hall problem.

Part 2 is Marilyn's version of "How to lie with statistics".

Part 3 is the most original. Here she uses the 1992 election to show how politicians use numbers to their own advantage. Clips are chosen from the major news sources and analyzed to show how they do this. Two examples are the statements from the Clinton campaign: "most people are working harder for less money than they were earning ten years ago" and "there is something wrong with our tax code, if your income went up 65 percent in the 1980's and your taxes went down 15 percent".

One of the most interesting contributions in the book is an appendix which gives a popular account written by Donald Granberg of an article he wrote with T. Brown, "The Monty Hall Dilemma," *Personality and Social Psychology Bulletin*, 1995, vol. 31, pp. 711-723. Ms. vos Savant gave Professor Granberg the mountains of correspondence she received after her analysis of the Monty Hall problem. Granberg and Brown analyze this correspondence to assess the concerns of the writers. They also report on experiments designed to estimate the proportion of people who start with a misconception of this problem and how hard it is to get these people to change their minds in repeated experiments. They do similar experiments with a different but equivalent version of the problem they call Russian Roulette. In this version there are two cars and one goat and Monte opens a door with a car, making this car no longer available to the contestant. Now the contestant should not switch and, unlike the Monte Hall version, most people get this right.

## **An Electronic Companion to Statistics by George Cobb**

From Chance News 6.08

An Electronic Companion to Statistics.  
by George Cobb and Jeffrey A Witner, Jonathan D. Cruyer  
with the assistance of Peter L. Renz and Kristopher Jennings  
Cogito Learning Media, New York, 1997  
1-800-WE-THINK  
think@cogitomedia.com  
Suggested retail price: \$29.95

The last two years have seen the development of some wonderful resources to supplement standard texts for teaching a basic statistics course including: activities from Richard Shaeffer and Allan Rossman, case studies from Samprit Chatterjee, a multi-media CD-ROM from Paul Velleman and now an electronic companion from George Cobb et al. These are all unique contributions---very different from each other---and produced by statisticians who have thought long and hard about what statistics is all about and how it should be taught.

Cobb's book and CD-ROM is meant to accompany any of the standard statistics texts to give students a chance to check out how they are doing and whether they really understand what they have learned. In order of topics, they follow books such as those of David Moore and others who "follow the modern distinction between exploration-and-description and inference". The 13 units, reviewing standard topics, can be easily changed to fit other models.

On the CD-ROM each unit is introduced with a short video from "Against All Odds" providing a real world example of the topic. Then there is a brief discussion of each topic and an opportunity for students to self-test their knowledge of this topic. The self- testing is livened up by "drag and drop" answers to fill-in-the- blanks and true-false questions. Students who are stuck can click on an important term to be reminded what it means or click on a "hint" button to get a suggestion how to start to think about the question. The exercises ask questions about actual studies and reports of studies in the media. They are beautifully designed to be sure the student really understands the topic. George and his colleagues have made sure that this project passes the "Cobb test" (see our quote for this issue).

One might think that the accompanying workbook, which treats the same topics using only the written word, would pale by comparison with the CD-ROM. George is not called the "intellectual" of the statistics reform group for nothing. By the use of words alone George reminds the students what the topics mean, how they relate, and gives students new ways to think about difficult topics. For example when reviewing the concept of probability distribution, he tells the student: It helps to try to become comfortable with four different variations on the one basic idea:

numerically, as a table listing all possible outcomes, together with the chance for each outcome;

visually, as a probability histogram, with one bar for each outcome and the area of the bar proportional to the chance of the outcomes;

physically, as a "box model" with numbered tickets as the outcome and the chance of each outcome given by the fraction of tickets in the box having that number on them, and;

predictively, as an answer to the question: What will happen if I repeat a particular chance process a very large number of times?

Here is another succinct comment that will be more than a review to most students:

Correlations summarize balloons. If your plot isn't balloon shaped, don't use a correlation.

The relations between the different statistical concepts are illustrated both in the workbook and on the CD-ROM by diagrams called concept maps. Thinking in terms of concept maps led George to depart slightly from tradition and put the topic of time series between describing distributions and describing relations. This is natural since time series are just a relation between two variables with one variable time.

Students who work their way through this review have learned that, by putting it all together, they have mastered a powerful tool for understanding important real life problems.

Well, once again we have to make a disclaimer. Like Paul Velleman, George Cobb was a Dartmouth undergraduate and a student in our probability course some years ago. We give him an A+ on his latest project.

### **"Workshop Statistics" by Allan Rossman**

From Chance News 5.10

Workshop Statistics : Discovery With Data and Minitab  
(Textbooks in Mathematical Sciences) Allan J. Rossman, Beth L. Chance / Paperback /  
Published 1998  
Amazon: \$39.95

Workshop Statistics : Discovery With Data and the Graphing Calculator (Textbooks in  
Mathematical Sciences) ~ Ships in 2-3 days  
Allan J. Rossman, et al / Paperback / Published 1997  
Amazon: \$29.95

Workshop Statistics : Discovery With Data  
Allan J. Rossman / Paperback / Published 1996 (Publisher Out Of Stock)

Note: We reviewed the 1996 version. .

