

**CHANCE News 13.03**  
**March 10, 2004 to April 17, 2004**

Prepared by J. Laurie Snell, Bill Peterson, Jeanne Albert, Charles Grinstead, and Myles McLeod with help from Fuxing Hou and Joan Snell. We are now using a listserv to send out notices that a new Chance News has been posted on the Chance Website. You can sign on or off or change your address at [here](#). This listserv is used only for this posting and not for comments on Chance News. We do appreciate comments and suggestions for new articles. Please send these to [jlsnell@dartmouth.edu](mailto:jlsnell@dartmouth.edu). Chance News is based on current news articles described briefly in [Chance News Lite](#).

The current and previous issues of Chance News and other materials for teaching a Chance course are available from the [Chance web site](#).

Chance News is distributed under the GNU General Public License (so-called 'copyleft'). See the end of the newsletter for details.

---

You know how dumb the average guy is? Well, by definition, half of them are even dumber than that.

J.R. "Bob" Dobbs

---

### Contents of Chance News 13.03

- [\(1\) Forsooth.](#)
- [\(2\) Dogs do resemble their owners, finds study.](#)
- [\(3\) The case of the Cherry Hill cluster.](#)
- [\(4\) Google numbers.](#)
- [\(5\) Chew On.](#)
- [\(6\) The margin of error is the easy part.](#)
- [\(7\) Tuning in on a telephone poll.](#)
- [\(8\) Betting against the pros.](#)
- [\(9\) When will a NHL team be mathematically eliminated from the playoffs?](#)
- [\(10\) The communication of risk.](#)
- [\(11\) Making de Mere turn over in his grave.](#)

---

Here is a Forsooth item from the April 2004 issue of *RSS News*:

In 1961, the average number of children born to an Australian woman was at its highest level since reliable records began in the 1920s, at 3.55. It was also the only occasion documented this century in which the 3.5 child average was exceeded - meaning women were statistically more likely to have four children than any other number.

Sydney Morning Herald  
8 February 2004

---

**Dogs do resemble their owners, finds study.**  
**NewScientist.com, March 31, 2004**  
**Shaoni Bhattacharya**

**Research report: Do dogs resemble their owners?**  
**Michael Roy and Nicholas Christenfeld**

Two researchers from the Psychology Department at the University of California, San Diego, conducted a study in which 28 student judges were asked to match photos of dogs with their owners. Each student was presented with a photo of a dog owner and photos of two dogs; one dog was the actual pet, the other an impostor. If more than half of the judges correctly paired a given dog with his or her owner, this was considered a "match."

Forty-five dogs "took part" in the study--25 pure-breds and 20 mutts. They were found with their owners at several local parks in San Diego and photographed against a variety of backgrounds. Overall, there were just 23 matches (as defined above.) However, the judges had an easier time with the pure-bred dogs: 16 matches, versus 7 for the mixed breeds.

The findings will be published in the May issue of *Psychological Science*. The researchers--Nicholas Christenfeld and Michael Roy--also examined related questions, such as:

- Does the length of time a dog and owner are together contribute to whether they look alike? (apparently, no)
- Which specific features, including hairiness, size, attractiveness, and perceived friendliness, contributes the most in determining a resemblance between dog and owner? (perceived friendliness was "the only feature they identified that might have been useful to the judges.")

It's hard not to be a bit skeptical about this amusing article, given how close the publication date is to April 1. However, several newspapers have run the story, and UCSD has mentioned the research in a press release on its website. Either way, it's a great way to make some basic statistical ideas fun and interesting.

**DISCUSSION QUESTIONS:**

- (1) If the judges are really just guessing, how likely is the outcome of 16 or more matches (out of 25)? How likely is 7 or fewer matches out of 20?
- (2) Do you think the title of the article is appropriate?
- (3) Comment on the research design of this study. Can you identify any potential problems or biases? What other ways might you try to investigate the resemblance between dogs and their owners?
- (4) According to the article, the difference between the pure-bred and mutt results is statistically significant. What does this mean? Also, consider this point in the context of question (1), above.
- (5) Do you see a resemblance?

### **Dan Rockmore and collaborator Digger on another research adventure**



---

#### **[The case of the Cherry Hill cluster.](#)**

***The New York Times Magazine*, March 28, 2004, p. 50**

**D. T. Max**

This article chronicles the efforts of Janet Skarbek to persuade state and federal health agencies to investigate whether mad cow disease could be responsible for the death of at least nine people, all of whom ate beef at the same location sometime between 1988 and 1992. Her quest began after the death of a friend in February 2000. The cause of death was determined to be Creutzfeldt-Jakob disease (C.J.D.), and was categorized as "sporadic" because no compelling direct cause could be determined. According to the article, in the U.S. the chance of developing this form of the disease is about 1 in a million, or roughly 250 to 300 cases per year.

Skarbek became suspicious when, three years later, she read about a woman from her home town who had recently died from C.J.D. and who had worked at the same place as her friend-- the Garden State Race Track in Cherry Hill, N.J. (The track is now closed.) Soon afterwards she located a man from a

neighboring town who had also died of C.J.D. in 2000. The man had held a season pass at the race track and had eaten there once a week. Since at the time there were 100 track employees and 1000 season pass holders, Skarbek reasoned that "the most often we should have seen one sporadic C.J.D. case was every 909 years. Here we've got three!" Eventually Skarbek identified nine people who had died of C.J.D. and who had some connection to the race track; moreover, they all had eaten beef there (usually hamburgers)--perhaps many times-- between 1988 and 1992. Skarbek became convinced that the "cluster" of victims had died not from sporadic C.J.D. but from variant C.J.D., the so-called human form of mad cow disease.

As one might expect, officials at the New Jersey Department of Health and The Centers for Disease Control haven't been terribly receptive to her ideas. Several reasons for this are explored in the article, including biological, political, and mathematical difficulties. For instance, to help explain some of the problems in identifying a "cluster", the author uses a coin-tossing example: "Flip a coin a hundred times and you should expect heads and tails to come up about even. But during those hundred flips you are likely to see a long run of all heads or all tails. If that's the only time you happen to be paying attention -- or if you happen to live near Cherry Hill, N.J. -- you may well think something strange is going on."

