STATS

THE MAGAZINE FOR STUDENTS OF STATISTICS

FALL 2005/WINTER 2006 • ISSUE # 44

STATISTICS in sports

Does Batting Average Measure Ability or Luck?

An NFL Cookbook: Quantitative Recipes for Winning

Is Lance Armstrong the Greatest Cyclist Ever?

Sponsor your students'

ASA MEMBERSHIP!

For only **$10** each of your students gets one year of ASA membership!

Students can become members of the American Statistical Association at the special rate of **$10 for one year** or **$20 for two years** of membership and only $25 per year thereafter.

New Students Receive a subscription to *Amstat News* and *STATS: The Magazine for Students of Statistics;* discounts on all ASA publications, meetings, and products; access to job listings and career advice; online access to the *Current Index to Statistics (CIS);* and networking opportunities to increase their knowledge and start planning for their futures in statistics.

SPONSOR INFORMATION Member ID____________________

Organization/Department____

First Name________________________ Last Name________________________

Address____

City________________________ State/Province________________________

ZIP/Postal Code________________________ Country________________________

Phone________________________ Email________________________

STUDENT MEMBERS *(Attach additional pages, if necessary, or email student contact information to asainfo@amstat.org)*

Name________________________
Address________________________
City________________________
State/Province________________________
Country____________ZIP/Postal Code____________
Phone____________Email____________

Name________________________
Address________________________
City________________________
State/Province________________________
Country____________ZIP/Postal Code____________
Phone____________Email____________

Name________________________
Address________________________
City________________________
State/Province________________________
Country____________ZIP/Postal Code____________
Phone____________Email____________

Name________________________
Address________________________
City________________________
State/Province________________________
Country____________ZIP/Postal Code____________
Phone____________Email____________

PAYMENT INFORMATION: **Total:** $10/1 yr. x _____ (# of students) and/or $20/2 yrs. x _____ (# of students) = $__________

□ Check/money order *(payable to the American Statistical Association in U.S. dollars drawn on U.S. bank)*

Credit Card □ VISA □ MasterCard □ American Express

Name on Card____

Card Number________________________________ Exp. Date______/______

Signature of Cardholder____

STATS

MAIL: American Statistical Association, Dept. 79081, Baltimore, MD 21279-0081
FAX: (703) 684-2037 **CALL:** 1 (888) 231-3473

STATS

The Magazine for Students of Statistics

Fall 2005/Winter 2006 • Number 44

STATS: The Magazine for Students of Statistics (ISSN 1053-8607) is published three times a year, in the winter, spring, and fall, by the American Statistical Association, 1429 Duke St., Alexandria, VA 22314-3415 USA; (703) 684-1221; fax (703) 684-2036; Web site *www.amstat.org*.

STATS is published for beginning statisticians, including high school, undergraduate, and graduate students who have a special interest in statistics, and is distributed to student members of the ASA as part of the annual dues. Subscription rates for others: $13.00 a year for members; $20.00 a year for nonmembers.

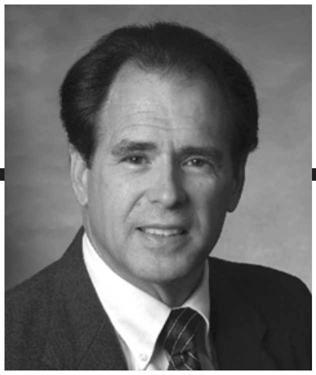
Ideas for feature articles and material for departments should be sent to the Editor at the address listed above. Material must be sent as a Microsoft Word document. Accompanying artwork will be accepted in graphics format only (.jpg, etc.), minimum 300 dpi. No material in WordPerfect will be accepted.

Requests for membership information, advertising rates and deadlines, subscriptions, and general correspondence should be addressed to *STATS* at the ASA office.

Features

Departments

EDITOR'S COLUMN



Paul J. Fields

Baseball is called the "Great American Pastime." It certainly deserves that title. However, it appears to me that sports statistics is an even bigger national pastime. Do you know someone who can tell you the batting averages of the top hitters in baseball history, or what was the longest baseball game ever played was, who played in it, where, and when? I do. I always have been amazed at sports fans' knowledge of sports statistics. Sports statistics are the ammunition used in the eternal debates about who's the best.

My introduction to statistics started with baseball—learning to calculate my batting average playing Little League baseball. I learned it was immensely important to be able to do that, and I learned how to do the calculations quickly to update my average after every game.

In his book *Numbers Game: Baseball's Lifelong Fascination with Statistics*, Alan Schwarz presents a history of baseball statistics that shows how baseball and numbers seem to be inseparable. It is fascinating to read about the origins of sports statistics. (2004, St. Martin's Press, ISBN 0-312-322224)

The questions in sports statistics are endless:

- Who had the most career hits in baseball? How many hits did he have?
- Where is the largest sports stadium in the world?
- What was the largest attendance ever at a sporting event?
- Are there more visits to hospital emergency rooms each year due to basketball or football?
- In the Mexico City Olympics, Bob Beaman broke the world record in the long jump by 2 feet! Can you calculate the probability of a man jumping an astounding 29 feet, 2.5 inches in 1968?
- During his career, Edwin Moses compiled a record of 107 consecutive wins in the 400-meter hurdles. Think about it. What are the chances of a streak that long?
- Speaking of streaks, did you know that Lou "Iron Man" Gehrig played in a record 2,130 consecutive baseball games?

In this issue of *STATS*, Jim Albert looks at the question, "Does a batting average measure a hitter's ability or luck?"

Matthew Strand shows how scatterplots can tell us about winning long-distance races, such as the NCAA Cross Country Championships and the New York City Marathon.

What does it take to win in professional football? Patrick Bartshe uses regression analysis to tackle this question as he describes the recipes for winning in the National Football League.

We welcome Schuyler Huck to the *STATS* editorial board. As the *STATS* Puzzler, he challenges our statistics skills by asking us to compute the batting averages of Bo, Jo, and Mo. See if you can hit the Puzzler's curve ball.

The boxer Mohammad Ali once proclaimed, "I'm the Greatest." How do you measure greatest? In July, Lance Armstrong won cycling's premier event, the Tour de France, for an unprecedented seventh straight time. Peter Flanagan-Hyde asks, "Is Lance the Greatest Cyclist Ever?" See his answer in AP Statistics.

The use of performance-enhancing drugs by professional athletes has been a hot topic this past year. One public opinion poll by the Associated Press showed two-thirds of the American public was in favor of denying a place in the Baseball Hall of Fame to anyone who used steroids. In a new *STATS* department, R U Simulating?, Bruce Trumbo shows how to use simulation to answer questions about sample sizes in public opinion polls. Try the statistical challenges he presents at the end. Then, send us your answers and you could win an American Statistical Association T-shirt!

In this issue's Ask *STATS*, Jackie Miller says, "Hey Coach, I have some questions..." She has some answers from sports statistics expert Robin Lock about basketball, golf, and—of course—baseball.

Taking a swing at the question of who's the best, Chris Olsen reviews *Baseball's All-time Best Sluggers* in Statistical µ-sings and shows the connection between analyzing the performance of baseball hitters and the performance of public schools.

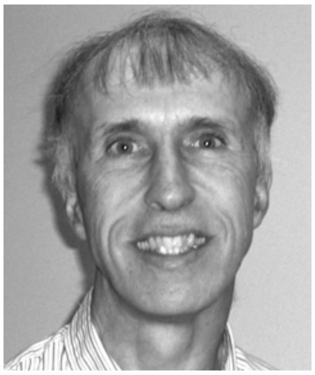
To find out more about sports statistics, visit the Statistics in Sports Section of the American Statistical Association's web site at *www.amstat.org/sections/sis*. Below are some more great places to find sports statistics and data to analyze:

www.sportsstats.com
www.infoplease.com/sports.html
www.bballsports.com
www.baseball-reference.com
www.baseball-almanac.com

So, "step up to the plate" and have some fun with sports statistics.

Paul J. Fields

Does a Baseball Hitter's Batting Average Measure Ability or Luck?



Jim Albert

In baseball, statistics are used to evaluate the performance of players. When people talk about the best hitter in baseball, it is common to think of the "best" as the player possessing the highest batting average. Ever since the early years of professional baseball (more than 100 years ago), the batting average has been the standard measure of hitting effectiveness. For instance, if you click on the "Stats" link on the official Major League Baseball (MLB) web site, *www.mlb.com*, you will see "MLB Leaders: Hitting"—the three players with the highest batting averages this season—at the top of the page.

Batting Average

When a batter faces a pitcher, it is called a plate appearance (PA). Ignoring some rare events, such as sacrifices and hit-by-pitches, there are two results of a PA: the player gets a walk (denoted by BB for "base on balls") or the player has an official at-bat (AB). When the player has an AB, two things can happen: the player strikes out (SO) or the player puts the ball in-play (IP). When a ball is in-play, there can be an OUT based on a fielder's play or there can be one of four base hits: single (1B), double (2B), triple (3B), or home run (HR). The outcomes of a plate appearance are shown graphically in Figure 1.