Rossman states that "Statistics is the science of reasoning from data." He has based his book on this belief and his philosophy that students learn statistics by doing it.

Rossman envisions the classroom as a laboratory where the instructor gives occasional explanations of basic ideas but mostly helps the students in a co-operative learning experience. The basic ideas of statistics are learned in terms of exploring data sets either provided by the text or generated by the students. Activities guide them in their explorations. Emphasis is placed on students learning to effectively communicate their findings. The data sets provided are from a variety of fields of study and many represent issues of current interest to students such as: student's political views, hazardness of sports, and campus alcohol habits.

The book is organized by subject into six units each of which has several subunits. Each subunit has the following items

- \* Overview: a brief introduction to the topic, emphasizing its connection to earlier topics.
- \* Objectives: a listing of specific goals for students to achieve in the topic.
- \* Preliminaries: a series of questions to get the students thinking about the issues and applications to be studied and sometimes to collect relevant data.
- \* In-class activities: the activities that guide students to learn the material of the topic.
- \* Homework activities.

The first two units introduce descriptive statistics and standard graphical displays of data and exploratory data analysis. The third unit deals with randomness and the last three with statistical inference. The students are assumed to have a computer or calculator available for their explorations.

The exploratory statistics units give the student a wonderful opportunity to try lots of different and interesting ways to look at the data, starting with a few basic graphics techniques such as stem-leaf plots, box-plots, histograms and scatter diagrams.

The tools for inference and test of hypothesis provided are limited to the standard normal, t-test and binomial tests. It might be better, having described statistics such as the sample mean, sample standard deviation or the chi-squared statistic, to have the students obtain, by simulation, the approximate confidence intervals, or P-values. Then, for example, if they were given the activity to design an experiment to determine if a fellow student can tell the difference between Pepsi and Coke they could explore different designs instead of being limited to the independent trials model suggested for Fisher's famous "tea testing experiment." They might even choose the design that Fisher used.

The use of the "bootstrap method" would also allow more adventuresome explorations. The instructor might have to help with some of the simulations but, after all, that's co-operative learning.

On the other hand, what makes this such a great book is that the author has limited himself to make it possible to get over the basic concepts of statistics at a reasonable level and to make a course based on the book very teachable. The book can also be used as supplementary material to liven up a more traditional course. In either case, Rossman's book shows that student's first introduction to statistics can be made the exciting experience it should be.

Editors (Snell's) comment: Here are some comments on the Rossman book from the biased point of view of a probabilist. As Rossman dramatically demonstrates, it is a lot more interesting and instructive to introduce statistical concepts in terms of real data related to serious issues. This presents a challenge to we who write probability books to discuss basic concepts of probability in terms of experiments and data corresponding to significant problems rather than the traditional experiments of tossing coins, rolling dice and drawing balls out of urns.

Of course, as a probabilist we were disappointed that there is essentially no discussion of basic probability concepts such as conditional probability and expected value. We think that it is a mistake to separate statistical reasoning and probabilistic reasoning so completely. After all, the greats such as Laplace, Fisher and Galton didn't; so why should we?

## **"Activity-Based Statistics" by Richard Scheaffer et. al.**

Activity-based statistics by Scheaffer, Gnadiseken, Watkins, Witmer  
"Instructor Resources" available from Springer-Verlag  
"Student Guide" available from Jones and Bartlett

From Chance News 5.10.

Under an NSF grant, the authors developed and tried out a large number of activities suitable for an introductory statistics course. In the "Student Guide", the authors give almost an encyclopedia of the activities they developed and tested. In the "Instructor's Resources", they provide these activities and discuss the art of using them in a statistics course.

In the Student Guide, each activity starts with a "scenario" which, in most cases, tells the student a real-world situation relevant to the activity. Then the objectives of the activity and a question that it will answer are given. Next, detailed instructions are given for carrying out the activity. Then students are given some "wrap up" questions and possible extensions of the activity.

The "Instructor Resources" provides the pages from the "Student Version" relating to each activity. It starts each discussion of an activity with general remarks about where and how it might be used in a statistics course. It then specifies what the students need to know and the materials needed to carry out the activity. Sample results from previous experiences with the activity are provided. Finally, you will find sample assessment questions to see what the students have learned from doing the activity.

There are many more activities than one would use in a single course, allowing instructors to pick those most suited to their own course. It is also possible to construct an interesting statistics course using primarily some of these activities.