### DISCUSSION QUESTIONS:

- (1) How did Skarbek arrive at "909 years"? What assumptions is she making?
- (2) Do you think the coin-tossing example is used effectively? Why or why not?
- (3) Do you think the U.S. government should be investigating more vigorously whether mad cow infected beef is in this country? Why or why not?

### [Google numbers.](#)

#### Sample project for Dartmouth Chance Course

#### Gregory Leibon

Greg Leibon taught the Dartmouth Chance Course this winter and made this sample project for his students. Since we have discussed the Benford's Law a number of times, we thought readers might also enjoy his project. Recall that the Benford's Law is that the first(leading) digit in "natural data" should have approximately the distribution given by base 10 logarithm of  $(1 + 1/d)$  for  $d = 1, 2, \dots, 9$ . Thus the leading digit is 1 with probability  $\log(2) = .301$ , 2 with probability  $\log(1.5) = .176$  etc. The complete distribution is:

First digit	Benford's probability
1	30.1%
2	17.6
3	12.5
4	9.7
5	7.9
6	6.7
7	5.8
8	5.1
9	4.6

For his project, Greg wanted to see if "natural numbers" on the web satisfied Benford's Law. He writes:

I wanted to understand numbers on the World Wide Web in which real live people were actually interested. In particular, I did not want to accidentally include numbers from data sets intended only for data mining purposes. To accomplish this, I included a piece of text in my search. I wanted to choose a natural piece of text, hence (for lack of a better idea) I used the word "nature". Thus, my Google Numbers are numbers that occur on a web page that also includes the word "nature".

I wanted my search to produce robust but reasonable numbers of results. This is because I wanted to leave myself in a position to actually examine the resulting hits in order to achieve a sense for how the numbers were derived.

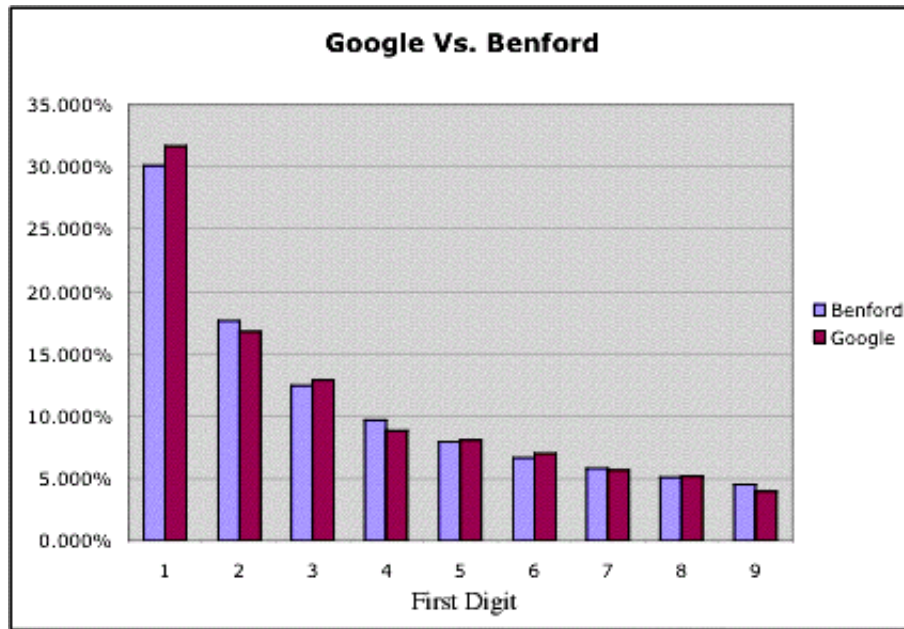
A little experimenting led Greg to the conclusion that searches for six-digit numbers and the word "nature" resulted in a reasonable number of hits. So he chose nine random five digit numbers and for each of these he added all possible leading digits. His first five-digit number was  $x = 13527$  giving him the 9 six-digit numbers 113527, 213527, 313527, ..., 913527. He then searched for each of these numbers and the word "nature" in Google and recorded the number of hits. Here is what he found:

$x = 13527$	occurrences
113527	136
213527	44
313527	35
413527	30
513527	27
613527	15
713527	9
813527	13
913527	8

He repeated this for his 8 other random five-digit numbers and combined the results to obtain:

Leading digit	count	Empirical Percent	Benford
1	645	31.65%	30.1%
2	342	16.78	17.6
3	262	12.86	12.5
4	181	8.88	9.7
5	164	8.05	7.9
6	143	7.02	6.7
7	115	5.64	5.8
8	105	5.15	5.1
9	81	3.97	4.6

This is a remarkably good fit. Here is his graphical comparison:



Greg wondered if he was just lucky or if there was some explanation for such a good fit. Looking for an explanation, he found that many of the numbers he observed could be considered the result of a growth process. As an example of such a growth process, consider the money you have in the bank that is continuously compounded. Then it is easy to check that the percent of time your money has leading digit  $k$  for  $k = 1, 2, 3, \dots, 9$  fits the Benford distribution. Greg remarks:

Hence, we would expect Google numbers to have a Benford distribution if they satisfied two criteria: first that every Google Number behaves like money with interest continuously compounded, and, second that the probability that a Google number is posted on the web is proportional to how long that quantity is meaningful.

We gave Greg an A on his project but you should read it [here](#) yourself. You can also see what Greg did in the Chance Course and some student projects [here](#).

### DISCUSSION QUESTION:

Repeat Greg's experiment replacing "nature" by a different word. Do you get similar results?

### Chew On.

***The New Yorker*, 9 Feb. 2004, The talk of the town; You don't say dept. Pg 22  
Ben McGrath**

### Evaluation of CDs and chewing gum in teaching dental anatomy.

#### [Abstract](#)

**K.L. Allen et al**

The authors of the study initially set out to compare the effectiveness of attending traditional lectures versus using a cd-rom to learn dental anatomy. Apparently the gum manufacturer, Wrigley, was willing to support the study, and for some reason (though perhaps one could guess) the researchers modified their design to also compare the effects of gum chewing on learning. Specifically, students were divided into



four groups: half attended a standard lecture and lab, half used a commercially available instructional CD and lab, and half of each of these groups were required to chew gum. The procedures were continued for three days, and then the students were given both a machine-graded, multiple-choice test as well as a practical exam.