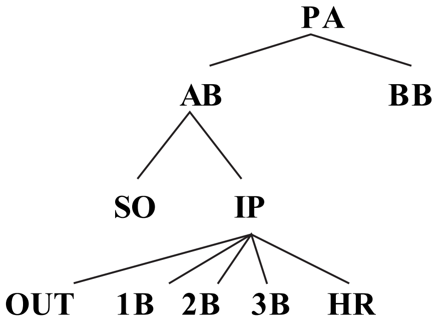


Figure 1. Outcomes of a player's plate appearances

A player's total hits (H) are the sum of 1B + 2B + 3B + HR. Therefore, a player's batting average is:

$$\text{Batting Average} = \frac{H}{AB}$$

Decomposing the Batting Average

In a plate appearance, a batter's first objective is to avoid striking out. The batter's second objective is to hit the ball in a place that is not reachable by the fielders. Using this analogy, Eric Bickel, in an article in the *Baseball Research Journal*, represents the batting average (AVG) as:

$$\begin{aligned} AVG &= \frac{H}{AB} \\ &= \frac{IP}{AB} \text{¥} \frac{H}{IP} \\ &= \left(1 - \frac{SO}{AB}\right) \text{¥} \frac{H}{IP} \end{aligned}$$

We see from this equation that a batting average is dependent on two ratios: the rate of striking out (SO/AB) and the rate of getting a ball to fall in for a hit (H/IP). Bickel looked at the pattern of the strikeout rate and the in-play batting average throughout the history of baseball. He showed that the chance of getting a batted ball to fall in for a hit has remained fairly constant throughout the last 100 years. In contrast, there have been dramatic changes in the strikeout rate over this same period.

Ability or Luck

Jim Albert and Jay Bennett in the book *Curve Ball* consider a number of alternative measures of hitting performance. One general theme of *Curve Ball* is the predominant role of luck, or chance variation, in the game

Jim Albert (albert@bgnet.bgsu.edu) is Professor of Mathematics and Statistics at Bowling Green State University. His research interests are in Bayesian modeling, statistical education, and the application of statistics in sports (especially baseball). Currently, he serves as editor of The American Statistician. *He enjoys playing tennis and is an avid baseball fan. He is patiently waiting for the Phillies to win the World Series.*

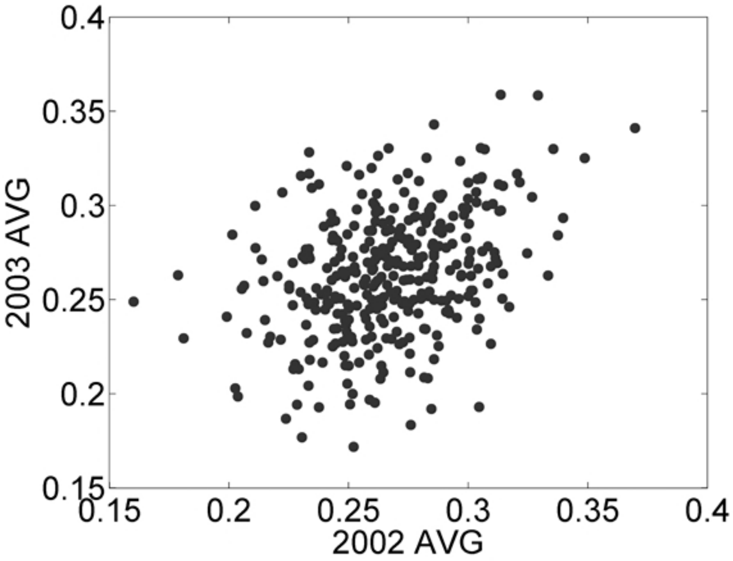


Figure 2. Scatterplot of the 2002 and 2003 batting averages of all players who had at least 100 at-bats each season

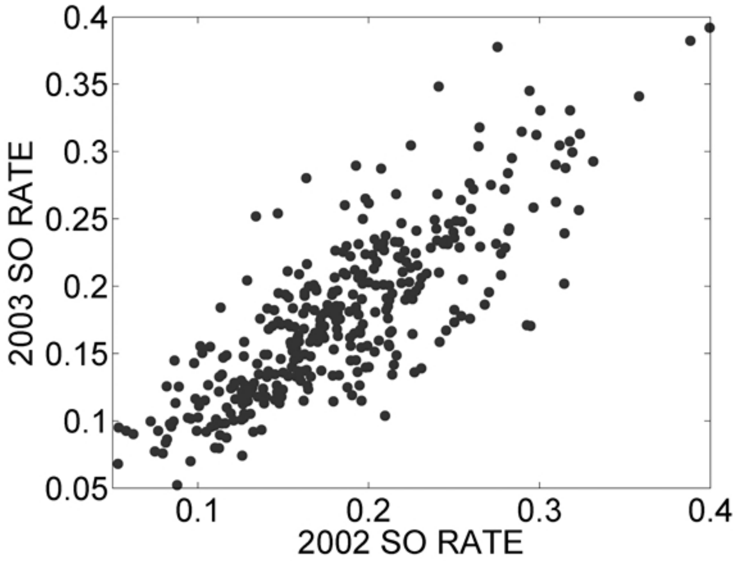


Figure 3. Scatterplot of the 2002 and 2003 strikeout rate of all players who had at least 100 at-bats each season

of baseball. So, let's take a deeper, more statistical look at the batting average. In particular, let's ask:

How useful is an observed batting average in describing the ability of a batter? In other words, how much of a batter's batting average is determined by skill and how much is determined by luck?

Generally, the variation among players' performances based on a given statistic, say the batting average, is due in part to differences in players' abilities to hit, with the remaining variation attributable to chance. One way of assessing how much 'ability' is contained in a hitting statistic is by exploring two years of hitting data. If a particular statistic is a good measurement of the ability of a player, we would expect players to have similar values of the statistic for two consecutive years. Let's look at two hitting statistics for all players with at least 100 at-bats for the 2002 and 2003 seasons. In Figure 2, we construct a scatterplot of the batting averages for the two seasons, and in Figure 3, we construct a similar scatterplot of the strikeout rates. Note there is a relatively weak association in the 2002 and 2003 batting averages. In contrast, there is a much stronger positive association between the 2002 and 2003 strikeout rates. This indicates the strikeout rate is more closely related to a batter's ability than the player's batting average.

Data Analysis

Based on the possible outcomes of a plate appearance and the decomposition of the batting average, there are a number of hitting statistics one can compute for baseball players. A player can be evaluated by means of:

- ***Walk Rate:*** the ratio of walks to plate appearances (BB/PA);
- ***Strikeout Rate:*** the ratio of strikeouts to at-bats (SO/AB);
- ***Batting Average:*** the fraction of at-bats that are hits (H/AB); or
- ***On-base Fraction:*** the ratio of hits and walks to plate appearances [(H + BB)/PA].

A player also can be evaluated by "in-play" statistics—such as *In-play Average*, *In-play Single Rate*, *In-play Doubles+Triples Rate*, and *In-play Home Run Rate*—found respectively by dividing the counts of hits, singles, etc. by the number of balls put in play.

Each of these eight hitting statistics measures a player's hitting ability to some extent. But, as we have seen, some statistics are better measures of ability than others.

Fitting a Random Effects Model

To evaluate the usefulness of hitting statistics, we want to know how much of the variation in a particular statistic for a given season is due to ability and how much is due to chance. Toward this goal, we fit the following random effects model.

Suppose we have N players with different abilities and we assume these abilities $p_1,\ldots, p_N$ come from a probability distribution g(p) (see Figure 4). We can think of these abilities as the true probabilities of a successful outcome. For example, if we are considering batting average and a given player has an ability of p_1, then p_1 represents Player 1's probability of getting a base hit. A second player with a better hitting ability than Player 1 would have a probability p_2 that is larger than p_1. Once the probabilities are known, the actual numbers of successes (say, base hits in our batting average illustration) $y_1,\ldots, y_N$ have binomial distributions, where y_i is distributed binomially with sample size n_i and probability of success p_i. This is similar to the results of tossing N coins, where the probabilities of heads of the N coins are variable and come from a probability distribution g. Either getting a hit or not when batting is similar to getting a head or not when flipping coins.

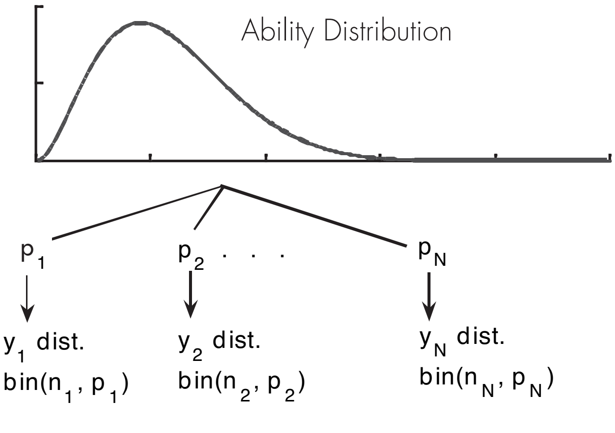


Figure 4. Graphical representation of a random effects model. The batting abilities $p_1,\ldots,p_N$ come from an ability distribution, and given a player's ability p_i, the number of observed successes y_i has a binomial distribution.

For all of our batting performance measures, a batter's ability will be a proportion that falls between 0 and 1. It is convenient to assume the ability distribution has the beta form:

$$g(p)\ \mu\ p^{K\eta-1}(1-p)^{K(1-\eta)-1},\ 0 < p < 1,$$

where η represents the mean or average ability of all players and *K* gives some indication of the spread of the abilities. Small values of *K* indicate there is a wide range of abilities among ballplayers, and large values of *K* indicate there are small differences in abilities. Consequently, a statistic with a relatively small *K* is a better measure of true ability.