The activities include a few old favorites, such as the German tank problem and standing coins on end, but, at least for us, most of them were new. We were pleased to see that a number of activities aimed at learning probability concepts are provided. Strangely, the important concept of conditional probability is again missing. It must have taken real will-power to omit the infamous Monty Hall problem, and we can appreciate not including this, but, surely, the activity of having students discover the probability of AIDS given a positive test would have been appropriate.

The activities range from very simple to somewhat complex. For example, to illustrate the idea of bias we find the very simple activity of asking the students to estimate the length of a piece of string 45 inches long. The distribution of the student's estimates will be centered near the more natural length of 36 inches.

To introduce the idea of trying to tell the effect on the outcome of an experiment of one factor when two factors affect the outcome, the authors describe a delightful but more complicated "frog activity". Students begin by constructing a frog from a square piece of paper. The students are randomized according to the four different possibilities of size and weight of paper and then they experiment to see how far their frogs will jump. We confess we did not believe jumping frogs could be constructed from a square piece of papers and so, with some difficulty and some help from my son, We followed the instructions and were delighted to find that our frog did indeed jump quite well. (We will try to put a video of our frog jumping on the web version of this chance news.

If you used some of these activities and the students enjoy them as much as we did the "frog activity" your course cannot fail!

### **ActivStats**

CD-ROM, Addison-Wesley, July 1997.

Paul Velleman

ISBN 0-201-31071-6, \$42.50

Concern is often expressed that most students do not get their first introduction to statistics from a statistician. This CD-ROM will go far toward improving this situation. Students who have ActivStats will learn what statistician Paul Velleman thinks statistics is all about.

They will not hear him an hour at a time but more like three or four minutes at a time. The segments are short because they are limited to a single concept. The goal is to motivate, explain, visualize and reinforce each individual concept before going on to the next -- something we can't afford to do in a lecture.

In a typical segment, Paul tells the students the importance of understanding the context of the data. He gives examples from advertisements and asks the students to answer the questions who?what? and why? relative to the collection of the data. Key points are presented on a virtual blackboard, and the students are encouraged to take notes.

In between Paul's segments, students are asked to carry out a variety of tasks related to the topic Paul has introduced. For example, students may be asked to analyze a data set using the built in statistical package Data Desk, read a paragraph from a standard text such as David Moore's "Basic Practice of Statistics," or test their understanding of the concept by trying to make a correct transformation between a set of words and a paragraph with missing words. Wrong words tumble back to the word source and final victory is rewarded by a pleasant cheer.

Students may also be invited to see a segment from the popular "Against all Odds" and "Decisions Through Data" video series. For example, in the unit on "understanding relationships", students view the segment relating to the Boston Beanstalk Tall Club. If they have a web browser on their computer and click on the WEB button, they will be taken to the Boston Tall Club's homepage.

When studying correlation, the WEB button takes them to articles in Chance News that use correlation. When studying confidence intervals, clicking on the TOOL button gives the student a chance to draw samples from a population and visualize the Central Limit Theorem.

A student making the mistake of clicking on the WORK button will be given a set of homework problems relating to the unit being studied. Clicking on the PROJ button provide them with a mini- project to carry out that relates to the topic being studied.

Authoring tools will allow instructors to add their own lessons to the end of a page of the lesson book. Those lessons could be as simple as some text to read, could be datasets that launch a statistics package (including one other than Data Desk, if they prefer), or a page with a URL – for example to administer a quiz or provide more information from a course home site.

It is hard to think of a subject which lends itself better to multi-media presentation than statistics. Having Paul Velleman, DataDesk, "Against all Odds", numerous activities, and the Web at your fingertips will present real competition to the conventional way of learning statistics.

It will be interesting to see how this CD-ROM will be used. For a Chance course, it would permit the instructor to spend most of class time discussing news articles and carrying out activities, letting Paul teach the basic concepts. One can even imagine that some students at Dartmouth, who pay about \$3000 to take an introductory statistics course, might choose to save \$2958 by learning their statistics from Paul!

ActivStats was available and tested this Spring with a Mac version. A fully cross-platform Mac/Win95/WinNT release will be available in July. Teachers can obtain now a preliminary version of the July release by contacting Bill Danon at [awi-info@aw.com](mailto:awi-info@aw.com) or 617-944-3700x2563 (e-mail preferred).

Disclosure: Paul was one of our favorite Dartmouth undergraduates who acted as John Kemeny's statistical advisor in introducing co- education at Dartmouth.

### **List of Undergraduate Textbooks**

<http://wwwcsc.cornell-iowa.edu/~acannon/stated/booklist.html>

This page, constructed by Ann Cannon, contains a partial list of textbooks available for use in two typical undergraduate courses: Introductory Statistics and Mathematical Statistics. The books listed in the supplements section might also be used as a primary textbook depending on the structure of the course. Books for which we have found reviews, have links to the bibliographic information for those reviews.