The results were mixed. The 29 students who chewed gum had an average score on the written test that was slightly higher than the average of the 27 non-gum chewing students: 83.6 versus 78.8. The difference on the written test between the CD students ( $n = 30$ ) and the lecture students ( $n = 26$ ) was smaller: 83.7 versus 81.3. (Standard deviations were not provided.) According to the study abstract, on the practical exam there were no differences between groups.

### DISCUSSION QUESTIONS:

- (1) The researchers write that "only the written examination average scores for the gum vs. no gum chewing groups showed differences which appear to be educationally meaningful, though not statistically significant." What do you think this means?
- (2) Do you think the study provides sufficient evidence that gum chewing helps students learn? Why or why not?
- (3) Do you think the study provides sufficient evidence that CD ROMs are as effective as lectures? Why or why not? For this and the previous question, how might a smaller versus a larger exam score standard deviation affect your answer?
- (4) The *New Yorker* article states that "fifty-six students took part in the pilot study, and, as Allen said, 'We really need a sample size of about two hundred to determine [the results] beyond a reasonable doubt.'" Why do you think Allen chose "about two hundred" as a preferred sample size? Do you think that "beyond a reasonable doubt" is an appropriate phrase to use in this context?

The next article was suggested by Joan Garfield

**Is your radio too loud to hear the phone? You messed up a poll.**  
***Wall Street Journal*, 12 March 2004**  
**Sharon Legley**

Legley begins her article with:

(Math-averse readers are allowed to skip this paragraph.) The sampling error represents the range of possible outcomes from a random, representative slice of the population. For practical purposes, it equals 1 divided by the square root of the number of people surveyed. If you poll 1,600 people, then the sampling error is  $1/40$ , or 2.5%.

The bulk of the article is devoted to non-sampling errors. Legley remarks that it is well known that polls tend to include too many women, too many whites, and too many older folks. If women are more likely than men to favor a particular candidate the poll will give a biased result. To avoid this, pollsters adjust for demographic factors by weighting the numbers so they match the census. The Gallup poll has a detailed description of how their polls are carried out available [here](#) where we read:

In most polls, once interviewing has been completed, the data are carefully checked

and weighted before analysis begins. The weighting process is a statistical procedure by which the sample is checked against known population parameters to correct for any possible sampling biases on the basis of demographic variables such as age, gender, race, education, or region of country.

Legley remarks that people are reluctant to indicate how wealthy they are and so this might be a more difficult to correct for though she suggests that zip codes might be a proxy for wealth. She continues:

Worse, polls likely undersample or oversample people in categories the census doesn't count, making an adjustment... virtually impossible. Prof. Gelman's favorite example is surly people. They're more likely to treat a pollster as they would a telemarketer, hanging up and therefore not having their views included. But we don't know how many surly people are in the voting population. If surly people lean toward one candidate, then a poll asking "whom are you most likely to vote for?" will underestimate his support.

Fritz Scheuren, at the *University of Chicago National Opinion Research Center* and president-elect of the *American Statistical Association* considers "no-response" the biggest source of non-sampling error commenting:

If pollsters really wanted to indicate how good their sample is, they'd skip the plus-or-minus-X% and reveal the no-response rate.

Legley says that pollsters don't call cell phones (the owner might be driving) and Scheuren comments

As more and more people have only a cell, you have a problem. Because no one knows how ditching one's land line correlates with political leanings, pollsters can't tell how omitting the cell-only population distorts reality.

We wondered why cell phones could not be used. An [article](#) by John Kamman in the online *Arizona Republic*, Dec. 30,2003, addressed this issue. According to this article, pollsters are indeed quite worried about the fact that they cannot contact those whose only phone is a cell phone. Kamman writes:

For a decade, Federal Communications Commission regulations have restricted pollsters from using modern dialing equipment to call cell phones. And even if they dial by hand, another rule prohibits them from phoning anyone who would have to pay for the call. A violation makes the caller vulnerable to a lawsuit with a penalty of at least \$500.

This issue was complicated by the 2003 FCC ruling giving customers the right to keep their regular phone number when switching phone companies, making it hard for pollsters to know if they are violating the government rules . The FAA says that the pollsters have ways of tracking changes of numbers from wired to wireless but pollsters say this is not so easy.

It is thought that the number of people who now rely only on cell phones is not large, but it obviously involves more younger people which would introduce a bias. It is also pretty obvious that as cell phone usage increases, pollsters will have to find some solution to this problem as well as to the problem of increased use of answering machines and other call-filtering devices. Earl de Berge, research director for the *Behavior Research Center* in Phoenix, suggests an old-fashioned remedy. He remarks: "I wouldn't be surprised to see more door-to-door polling."

It was not clear to us what Legley meant by "a random, representative slice of the population" (see first paragraph). This seems related to a question asked on sci.stat.math: "[Is there a statistical definition of](#)



representative sample?" In this discussion there was little agreement on what representative sample might mean. Some said that it simply meant a random sample but others that it meant a stratified-random sample. Others thought that there is no proper definition. Here is one response:

J Dumais

Organization: Wanadoo, l'internet avec France Telecom

Date: Fri, 3 Nov 2000 19:18:01 +0100

Once upon a time (up to 1925 or about), there was "representative sampling" meaning, "creating a representation of the population in the sample" i.e. quotas. The caricature is that to obtain the so-called representation, you \*have to\* get those 5 middle-age out-of-work with post-secondary education living in towns of less than 20000 people, otherwise you don't have the "true" representation.

With Neyman's theory on disproportional allocation to stratified random samples, "representative sampling" sort of faded away, at least among (mathematical) statisticians dealing with survey sampling.

I have searched in more than 20 textbooks, never coming up with any satisfying "modern" definition of "representative sampling". Basically, (mathematical) statisticians have dropped the notion altogether. When I hear it, I'll question the speaker (if I can) as to what he/she actually means, and it more or less boils down to unbiasedness.

One writer said that, in certain situations, a "representative sample" was required by the FDA so we asked Susan Ellenberg at the FDA if they have a definition of a "representative sample". She replied:

"Representative sample" is used primarily to describe material a product manufacturer must supply to the FDA for testing. It is actually defined as follows, in Volume 21, Part 210 in the Code of Federal Regulations:

(21) Representative sample means a sample that consists of a number of units that are drawn based on rational criteria such as random sampling and intended to assure that the sample accurately portrays the material being sampled.