We fit the above random effects model and estimate the quantities *K* and η for each of the eight hitting measures using data for all players in the 2003 baseball season with at least 100 at-bats. Table 1 displays the

Fitted Ability Distribution					
Statistic	***K***	**5th Percentile**	**50th Percentile**	**95th Percentile**	**Order**
SO rate	45	0.094	0.173	0.278	More Ability
IP HR rate	70	0.012	0.039	0.088	
BB rate	81	0.042	0.083	0.143	
BB rate (w/o Bonds)	85	0.040	0.079	0.136	
OBP	209	0.278	0.330	0.384	
OBP (w/o Bonds)	230	0.280	0.330	0.382	
IP AVG	408	0.288	0.326	0.365	More Luck
AVG	486	0.235	0.267	0.301	
IP 2+3 AVG	495	0.055	0.072	0.093	
IP S rate	530	0.186	0.215	0.245	

Table 1. Ability distributions for fitted random effects models for each of eight batting statistics for all MLB players in the 2003 season. [Note: We did two analyses for the walk data (BB)—one including Barry Bonds and one not including Bonds—as Bonds' walk rate is much larger than the remainder of the values and his data have a significant influence on the fit. Likewise for the on-base percentage (OBP) statistic.]

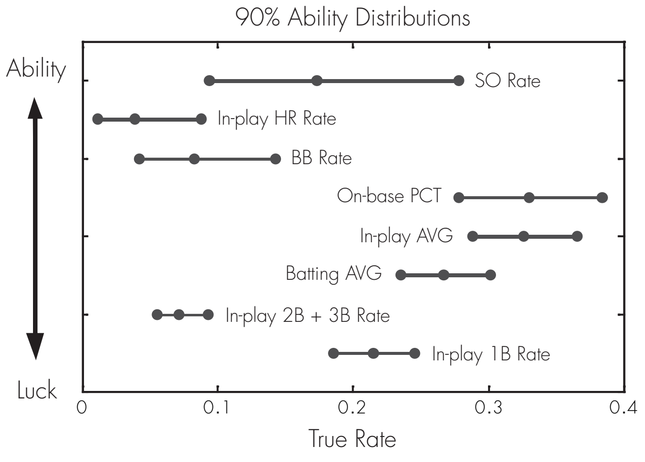


Figure 5. Graph of 5th, 50th, and 95th percentiles for fitted ability distributions for each batting statistic for all MLB players in the 2003 season

estimated value of *K* for each of the model fits. In addition, this table gives the 5th, 50th, and 95th percentiles of the fitted ability distribution. The 50th percentile (the median) is an estimate of the ability for a typical player and the 5th and 95th percentiles are useful for finding an interval where 90% of the player abilities fall. For the SO rate, we see the 5th and 95th percentiles are 0.094 and 0.278. Therefore, 90% of the players have 'true' strikeout rates between 9% and 28%. Figure 5 displays these ability distributions using error bars.

Table 1 and Figure 5 are useful in comparing the ability dimension of each of the eight hitting statistics. In fact, Table 1 and Figure 5 sort the statistics with respect to ability—statistics at the top are more indicative of batting skill and the statistics at the bottom are more reflective of chance variation.

At the top of Figure 5, we see there is a wide variation in the true strikeout rates for the players—90% of the players strike out between 9% and 28% of the time. Similarly, players seem to have varying abilities to hit a home run or draw a walk. In contrast, the other batting measures toward the bottom seem to be primarily chance-driven, which means that much of the variation in the players' values is due to basic binomial (coin-tossing) variation.

The most extreme measure of this type is the IP singles rate. Ninety percent of the true IP singles rates fall between .186 and .245, which is a rather short interval. This means it is relatively difficult for a batter to use his skill to aim his hits so they fall in for singles. Likewise, we note from Table 1 that the true IP doubles + triples rates and true batting averages have relatively short intervals. Most of the variation we see in these particular statistics across players is due to chance. So, it is hard to detect players' batting abilities by looking at these measures.

Concluding Remarks

The batting average is difficult to interpret because it confounds two hitting characteristics: the propensity of striking out and the ability to make a batted ball fall in for a hit. For this reason, it is difficult to understand what it means for a player to possess a batting average of, say, .320. How much does this say about the player's hitting ability?

We have that the batting average is a relatively poor measure of a batter's ability and that there are superior measures of hitting ability, such as a player's strikeout rate, in-play home run rate, and walk rate. Throughout one season, these measures are more ability-driven than the batting average.

To see a more extensive analysis of batting averages, read "A Batting Average: Does It Represent Ability or Luck?" at *http://bayes.bgsu.edu*.

References:

Albert, J. and Bennett (2003). *Curve Ball: Baseball, Statistics, and the Role of Chance in the Game.* Springer-Verlag.

Bickel, E. (2004). Why it's so hard to hit .400. *Baseball Research Journal*, 32:15-21. ■

Every Scatterplot Tells a Story:
a Look at the Performance of Long-distance Runners

Matthew Strand

While the typical person is probably familiar with long-distance running and perhaps even has participated in a local 5- or 10-kilometer road race to support a charity, it takes a special breed to actually become a competitive long-distance runner. Although the race objective is quite simple—finish as fast as possible—the preparation that goes into striving for this goal is not simple.

There are both physical and mental components to training. Runners spend many months working out to prepare for key races, and mental toughness allows them to fight through physical discomfort in hard training runs or races. However, another key to good performance is pace strategy. That is, knowing how fast to start the race in order to maintain pace and have a strong finish.

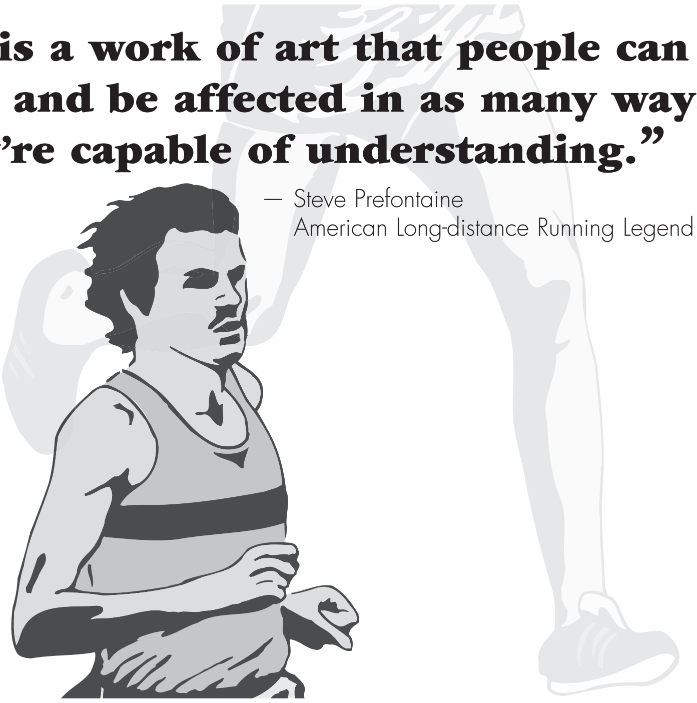
Runners can gain psychological advantage by starting hard and running with other lead runners, but going out too hard can leave them burned out in the second half of the race. On the other hand, starting too slowly may leave a runner demoralized or spending too much energy trying to pass other runners during a race. Reviewing their *split times*—the time it takes a runner to complete a certain portion of the course—is critical for runners to know in order to plan for future races.

With advances in technology, many races now clock times at various points in the race for each runner in addition to the overall finish time. A common way this is achieved is by having each runner place an electronic chip on his or her shoe. Censors under mats at designated locations then identify the runners and record their times as they pass each location. Using these times, runners can see not only how they progressed throughout the race, but they can compare their pace performance with others.

Two races with split time data collection are the NCAA Division I Cross Country Championships and the New York City Marathon. Both races are run in November, when temperatures are bearable, if not pleasant, for long-distance running. The NCAA meet fields some of the top collegiate runners from across the country. In fact, many non-U.S. citizens also compete in these races, as they have gained scholarships to attend U.S. colleges. Some of the runners who participate in this race even go on to compete at the world-class level. The NYC Marathon has a broader range of runners, from the elite to those who simply hope to finish the race.

"A race is a work of art that people can look at and be affected in as many ways as they're capable of understanding."

— Steve Prefontaine
American Long-distance Running Legend



NCAA Cross Country Championship

In a recent championship, 251 men finished a 10-kilometer race and 254 women finished a 6-kilometer race. Both men's and women's races included 31 teams in addition to other qualifying individuals. Times were recorded at the 5k, 8k, and 10k points for the men

Matthew Strand (strandm@njc.org) is a biostatistician working at the National Jewish Medical and Research Center. He also teaches at the University of Colorado Health Sciences Center. He is an avid long-distance runner and enjoys interpreting data.

and at the 3k and 6k points for the women. Figure 1 is a scatterplot of each runner's time in the second half versus the first half of the race for both men and women. The graphs show what one might expect: runners who ran faster first-half splits tended also to run faster second-half splits.

Using Two-dimensional Scatterplots To Plot Three Related Variables

Although a two-dimensional scatterplot is used to plot two variables, any other variable that is a linear combination of those two also will be on that plot. Thus, a two-dimensional scatterplot can be used to graph three or more variables that span two dimensions. One of the keys to making a good graph is to identify the additional variables of interest without unnecessarily complicating the graph. For Figure 1, time in the first half (x) and time in the second half (y) can be added to obtain the overall time ($x + y$), which is represented by a solid line that runs 45° counterclockwise from x, as the two primary axes are plotted with identical scales. The dashed lines are used to indicate the overall time on the third axis.

Another variable of interest for the race data is the difference between first- and second-half splits ($y - x$), which is one measure of pace performance. This axis also lies on the plot of y versus x and runs 45° counterclockwise from y on Figure 1. The overall time axis, itself, is perpendicular to the split difference axis and is the line $y - x = 0$, which indicates equal or 'even' split times. It is apparent from the graphs that the fastest runners had split times close to even. Some of the slowest finishers had some of the most uneven splits, as their points lie the furthest above the even-split line.

Figure 2 shows a graph that brings the split difference ($y - x$) into more prominent view by using it as the vertical axis. Overall time is placed on the horizontal axis. A solid line is used to indicate $y - x = 0$, the line of even splits, making it easy to see the runners who slowed down and those who sped up in the second half of the race and therefore ran 'negative splits.'

In this scatterplot, the first 5k time (x) axis runs from upper left to lower right. Dashed lines run perpendicular to x and are included to show times of equal value on this axis. Consequently, the dashed lines show groups of runners that came through the halfway point together. Any dashed line that is parallel to those already displayed also could be added to the graph to indicate constant halfway times. For example, a runner interested in seeing his or her relative location at the halfway point could draw a parallel dashed line over his or her point on the graph. Any other points on that line would represent the other runners around him or her at the halfway point.

An advantage Figure 2 has over Figure 1 is that although the same data are being plotted, there is greater separation of points in Figure 2; hence, the location of individual points relative to others is easier to see. This is especially useful for runners who want to compare their performance with the performance of others. The greater separation of points stems from the fact that split difference and overall time are more weakly correlated (r = 0.7 and 0.6 for the men and women, respectively) than first-half time and second-half time (r = 0.9 for both men and women).

Interpreting the Scatterplots

The similarity of trends in Figure 2 between men and women is quite remarkable. The majority of times are above the even-pace line, showing that most run the first half faster than the second. For the men, only 6% (15 out of 250) of the runners ran the second half faster. Similarly for the women, only 7% (17 out of 254) ran faster in the second half. Although the differences in split times and overall times are correlated, this does not imply a causal relationship, as it is likely that those with greater ability set the tone for the race while many

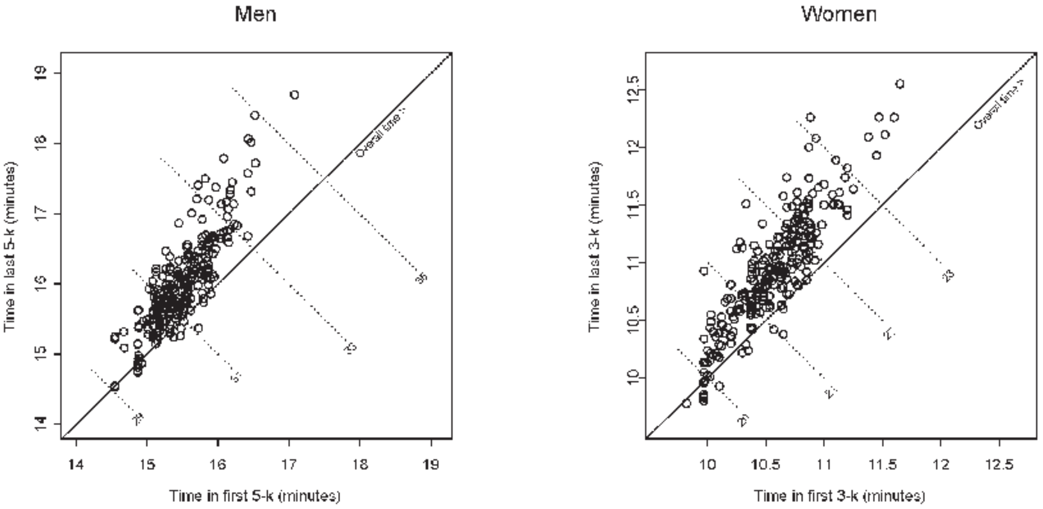


Figure 1. Time in the second half of the race versus time in the first half of the race for men and women in a recent NCAA Division I Cross Country Championship. Each circle represents a runner. The angled solid line represents the axis for overall time. The dashed lines indicate constant values on this axis in minutes. Points above the angled solid line are runners who ran faster in the first half of the race compared to the second half.

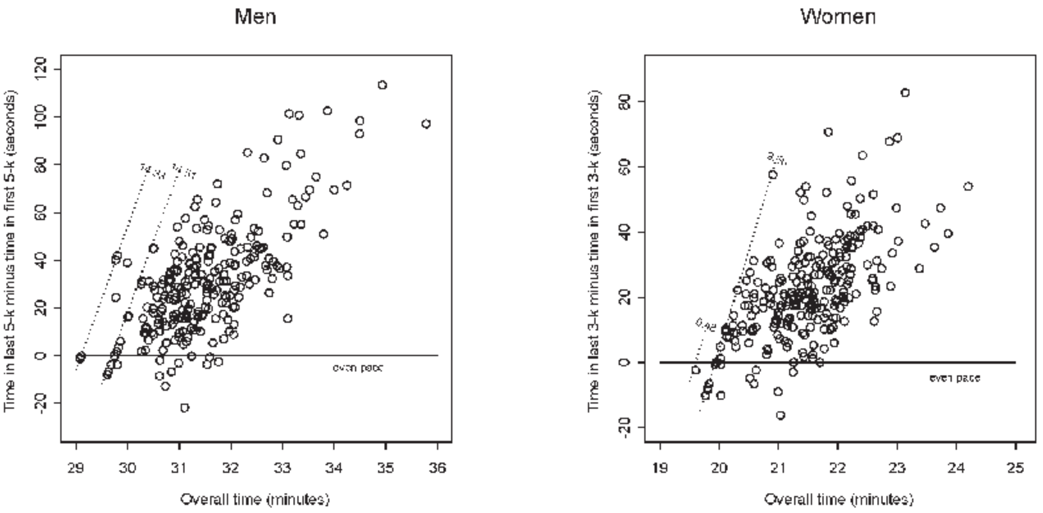


Figure 2. Split-difference times (second-half time minus first-half time) versus overall time for men and women in a recent NCAA Division I Cross Country Championship. Each circle represents a runner. The solid line indicates even split times. The dashed lines indicate equal times at the halfway point labeled in minutes:seconds. Points above the horizontal even pace line are runners who ran faster in the first half of the race compared to the second half.

of the others were just trying to keep up as long as they could. Also, some runners with a large disparity in their split times (particularly for the men) still finished in the top 25.

The progression of the race is arguably easier to see in Figure 2 than in Figure 1. Runners tend to run in groups called 'packs.' The dashed lines indicate that some runners ran in packs for part of the race, but then dispersed somewhat in the second half. In Figure 2, for the men, the first dashed line (14:33) shows that four runners were together at the 5k point. Two of these runners maintained their pace to the end and finished together. The other two slowed down.

Another major pack with 13 runners (along line 14:51) was about 18 seconds behind the lead pack at the halfway point. The two runners in the lead pack who slowed down apparently were caught by some of the runners in the second pack who finished with faster overall times. The two runners between the first and second packs at the halfway point also were caught by some of the runners in the second pack.

For the women, it is clear that one runner was alone in the lead at the 3k point, and she was able to maintain her pace and win the race. A major pack followed (9:58) 10 seconds behind the leader. Some of the runners in this pack were 'pulled up' by the lead runner and actually sped up in the second half in an attempt to catch her.

Note that all of these interpretations were made simply by looking at the scatterplots, as the author did not attend the race. Even without attending the race, we can see what happened. The scatterplots tell the story! Now, let's look at another race.

NYC Marathon

In the New York City Marathon races, there are typically more than 30,000 finishers each year. In two recent consecutive years, the times among the top runners were comparable, as 250th place was 2:48:25 the first year and 2:48:58 the next year. Table 1 shows the average cumulative split times for the top 250 finishers in the two years and, for comparison, what the splits would be for an even pace. As with the cross country times, these data exhibit faster start times, with a level of consistency between years that seems remarkable. Some slowing is noticeable in the last quarter of the race.

Cumulative Split Distance	First-year Average Cumulative Split Time	Second-year Average Cumulative Split Time	Even Pace Time to Finish at 2:38
10-k	0:35:52 (-1:34)	0:35:53 (-1:33)	0:37:26
21.1-k	1:15:53 (-3:07)	1:15:41 (-3:19)	1:19:00
32.2-k	1:58:02 (-2:32)	1:57:36 (-2:58)	2:00:34
42.2-k	2:38:14 (+0:14)	2:37:45 (-0:15)	2:38:00

Table 1. Average cumulative split times for the top 250 finishers of the NYC Marathon in two consecutive years. Differences relative to the even pace time shown in the far right column are given in parentheses. The even pace times are based on a 2-hour and 38-minute finish, which was the average finish time of the top 250 runners over the two years. A negative time compared to the even pace time means the runners were running faster than the even pace time at that split distance. So, a negative means running 'ahead' of even pace.