I have to say I don't think this is the world's greatest definition, but the latter part of the sentence conveys the basic intent. This section of the regulations was initially published in 1978 and was most recently revised in 1993, so the wording is somewhere between 11 and 26 years old.

### DISCUSSION QUESTION:

Well, Google found "representative sample" about 381,000 times so apparently it is not dead. What do you think "a representative sample" means?

### The facts don't matter.

**PBS Ira Glass: *This American Life*, March 12, 2004, Episode 260, Act 2, 43:06, 15 min Sarah Koenig**

In the last segment of this program Sara Koenig visits a John Zogby polling operation to get some idea how the questions are asked in their telephone polls and how they are answered. This poll was described

by Zogby as:

Zogby International conducted telephone interviews of a random sampling of 600 likely primary voters statewide over a rolling three-day period. All calls were made from Zogby International headquarters in Utica, N.Y., from Friday, February 13th through Sunday, February 15. The margin of error is +/- 4.1 percentage points. Slight weights were added to age, race, union and gender to more accurately reflect the voting population. Margins of error are higher in sub-groups.

You will enjoy listening to this segment of the program. We describe the first few minutes to encourage you to do so.

We hear a Zogby interviewer named Boden carrying out his first interview of the day. Sara reports that Boden is interviewing a women, aged 41, union member, separated, college graduate, white, conservative, making between 25 and 50 thousand dollars a year.

Boden: Good afternoon, my name is Boden and I am calling for Zogby International. Today we are doing a poll of Wisconsin voters for Reuters/MSNBC news.

Boden: And how likely are you to vote in the national elections: very likely, somewhat likely, or not likely?

Answer: Very likely

Boden: And the Democratic candidates for 2004 are: Howard Dean, John Edwards, John Kerry, Dennis Kucinich, and Al Sharpton. If the primary were held today, for whom would you vote out of these Democrats?

Answer: Not sure

Boden: O.K. You're sure you're not sure or might you be leaning towards one?

Answer: Not sure

Boden: You're not sure at this point? O.K.

Sara: Boden entered "undecided" in his computer and his screen comes up with this question which is basically the same question asked in a different way.

Boden: And if you **had** to choose today, if you **had** to choose, which candidate might you just be leaning towards: Dean, Edwards, Kerry, Kucinich or Sharpton? Just the slightest leaning towards one of them if you had to choose today.

Answer: Dean

Boden: O.K. Thank you.

This illustrates how hard they work to get an answer. Sara reports that almost 10,000 calls had to be made to get the desired 600 responses.

After the election Sara called a sample of those who responded to the poll to ask them about their experience with the poll and who they finally voted for.

Sara also discusses what she had learned from observing this poll being with Daniel Yankelovich, a pioneer in modern polling who has his own polling company and is the author of the well known book: "Coming to Public Judgment".

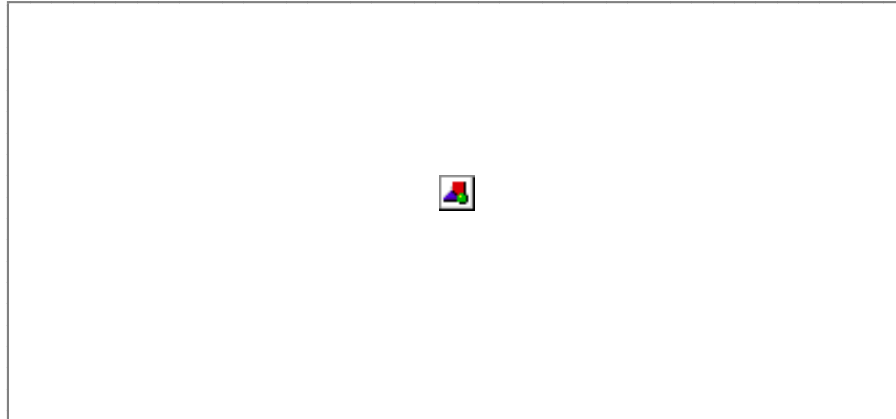
This would be a great program for students to listen to though it might not increase their faith in political

polls.

### DISCUSSION QUESTIONS:

(1) What do you think about the interview?

(2) Here are the results of the Zogby poll released February 16, 2004:



And here are the results of the election held on February 17, 2004:

#### Results: 99% Precincts Reporting

Candidate	Votes	%	Delegates
<b>Kerry *</b>	327,245	40%	30
<b>Edwards</b>	282,978	34%	24
<b>Dean</b>	150,501	18%	13
<b>Kucinich</b>	27,215	3%	0
<b>Sharpton</b>	14,647	2%	0
<b>Clark</b>	12,663	2%	0
Uncommitted	1,136	0%	0

\*projected winner

Were the results within the margin of error? If not, why not?

---

### Betting against the pros.

#### Anonymous

Betting on sports is illegal in many parts of the United States, but nevertheless, millions of Americans bet billions of dollars each year. Many of the betting sites on the Internet are located offshore, beyond the reach of U.S. law. At these sites one can bet on almost any sport. Typically, the site will take a piece of the bet, called the vigorish, or vig. At many sites, the vig is 10%, meaning, for example, if one bets \$11 and wins on an even-money bet, one will only receive \$10 in winnings. (One could argue about whether the vig in this case is 10% or 9.09%, but we won't.) This means that in order to break

even when making bets at such sites, one must win 11 bets out of every 21, on the average. This works out to a winning percentage of 52.38%.

In basketball games, in order to make the bets even the gambling sites give points to the underdog team. In order to win a bet on the favored team, that team must win by more than the number of points assigned to their opponent. Conversely, a bet on the underdog will win as long as that team either loses by fewer points than it has been given (or the underdog wins the game). The team that represents the winning bet is said to have 'beat the spread.' This point spread is set by people known as bookies. Their job is to set the spread so that about half of the bettors will bet on each side, thereby limiting the gambling sites' exposure (and guaranteeing income for the sites, by way of the vig).