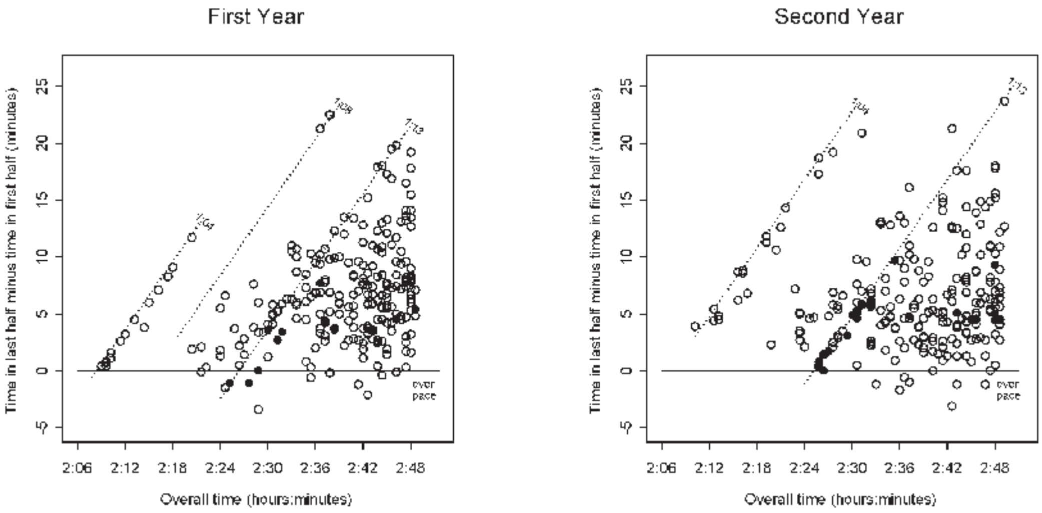


Figure 3. Split difference times (second half time minus first half time) versus overall time for the top 250 finishers in two consecutive NYC Marathon races. Each circle represents a runner: Open = Men and Closed = Women. The solid line indicates even split times. The dashed lines indicate equal times at the halfway point labeled in hours:minutes. Points above the horizontal even pace line are runners who ran faster in the first half of the race compared to the second half.

Figure 3 shows split difference versus overall time for the top 250 finishers for each of the two races. The most striking feature of the two graphs is the clear separation of points on the left-hand side. These 15 or so runners primarily comprise the elite men, who are competing not only to win, but for prize money. Recently, the total purse climbed to more than $500,000, with prize money guaranteed to the top 10 men and women. The top five finishers in the first year, shown in Figure 3, had fairly even half splits, while in the second year, even the winner slowed down substantially in the second half.

In the first year, there was a tight lead pack at the halfway point (1:04). These runners dispersed in the second half, although even the runner in this group who slowed down the most (12 minutes slower in the second half) was not caught by those behind the elite pack. The two runners who followed at roughly 1 hour and 8 minutes at the half slowed down tremendously in the second half and were passed by 50 or 60 runners. In the second year, the lead pack at the halfway point was not as tight as in the previous year, as seen by the greater spread along the dashed line at 1:04, and many of the following runners passed more of the lead runners before the finish.

The elite runners are strongly motivated by placing for prize money, which may partially explain the dispersion among the top runners in the second half of the race in both years. Specifically, runners may slow down if they are not likely to be passed, or they fall out of contention for prize money and lose their incentive for a strong finish.

In Figure 3, women are shown as closed circles. Contrary to the men, there was a tighter lead pack for the women the second year than in the first. The leader for the women can be seen along the 1:13 dashed line. The mix of open and closed circles along the 1:13 dashed line indicates that many men ran with the elite women for at least part of the race.

For Runners

Graphs of split times can be informative to runners in helping to plan their pace strategies. There is an old adage, “even pace wins the race.” Figures 2 and 3 support that assertion. However, other factors may

affect whether a runner needs to start a little faster or slower. Such factors may include competitors' tactics, weather conditions, course characteristics, or individual physiology. By examining split and finish times in relation to those of other runners, a runner may have a better idea about whether his or her pacing was reasonable in the last race or whether he or she should try a different approach for the next race.

For teams competing in cross country, coaches and their runners also can use graphs to plan for future races, where runners on the team are identified with a unique symbol. Coaches often encourage teammates to run together. The graph can show not only how closely runners on a team finished, but their split consistency as a group and how they progressed through the race together.

For Statisticians

Graphs are powerful tools that can be used to summarize a large amount of data into a picture. Good graphs can be quite simple, where one or two ideas can be conveyed quickly, or quite complex, where many ideas can be extracted with careful study. The graphs presented here have elements of both.

Although scatterplots are typically two-dimensional, a third variable sometimes can be identified on the same plot, as we have done in this analysis. Such graphs should be distinguished from three-dimensional visual plots, which are plotted in two dimensions with the illusion that they are in three-dimensional space. These graphs do have an intuitive feel, but identifying where individual points actually lie often can be difficult.

Researchers should not pigeonhole graphs for certain types of data. For example, one does not always have to use a line graph for repeated measures data or data collected over time, such as data presented in this article. Many graphs actually do tell stories, and one can actually relive events to some degree by viewing a well-constructed graph.

Additional Reading:

Here are some references about graphs and running for additional reading and further study:

Gambaccini, P. and Miers, C. 1994. *The New York City Marathon: Twenty-five Years 1969–1994*. Rizzoli. ISBN 0-847-818152

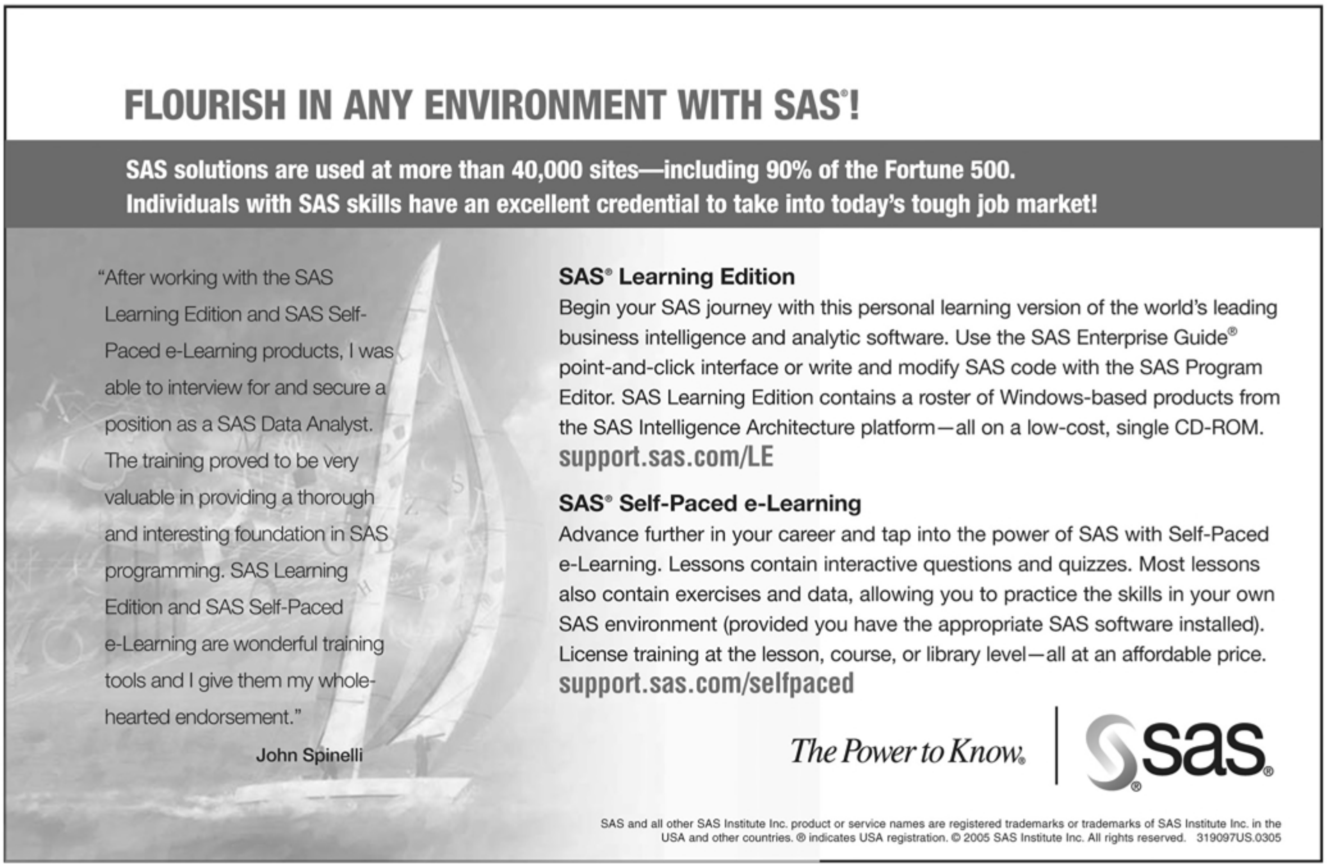
Jordan, Tom. 1997. *Pre: the Story of America's Greatest Running Legend, Steve Prefontaine*. 2nd edition, Rodale Press, Inc. ISBN 0-875-964575

Lear, Chris. 2003. *Running with The Buffaloes: a Season Inside with Mark Wetmore, Adam Goucher, and the University of Colorado Men's Cross-country Team*. The Lyons Press. ISBN 1-585-748048

Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. 2nd edition, Graphics Press. ISBN 0-961-392142

Acknowledgement:

The data presented in this article were used with the permission of the National Collegiate Athletic Association (NCAA) and the New York Road Runners. ■



An NFL Cookbook:
Quantitative Recipes for Winning

Patrick Bartshe

Winning is everything in the National Football League (NFL), and most coaches would agree there is not one variable that accounts for success. It is the combination of variables that comprise a winning team. For example, when baking a cake from scratch, you need more ingredients than just flour and water. A cup of brown sugar, two eggs, or a teaspoon of vanilla also may be needed to make a respectable cake. Likewise, a good football team needs a heap of team chemistry, a dash of good running game, two cups of solid defense, a teaspoon of special teams, and maybe a large scoop of Payton Manning.

NFL Analysis

The NFL has two conferences: the National Football Conference (NFC) and the American Football Conference (AFC). The NFC and AFC champions meet each year in the Super Bowl to determine the NFL Champion. Data for the NFL are available on the NFL web site at *www.nfl.com*.

Offense	Defense	Special Teams
Passing Yards Gained	Passing Yards Allowed	Field Goals Scored
Rushing Yards Gained	Rushing Yards Allowed	Kick-off Return Average
Yards Penalized	Yards Penalized	Punt Coverage Average
Interceptions Thrown	Interceptions Caught	
Fumbles Lost	Fumbles Recovered	
	Sacks Made	

Table 1. Predictor variables evaluated using regression analysis

Patrick Bartshe (patrick.bartshe@asu.edu) is a doctoral student at Arizona State University who is majoring in educational psychology. His interests include real estate, sports, and spending time with family. After graduation, he hopes to find his way into academia.

We used multiple regression analysis to examine data from three NFL seasons spanning the years 2002–2004. We built regression models for the NFC and AFC separately in order to describe the aspects of the game emphasized in each conference and to determine the recipe for success as practiced by the teams in each conference. The response variable was the number of wins each team had in the regular season. The predictor variables were selected to describe the performance of the offense, defense, and special teams units of each football team. Table 1 lists the predictor variables evaluated.

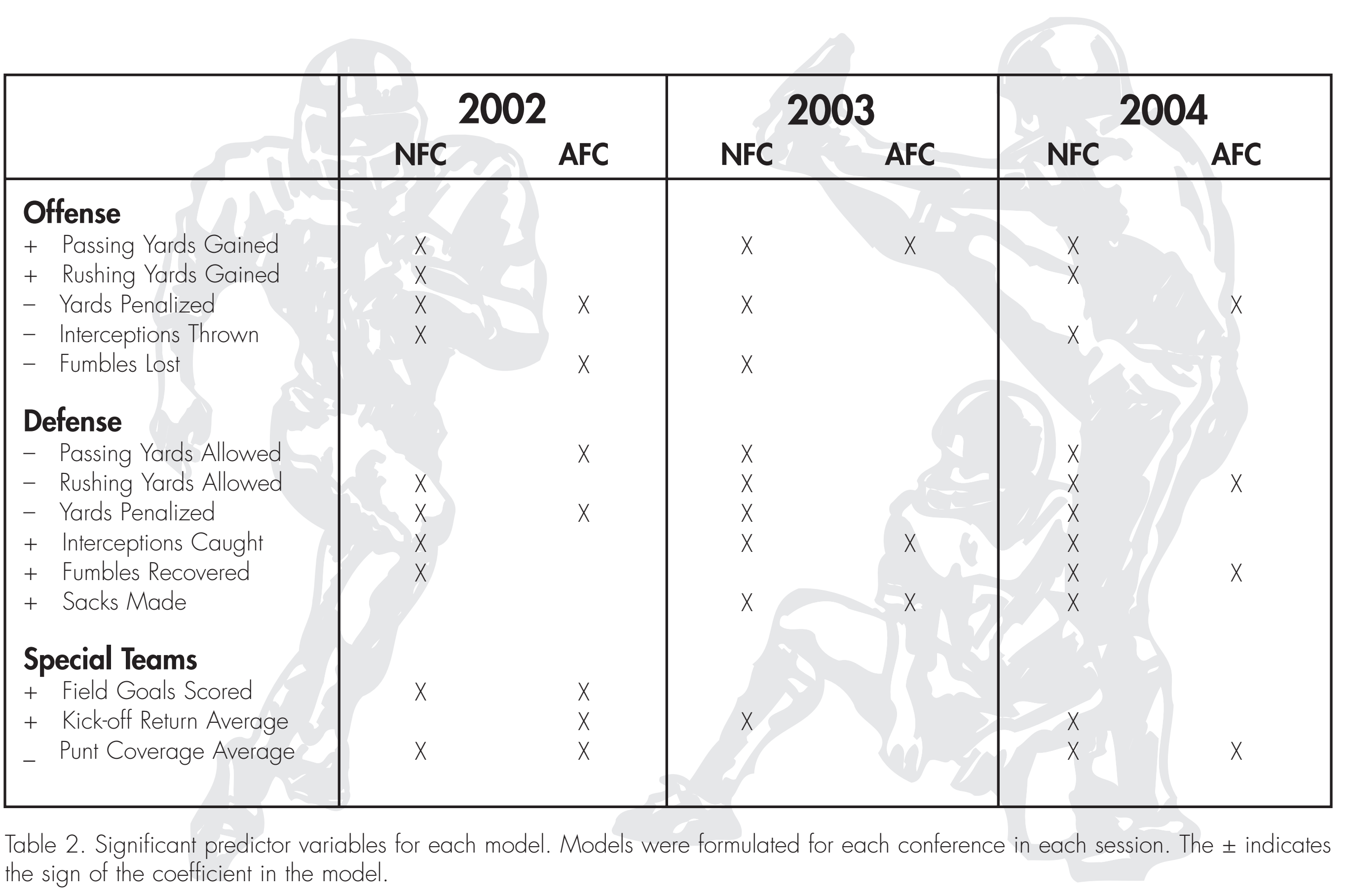
The selection criteria for including a predictor variable in the regression models were the F-test, adjusted R^2, *Mallows* C_p, and the concept of parsimony—simple models are preferable to complex models. Table 2 shows the predictor variables judged to be significant for each model. These are the variables essential to winning football games during the 2002 through 2004 seasons. We can think of Table 2 as an "NFL Cookbook," with the recipes for success during the last three seasons.

The 2002 NFL Season

Based on the regression models, the 'recipe for success' in the NFC during the 2002 season might be described as follows:

Formulate a balanced offensive attack of both rushing and passing plays with a low-interception quarterback. Make sure the offensive line focuses on the snap-count and is careful to avoid holding and illegal blocks. Add a defense that can shut-down the run, create turnovers, and refrain from penalties. Mix in a good field goal kicker and an effective punt coverage scheme. The result will look a lot like the Super Bowl Champion Tampa Bay Buccaneers.

The NFC model during the 2002 season was significant, as shown by an F-statistic of $F_{10,\,5}$ = 105.52 (p < .001) and adjusted R^2 = .9859.



	2002		2003		2004	
	NFC	AFC	NFC	AFC	NFC	AFC
Offense						
+ Passing Yards Gained	X		X	X	X	
+ Rushing Yards Gained	X				X	
– Yards Penalized	X	X	X			X
– Interceptions Thrown	X				X	
– Fumbles Lost		X	X			
Defense						
– Passing Yards Allowed		X	X		X	
– Rushing Yards Allowed	X		X		X	X
– Yards Penalized	X	X	X		X	
+ Interceptions Caught	X		X	X	X	
+ Fumbles Recovered	X				X	X
+ Sacks Made			X	X	X	
Special Teams						
+ Field Goals Scored	X	X				
+ Kick-off Return Average		X	X		X	
_ Punt Coverage Average	X	X			X	X

Table 2. Significant predictor variables for each model. Models were formulated for each conference in each session. The ± indicates the sign of the coefficient in the model.

To win games according to the model for the AFC teams, the offense needed to hold on to the ball and avoid penalties, the defense needed to be effective against the pass and also refrain from committing penalties, and the special teams needed to play a solid game plan with a steady field goal kicker. For the AFC model, the F-statistic was $F_{7,\,8} = 26.98$ ($p < .001$) and adjusted $R^2 = .9238$.

The 2003 NFL Season

The regression model shows the successful NFC team had a coordinated offensive line that could give the quarterback time to pass to sure-handed receivers and that the running backs held on and protected the ball. Also, the defense was disciplined (did not commit penalties) and ready for pass or run plays. They apparently applied pressure on the quarterback, catching him in the backfield or causing him to make mistakes, such as throwing interceptions. Finally, the special teams were characterized by a solid kick-off return man who had effective blocking in front of him. The statistics for this model were $F_{9,\,6} = 36.30$ ($p < .001$) and adjusted $R^2 = .9549$.

In the AFC, the offense had a good pass-protection offensive line, sure-handed receivers, and an effective quarterback; the defense emphasized sacking the quarterback and making interceptions. This was the recipe used by the Super Bowl Champion New England Patriots. The model statistics were $F_{3,\,12} = 13.82$ ($p < .001$) and adjusted $R^2 = .7195$.