In 1989, Colin Camerer wrote a paper containing statistical evidence that the bookies over-emphasized streaks, in that they tended to assign more points than they should have to teams with long losing streaks, and they tended to assign more points to the opponents of teams with long winning streaks than they should have. For example, in the three years' worth of data that Camerer collected, teams who had won at least three games in a row only managed to beat the spread 45.6% of the time against teams with either smaller winning streaks or any type of losing streak. (The reason for this last proviso is, Camerer argued, that if two teams with winning streaks play each other, the one with the longer streak is the 'hotter' team, in the eyes of the bookies and the betting public.)

Thus, had someone bet against the 'hot' team, because it was thought that the bookies were over-valuing the streak, the bettor would have won 53.4% of the bets. The corresponding results involving 'cold' teams were as follows. Teams with at least three-game losing streaks beat the spread against less 'cold' teams 340 out of 643 games, for a winning percentage of 52.9%.

As usual, we have to check whether the above observations are significant. There were 698 games that formed the first observation, and if the point spreads were set so that each team had a 50% chance of winning in a given game, then one would expect to see about 349 of the 'hot' teams and 349 of their opponents win. In fact, 45.6% of 698 is 318, and the probability that one would see 318 or fewer wins by the 'hot' teams is about 1%. The corresponding p-value for the observation involving 'cold' teams is about 7%.

Your intrepid Chance News editors decided to see what the corresponding data looked like in the current NBA (National Basketball Association) season. The results are not pretty. Suppose, for example, that we bet against the teams with winning streaks of at least three games (if they are playing a team that is not as 'hot' as they are). Through March 24, 2004, we would have lost 128 out of 242 bets, for a winning percentage of 47.1%. If one had bet on teams with at least three-game losing streaks, figuring that the bookies were undervaluing these teams, one would have won 111 out of 243 bets, for a winning percentage of 45.7%. The p-values of these observations are .18 and .09.

What is interesting is that now the teams on winning streaks are beating the spread more than half the time, while the teams on losing streaks are failing to beat the spread more than half the time, which is the reverse of what happened in Camerer's data set.

One could interpret the new data to mean that the bookies have learned over the years not to over-emphasize streaks. In fact, one might be tempted to bet in favor of 'hot' teams and against 'cold' teams, based on the current data.

As we were perusing the Web to find out what was happening in the world of sports betting, we came across a website with a fairly impressive record. This website had made a study of professional basketball games and had determined that the average number of 'possessions' that a team has over the course of a

season is well correlated with that team's winning percentage. The number of possessions of a team in a game is defined by the following formula:  $\text{Number of Possessions} = \text{Field Goal Attempts} + \text{Turnovers} - \text{Offensive Rebounds} + (\text{Free Throws}) \times .44$ .

Ignoring the .44 for the moment, we can interpret this formula as follows. If a team takes the ball downcourt, one of three things can happen: it can attempt a field goal, it can make a turnover, or there can be a foul called. Once a field goal is attempted, the team keeps the ball only if it misses the field goal attempt and gets an offensive rebound, or it gets fouled. So, for example, if a team takes the ball downcourt and attempts three field goals, getting two offensive rebounds in the process, this sequence contributes 1 to the above sum. The .44 seems to have been computed by a linear regression of the above summands against the team's scores.

After this was posted reader Ann Azevedo suggested the following alternative explanation for the .44:

As the .44 is only a multiplier for free throws, and, since the number of free throws attempted in any given possession can vary:

1 free throw - if a free throw is awarded after a made basket- if the front end of a 1-and-1 free throw is missed or

2 free throws - awarded for a foul on a missed 2-point field goal- awarded for a foul not on a field goal attempt when in the double-bonus (10 or more team fouls)- awarded for making the front end of a 1-and-1, and thus being awarded a second free throw or

3 free throws - awarded for a foul on a missed 3-point field goal

I would think the 0.44 is a factor for correcting the number of free throws to the number of possessions they represent.

The website uses this information to predict the score of each game. This predicted score can be compared to the point spread published by the bookies at the gambling websites. For example, suppose San Antonio is playing Philadelphia, and the website predicts that the score will be 101-90, with San Antonio winning. If the bookies are giving fewer than 11 points to Philadelphia, the website suggests a bet should be made on San Antonio.

The website also separates out some games that are labeled 'best bets.' These are games in which the website's point spread differs from the bookies' spread by 4.5 points or more. So for example, suppose that the website predicts that Milwaukee will beat New Orleans by 10 points, but the bookies are only giving 5 points to New Orleans. The difference between the two point spreads is 5 points, so this is a 'best bet' on Milwaukee.

In the 2002-03 season, the website was right on 52.2% of all of the games and on 55.9% of the best bets (there were 295 best bet games, and the website predicted the correct team in 165 of these games). The p-values of these observations (against the null hypothesis that the website does no better than 50%) are .10 and .02. When we discovered this website (in December of 2003), the website's winning percentages for all games and best bet games were about 53% and 58%, respectively. Thus, they were similar to the percentages recorded in the preceding season.

To further the cause of applied statistics, two of us got together and put a small amount of money (each contributed \$100) down on the ensuing best bet as predicted by this website. Before we started betting, the website's win-loss record on best bet games was 90-73, for a winning percentage of 55.2%.

What happened next can be only classified as an unmitigated disaster. In the ensuing two months, before we went broke, the best bet record was 45-55, and it's not very hard to figure out what that winning percentage is. In fact, since January 1, 2004, the website's best bet record is 65-87, for a winning percentage of 42.8%, and their overall record is 238-303, for a winning percentage of 44.0%. Luckily, some of us have quite a few years of gainful employment left to make up for our losses in this endeavor.

## REFERENCES:

Camerer, C. F. (1989a). Does the basketball market believe in the 'hot hand'? *American Economic Review*, 79, 1257-1261. Available from [JSTOR](#).

---

Gerry Grossman suggested this article:

**[1-2-3, NHL playoff teams will be ...](#)  
denews.com, MoreSports, Sunday March 7, 2000**

This article describes a mathematical research project that Oakland University student Dan Steffy and Professor Eddie Cheng are carrying out to determine, on any given day, which teams in the NHL have been mathematically eliminated from the playoffs. This turns out to be a hard problem--in fact it is an NP-complete problem. Thus it is not surprising that newspapers sometimes get it wrong. The aim of this project is to implement an algorithm to answer this and related questions and to make the results available to the media.