The 2004 NFL Season

For success in the NFC in 2004, the model says teams had an offense with a balanced attack of running and passing plays, orchestrated by a patient quarterback; the defense was able to do everything right and was prepared for anything. Also, special teams emphasized the running aspects of their game. The statistics for this model were $F_{11,\,4} = 29.16$ ($p = .003$) and adjusted $R^2 = .9538$.

The AFC model suggests the offensive line on effective teams was mindful of the snap-count and administered clean blocks, the defense shut down the run and tried to strip the ball, and the special teams had a good punter who could kick the ball high and far enough that the coverage team could minimize the return. The New England Patriots repeated as Super Bowl champions with this recipe. This model's statistics were $F_{4,\,11} = 30.43$ ($p < .001$) and adjusted $R^2 = .8870$.

Conclusions

This investigation was an observational study and gives a descriptive analysis of the data, rather than a prescriptive evaluation. The results help us understand what happened, rather than what could have happened. Nonetheless, the models give a perspective on what was required for success in each conference of the NFL during the last three seasons.

Although the interpretations presented here are highly subjective, the goal of this analysis was to get involved with the game quantitatively. Consequently, one major finding was the great fun it was to investigate this kind of data and to look for possible meaning in the results. It would be fun to test other variables and combinations of variables in similar models for football or other sports. Try it and see what you find and how much fun you have in the process. ■

STATS **Puzzler**

Schuyler W. Huck

Bo, Jo, and Mo

The three Smith brothers—Bo, Jo, and Mo—were highly talented baseball players. Each made it to the Major Leagues, although they played on different teams. These brothers were highly competitive, both on and off the field, and each enjoyed beating out the other two when it came to statistical measures of their baseball accomplishments. Though many measures exist for doing this, Bo, Jo, and Mo agreed that one's batting average is the *best* indicator of player performance.

Imagine the excitement in the Smith family when it turned out each of the brothers had a batting average of exactly .300 going into the final game of the season. How they did in their final games would determine who would claim 'bragging rights' during the long off-season.

During their season-ending games, they each had five official trips to the plate. However, that was the only similarity among them because Bo went three-for-five, Jo went four-for-five, and Mo went five-for-five. In baseball statistics vocabulary, when a player goes "X-for-Y" it means the player had X hits in Y times at bat.

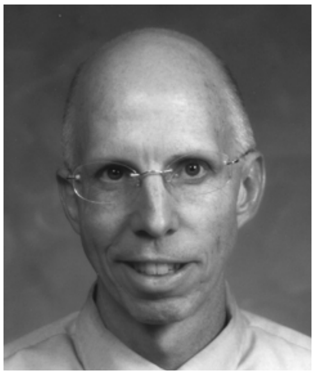
Which one of these three players do you think ended the season with the highest batting average? Do you have enough information to answer the question? If not, what information do you need?

After you have your answers, turn to Page 19 to see the *STATS* Puzzler's answer.

Schuyler W. Huck (shuck@utk.edu) teaches applied statistics at the University of Tennessee. He is the author of Reading Statistics and Research*, a book that explains how to read, understand, and critically evaluate statistical information. His books and articles focus on statistical education, particularly the use of puzzles for increasing interest in and knowledge of statistical principles.*

Is Lance Armstrong the Greatest Cyclist Ever?

Data Investigations with the Tour de France



Peter Flanagan-Hyde

In July 2005, Lance Armstrong won the Tour de France for the seventh consecutive year, an unprecedented victory in the grueling bicycle race that dates back to 1903. No other rider has won more than five tours, and no rider has maintained the consistency and durability Armstrong has shown. Indeed, no other rider has ridden a faster average pace than Armstrong did to earn his last victory, which has spurred some to say he is the greatest endurance cyclist ever.

Peter Flanagan-Hyde (pflanaga@pcds.org) has been a math teacher for 27 years, the most recent 15 in Phoenix, Arizona. With a BA from Williams College and an MA from Teachers College, Columbia University, he has pursued a variety of professional interests, including geometry, calculus, physics, and the use of technology in education. Flanagan-Hyde has taught AP Statistics since its inception in the 1996–1997 school year.

Add to this Armstrong's personal story of winning the World Championship of Cycling in 1993 and then being stricken with cancer. He managed to climb back—not just to participation, but to greatness. The compelling tale of Armstrong will long endure, but do the data show his accomplishments to be special when compared to past champions?

First, let's point out that the question "Who is the greatest ever?" is ubiquitous in sports. Who is the greatest home run hitter of all time—Henry Aaron, Babe Ruth, or Barry Bonds? Who is the greatest basketball player ever? Is it Michael Jordan, Shaquille O'Neal, or Wilt Chamberlain? How about the top golfer—Jack Nicklaus or Tiger Woods? Each of these sports figures has some unique accomplishment that can be touted as so extraordinary that it could earn them the title of greatest ever. Questions such as these often reveal strongly held opinions among sports fans. Statistical graphics can be helpful when exploring the data and looking for insight during these debates.

Let's take a look at Armstrong's accomplishments throughout his career in cycling. The format of the Tour de France has evolved somewhat over the course of its existence, making comparisons to the early years especially difficult. As grueling as the current race is today, the first tours featured even longer stages of more than 400 km, as opposed to the current tour with stages averaging fewer than 200 km. We therefore will confine our attention to what might be called the modern tour, beginning in 1953, when the prize money for placing second and third was increased, making the tour a more competitive event.

In this exploratory process, we can examine distributions, relationships, and special cases to find meaningful relationships in a large set of data.

Examining Distributions

The point of a bicycle race is to cover a specified distance in as little time as possible, so perhaps Armstrong's winning times will reveal some distinguishing characteristics. Figure 1 is a display of the winners' times with the shading indicating Armstrong.

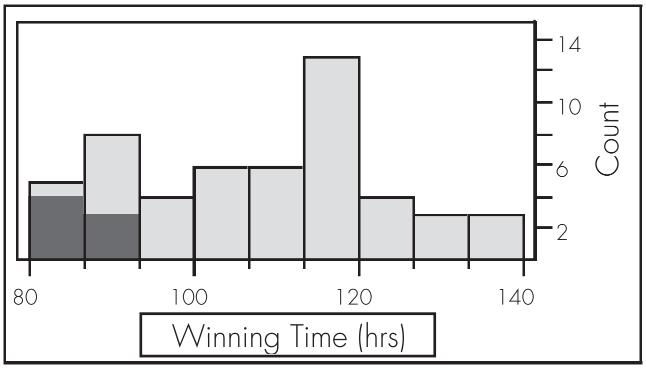


Figure 1. Distribution of winning times on the Tour de France

Quite impressive! His seven winning times are all among the 10 fastest times ever recorded. How about the average pace for each of these races? This looks equally impressive in the second histogram (Figure 2), with Lance taking many of the top spots for winning pace. In fact, he posted the top five fastest times.

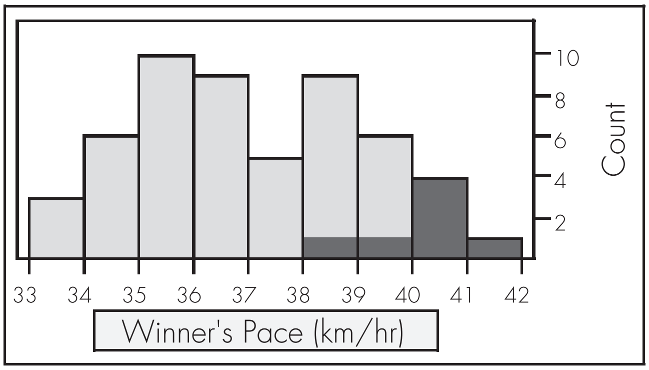


Figure 2. Distribution of winner's pace on the Tour de France

Figure 3 shows the winning margin for each of the tours, but in this case Armstrong's performances do not stand out.

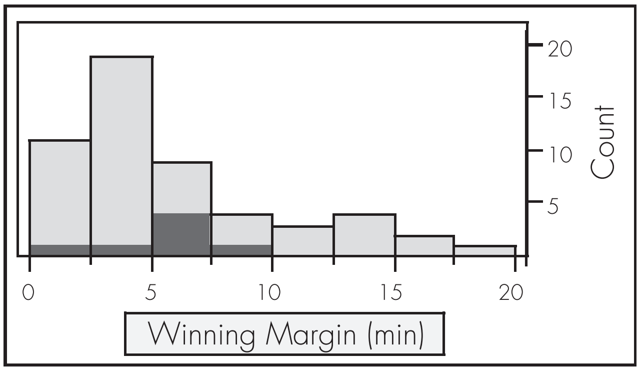


Figure 3. Distribution of winning margins on the Tour de France

This bit of dissonance provokes some puzzlement. How can Armstrong ride faster than nearly any other winner, yet end up with relatively small margins of victory? Does this tell us something about his competition? How about the length of the race? Let us look at the distribution (Figure 4) of the length of each of the complete Tour de France.

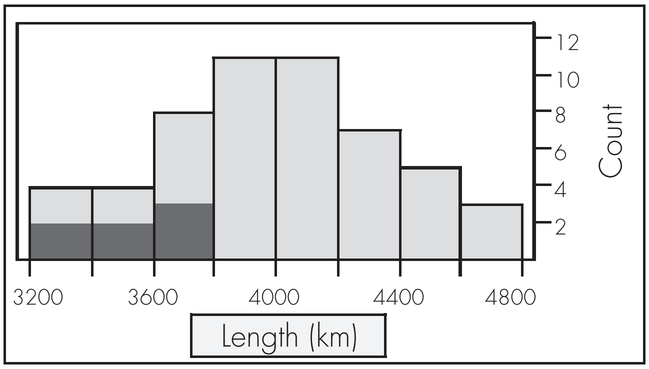


Figure 4. Distribution of race length on the Tour de France

In this graph, something that isn't necessarily common knowledge becomes abundantly clear: Armstrong has ridden in some of the shortest modern tours. Was he just fortunate to have been riding in years that had shorter courses, or is there more to the story? Let's see!