Here is what Steffy and Cheng say about their project:

In the NHL tournament there are 30 Teams, which are arranged in five team divisions. There are two conferences, the East and West, each containing 3 divisions. Teams are awarded two points for a win, one point for a tie, and zero points for a loss, and one point for an over time loss. After the regular season is completed eight teams are chosen from each conference to advance to the tournament. In each conference, qualifying teams are made up of the leader of each division, the team in each division with the most points, and the next five highest ranked teams in the conference according to score. Special rules are also employed to break ties between teams who have the same amount of points.

We developed a mixed integer program formulation for the following two problems -- the Guaranteed Qualification Problem: is a given team guaranteed a place in the finals, and the Possible Qualification Problem: can a given team have a chance of qualifying.

In addition we modified our formulations to solve the Guaranteed Qualification and Possible Qualification problem for: Division Leader Status, Conference Leader Status, and Presidents' Trophy.

You can see the current standings as computed by their program [here](#) and more about their project [here](#).

Of course, this isn't a statistics or probability question, but it could become one if we tried at the same time to estimate the probability that a given team reaches the finals. This might be of greater interest to the bookies.

## DISCUSSION QUESTION:



How would you estimate the probability that, at a given time, a particular team would reach the finals?  
Would the program of Steffy and Cheng help you?

---

Norton Starr brought our attention the following announcement from the *Royal Statistical Society*:

**The communication of risk.**  
***Statistics in Society* (JRSS series A), June 2003**

This issue of the *Journal of the Royal Statistical Society Series A* features a collection of short papers on the communication of risk, with guest associate editors D. R. Cox and S. C. Darby:

<b>Title and authors</b>	<b>Page number</b>
<i>The communication of risk</i> D. R. Cox and S. C. Darby	203
<i>Introduction to the papers on 'The communication of risk'</i> A. F. M. Smith	205
<i>Human immunodeficiency virus risk: is it possible to dissuade people from having unsafe sex?</i> J. Richens, J. Imrie and H. Weiss	207
<i>Communicating risk—coronary risk scores</i> I. M. Graham and E. Clavel	217
<i>Tobacco—the importance of relevant information on risk</i> S. C. Darby	225
<i>Tobacco: public perceptions and the role of the industry</i> D. Simpson and S. Lee	233
<i>Communication of risk: health hazards from mobile phones</i> D. R. Cox	241
<i>Crime victimization: its extent and communication</i> P. Wiles, J. Simmons and K. Pease	247
<i>Accidental fatalities in transport</i> A. W. Evans	253
<i>Communicating the risks arising from geohazards</i> M. S. Rosenbaum and M. G. Culshaw	261

The RSS comments:

To assist in the wider appreciation of the issues raised, a short commentary on the papers has been commissioned from the well-known science writer, Geoff Watts.

This commentary is available [here](#).

These articles are all very good and are available electronically if your library subscribes to the journal. They are short articles which describe the issues involved and any of them could be the basis for an interesting discussion in a statistics class. We illustrate this by discussing the article on coronary risk

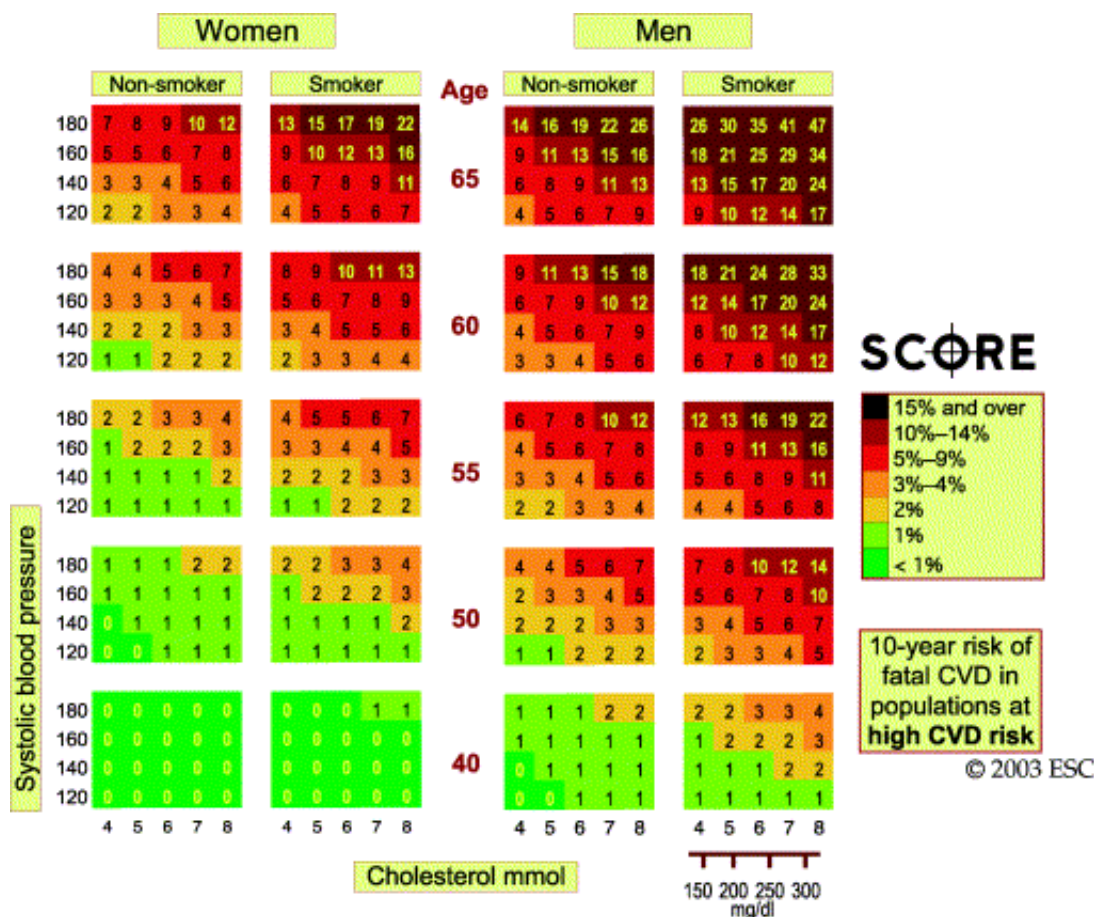
scores.

Heart disease is the largest cause of deaths in adults in their middle years and older in most European countries. Risk factors such as smoking, high cholesterol levels, and high blood pressure are well known for heart disease. A physician is typically not in a position to assess the overall risk when several risk factors are taken into account. To assist in this a number of charts, score cards, computer programs etc. have been developed to assist the physician. There is a bewildering array of these tools using different risk factors and based on different data.

In 1994 the *European Society of Cardiology* developed a chart from which the probability of coronary heart disease in the next ten years could be determined, given the patients age, sex, blood pressure, total cholesterol level and smoking status. This was widely used in Europe. However, experience with this chart showed some problems. It was based on a relatively small data set of about 5000 people from the Framingham study so some risk combinations had to be based on very little data. Also, in some countries the risk of heart disease is lower than others. One such country is Italy. Researchers in Italy constructed a chart just like the chart that was being used in Europe but based on data from their country. The resulting chart looked quite different from the one they had been using.

So, in 1998 a new chart called SCORE was developed and is now in use. The new chart was designed to estimate the probability of death in the next ten years from any cardiovascular event. It was based on 12 European cohort studies which involved over 200,000 subjects and contain some 3 million person-years of observation and more than 7000 fatal cardiovascular events.

Here is the resulting chart:



## Ten year risk of fatal CVD in high risk regions of Europe by gender, age, systolic blood pressure, total cholesterol and smoking status.

You can see a similar chart for low risk European countries [here](#).

Here are the instructions on how to use the chart as given in De Backer et al [2].

- The low risk chart should be used in Belgium, France, Greece, Italy, Luxembourg, Spain, Switzerland and Portugal; the high risk chart should be used in all other countries of Europe.
- To estimate a person's total ten year risk of CVD death, find the table for their gender, smoking status and age. Within the table find the cell nearest to the person's systolic blood pressure (mmHg) and total cholesterol (mmol/l or mg/dl).
- The effect of lifetime exposure to risk factors can be seen by following the table upwards. This can be used when advising younger people.
- Low risk individuals should be offered advice to maintain their low risk status. Those who are at 5% risk or higher or will reach this level in middle age should be given maximal attention.
- To define a person's relative risk, compare their risk category with that of a non-smoking person of the same age and gender, blood pressure less than 140/90 mmHg and total cholesterol less than 5mmol/l(190mg/dl).
- The chart can be used to give some indications of the effect of changes from one risk category to another, for example when the subject stops smoking or reduces other risk factors.

So let's see what my (Laurie's) risk is. Laurie is a male, does not smoke, has systolic blood pressure 130 and total cholesterol 226. If Laurie were in his fifties he would have only a 1% chance of dying of a cardiovascular event. If he were in his sixties he would have a 5% chance. However Laurie is in his 70's, so over the chart. Thus he will have to live with (die with?) the 5%. Note that he could probably get it down to 2% by taking more pills to lower his cholesterol.

Armed with this ammunition Laurie asked his friendly doctor if he should start taking more pills. His doctor drew several sharply rising curves saying: The first curve is for the risk of dying of a heart attack as a function of age, the second for dying of cancer, the third for dying of parkinson's disease, etc. Taking away one of these won't help all that much and many people consider a heart attack as a pretty good way to go! So Laurie is not taking more pills.

### REFERENCES:

[1] Conroy et al, [Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project](#). *European Heart Journal*. 2003;24:987-1003.

[2] De Backer et al. [European guidelines on cardiovascular disease prevention in clinical practice](#). *European Heart Journal*, 2003;24: 1601-1610

### DISCUSSION QUESTION:

What do you think of Laurie's doctor's analysis?

**Ask Marilyn**  
**Parade Magazine, 28 March, 2004**  
**Marilyn vos Savant**

Readers of Marilyn's column now know something about the history of probability from the following multiple choice question that she provided at the end of this column:

In the 17th century, French mathematicians Pierre de Fermat and Blaise Pascal together developed modern probability theory in the course of what activity?

- A. Playing spin-the-bottle with a milkmaid.
- B. Predicting the succession of King Louis XIV.
- C. Counterfeiting the first scratch-off lottery tickets.
- D. Answering a gambler's query about why he lost money on dice.

But, as usual, we ask "Does she have it right?"

In his article "Pascal and the Invention of Probability Theory" [1] Oystein Ore writes:

Most textbooks on probability feel obliged to include a brief account of the history of the subject. Their descriptions of this process of initiation usually run somewhat in the following vein: "In the year 1654 a gambler name de Mere proposed to Pascal two problems which he had run across in his experiences at the gaming table".

It is likely that the distinguished Antoine Gombaud chevalier de Mere, sieur de Baussay, would turn in his grave at such a characterization of his main occupation in life. He certainly considered himself a model of courtly behavior and taught his esthetic principles elegantly to the *haut monde* as one may see from the frontispiece of his collected works. His writings...have secured him a permanent niche in the French literature of the seventeenth century.

One of the problems proposed to Pascal was a "dice problem" and the second was "the problem of points" which had a long history going back to the 15th century. Pascal consulted Fermat about these two problems and their correspondence is considered by many to be the beginning of probability theory. Their letters, translated into English, can be found in F.N. David's book "Gods, Games, and Gambling"[2] and are available on the web [here](#) as part of the University of York's [Materials for the History of Statistics](#).

An example of what Ore thinks might make de Mere turn in his grave can be found in Grinstead and Snell's book [Introduction to Probability](#). Here we read:

It is said that de Mere had been betting that, in four roles of a die, at least one six would turn up. He was winning consistently and, to get more people to play, he changed the game to bet that, in 24 rolls of two dice, a pair of sixes would turn up. It is claimed that de Mere lost with 24 and felt that 25 rolls were necessary to make the game favorable.

But this is called the "de Mere legend" by Maistrov [3 ] in his history of probability book. He states that the legend in detail is presented in a story by Khinchin and Yaglom titled "The story of the Knight de

Mere." Maistrov provides this story with some omissions. He also provides references to the original story (in Russian). Maistrov argues that de Mere did not turn to Pascal with a problem from an actual gambling experience, but with a purely theoretical question.

So the question that we should ask Marilyn is: did Pascal pose the dice problem because he was a gambler or because he was interested in probability theory?

Since the first letter from Pascal to Fermat is missing, we have to try to determine this from later letters. In Fermat's reply to the missing first letter, he gives his way to determine the chance of getting a six in a specific number of rolls of a die and seems to suggest that Pascal got it wrong. Fermat:

Fermat to Pascal  
1654 [undated]  
Monsieur

Sir, if I undertake to make a point with a single die in eight throws, and if we agree after the money is put at stake, that I shall not cast the first throw, it is necessary by my theory that I take  $1/6$  of the total sum to be impartial because of the aforesaid first throw.

And if we agree after that that I shall not play the second throw, I should, for my share, take the sixth of the remainder that is  $5/36$  of the total. If, after that, we agree that I shall not play the third throw, I should to recoup myself, take  $1/6$  of the remainder which is  $25/216$  of the total.

And if subsequently, we agree again that I shall not cast the fourth throw, I should take  $1/6$  of the remainder or  $125/1296$  of the total, and I agree with you that that is the value of the fourth throw supposing that one has already made the preceding plays.

But you proposed in the last example in your letter (I quote your very terms) that if I undertake to find the six in eight throws and if I have thrown three times without getting it, and if my opponent proposes that I should not play the fourth time, and if he wishes me to be justly treated, it is proper that I have  $125/1296$  of the entire sum of our wagers.

This, however, is not true by my theory. For in this case, the three first throws having gained nothing for the player who holds the die, the total sum thus remaining at stake, he who holds the die and who agrees to not play his fourth throw should take  $1/6$  as his reward.

And if he has played four throws without finding the desired point and if they agree that he shall not play the fifth time, he will, nevertheless, have  $1/6$  of the total for his share. Since the whole sum stays in play it not only follows from the theory, but it is indeed common sense that each throw should be of equal value.

I urge you therefore to write me that I may know whether we agree in the theory, as I believe we do, or whether we differ only in its application.

I am, most heartily, etc.,  
Fermat.

Fermat is calculating the probability of obtaining a 6 for the first time on the  $k$ th toss when you toss a die 8 times. Thus if you roll the die 8 times the probability that it comes up in the first three rolls is  $1/6 + 5/36 + 125/1296$ . Today we would use the concept of independence to calculate the probability that you

don't get a six by the third roll as  $(1-1/6)^3$  and subtract this from 1 to get the probability that you do get a six in the first 3 rolls.

In a letter dated Wednesday July 29, 1654, responding to Fermat's letter, Pascal writes:

I have no time to send you the proof of a difficult point which astonished M. (de Mere) so greatly, for he has ability but he is not a geometer (which is, as you know, a great defect) and he does not even comprehend that a mathematical line is infinitely divisible and he is firmly convinced that it is composed of a finite number of points. I have never been able to get him out of it. If you could do so, it would make him perfect. He tells me then that he has found an error in the numbers for this reason .

If one undertakes to throw a six with a die, the advantage of undertaking to do it in 4 is as 671 is to 625. If one undertakes to throw double sixes with two dice there is a disadvantage of undertaking it in 24 throws. But nonetheless, 24 is to 36 (which is the number of faces of two dice)<sup>2</sup> as 4 is to 6 (which is the number of faces of one die).

This is what was his great scandal which made him say haughtily that the theorems were not consistent and that arithmetic was demented. But you will easily see the reason by the principles which you have.

<sup>2</sup>[Clearly, the number of possible ways in which two dice can fall.]

Ore explains de Mere's reasoning as:

Pascal does not understand de Mere's reasoning, and the passage also has been unintelligible to the biographers of Pascal. However, de Mere bases his objection upon an ancient gambling rule which Cardano also made use of: One wants to determine the critical number of throws, that is, the number of throws required to have an even chance for at least one success. If in one case there is one chance out of  $N_0$  in a given trial, and in another one chance out of  $N_1$ , then the ratio of the corresponding critical numbers  $n_0$  and  $n_1$  is as  $N_0:N_1$ . That is we have

$$n_0:N_0 = n_1:N_1.$$

Well, that's all the evidence we have. We leave it to the reader to decide if de Mere was asking a theoretical question or a question based on his gambling experience.

## REFERENCES

[1] Oystein Ore. Pascal and the Invention of Probability Theory, *American Mathematical Monthly*. v. 47, p. 409-19, May 1960. Available from [JSTOR](#).

[2] David F.N., Games, Gods & Gambling: A History of Probability and Statistical Ideas, Dover, 1998.

[3] Maistrov, L. E. (Leonid Efimovich). Probability theory; a historical sketch. Translated by Samuel Kotz from Teoriia veroiatnostei. Academic Press, New York, 1974.

## DISCUSSION QUESTIONS:

(1) Show that betting that two sixes will turn up in 25 rolls of a die is an unfavorable bet and betting that



two sixes will turn up in 24 rolls of two dice is an unfavorable bet.

(2) In his article Ore writes:

de Mere believed that the smallest advantageous number of throws should be 24. As the matter has been presented, he turned to Pascal because his own experiences had shown him that 25 throws were required. This is an unreasonable explanation. The difference between the probabilities for 24 and 25 throws is so small, as we have just seen, that to decide experimentally that one of them is less than  $1/2$  would, according to modern statistical standards, require at least 100 sequences of trial, which in turn would involve several thousand individual throws with the two dice. Besides, the dice would have to be specially made in order to show no bias; the usual bone cubes turned out by the diciers of Paris would be much too inaccurate. To prepare special equipment of this kind and to keep the tedious records involved was evidently contrary to the chevalier's temperament.

Determine either by simulations or theoretically, or both, that betting that 2 sixes will come up in 24 rolls is not a favorable game while with 25 rolls it is.

(3) Do you think Marilyn got it right?

---

Copyright (c) 2004 Laurie Snell

This work is freely redistributable under the terms of the GNU [General Public License](#) published by the [Free Software Foundation](#). This work comes with ABSOLUTELY NO WARRANTY.

**CHANCE News 13.03**  
**March 10, 2004 to April 17, 2004**

---