Examining Relationships

Is there a relationship between race length and the winner's pace? There is certainly some logic in suggesting that shorter races should have faster average speeds—less fatigue perhaps. Figure 5 is a scatterplot with a least-squares regression line fit to the data. The points that correspond to Armstrong are solid dots.

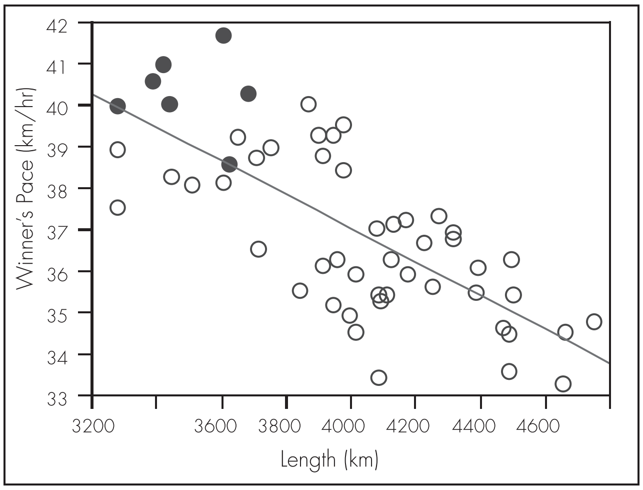


Figure 5. Winner's pace versus race length on the Tour de France

It is easy to see a clear, linear relationship between the length of the race and the winner's pace. The points for Armstrong are all either above the line or on the line in most cases. So, Armstrong was faster than would be expected from the length of the race. Perhaps he was just good (a distinct possibility), but there might be other reasons why he rode faster than expected. Here is where the real value of our exploration begins.

What other factors might influence average pace? The makeup of the course (more mountains one year than the next, or more time trials) or the impact of the technology of bicycles (lighter, more aerodynamic now) might be factors. No information was found about the makeup of the course, but as a surrogate for the factor of technology, the year of the tour can be used. Figures 6 and 7 show the winner's pace and the length of the race by year.

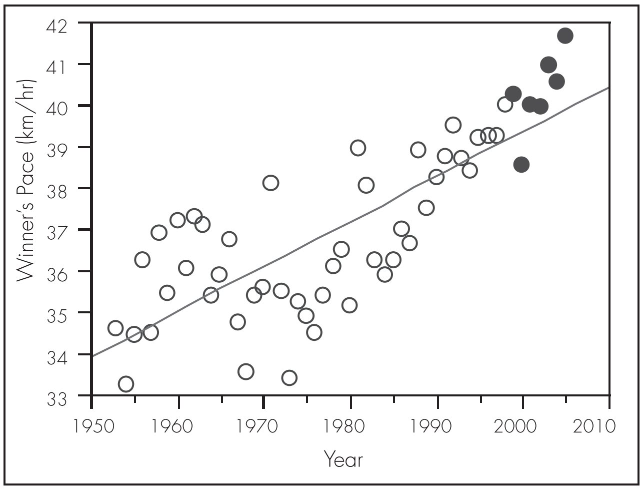


Figure 6. Winner's pace by year on the Tour de France

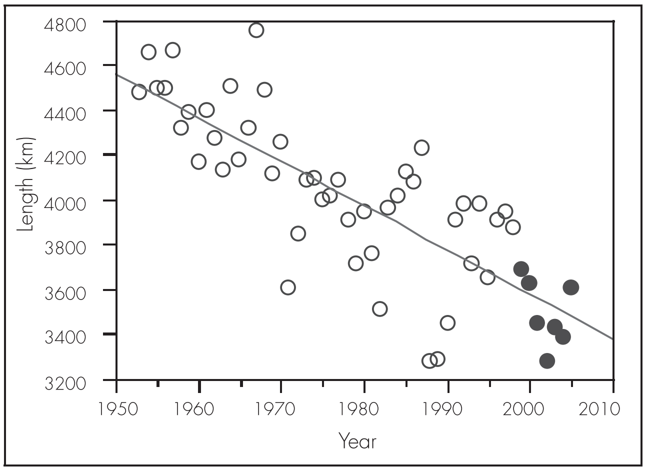


Figure 7. Race length by year on the Tour de France

In these graphs, there are clear associations between the year and both the winner's pace and the length of the race. Since the year associates strongly with both length and pace, it is impossible to say which of these factors 'explains' why some tours are ridden faster than others. Is it that some are longer, that there are better bicycles today, or a combination of these and other factors? We can't say!

Now let us turn to a group of riders widely regarded as the best to have ridden on the tour: the riders who have won the most races. These are special cases.

Examining Special Cases

Figure 8 is a boxplot that shows the number of wins for each of the riders on the tour who have finished in the top three places (i.e., on the podium).

The riders who have won the Tour de France five times are Jacques Anquetil, Eddy Merckx, Bernard Hinault, and Miguel Indurain. These riders are marked with an "X" as outliers in the boxplot. The three-time winners also are shown as outliers with the open circles. Who is that with the solid circle? Armstrong with seven wins.

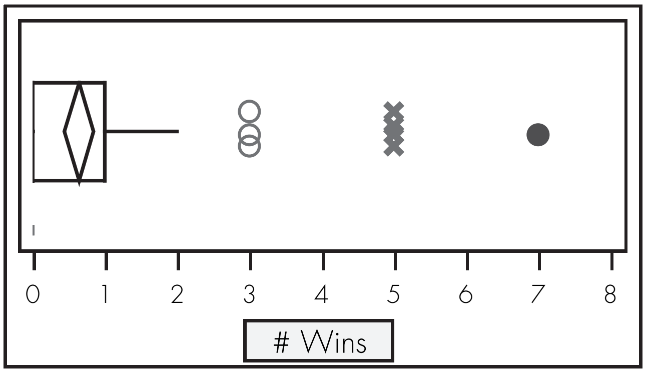


Figure 8. Boxplot of wins per rider on the Tour de France

Many cycling enthusiasts regard Merckx as the most dominant rider of any era. He has a reputation for going all out, all the time, whereas both Armstrong and Indurain are known as more strategic riders who carefully calculate just what it takes to win. To compare these exceptional riders, Figure 9 presents the winning margins of each of their victories on the tour in parallel dotplots. We can see that the winning margins for Armstrong and Indurain are generally lower than for Merckx.

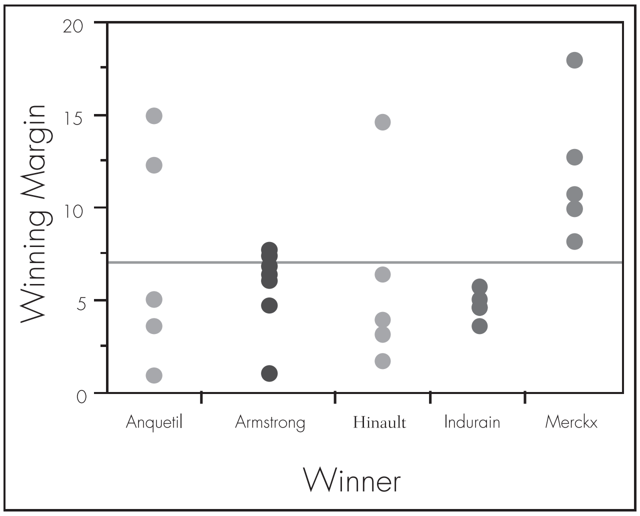


Figure 9. Winning margins of riders with five or more wins on the Tour de France

Are there other riders who should be considered as the greatest based on a combination of wins plus second- and third-place finishes? Figure 10 shows a distribution of the number of times riders have appeared on the podium. In this case, Armstrong's success is not the most extreme. That honor belongs to Raymond Poulidor, who had three second-place finishes and five thirds. Armstrong's seven appearances on the podium are matched by several other riders, including Hinault.

The last graph (Figure 11) shows a way to give more weight to victories by assigning points to each type of finish. If we assign five points for a first-place victory,

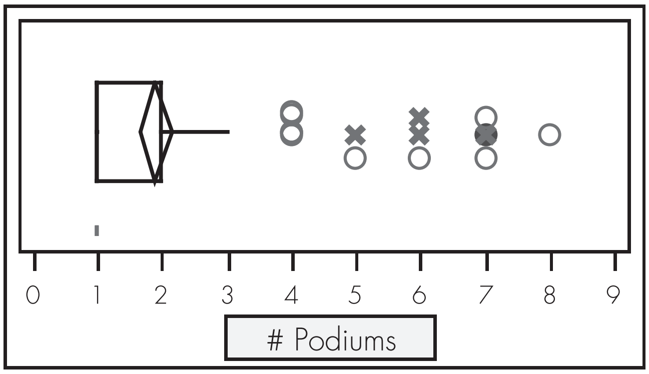


Figure 10. Distribution of appearances on the podium per rider on the Tour de France

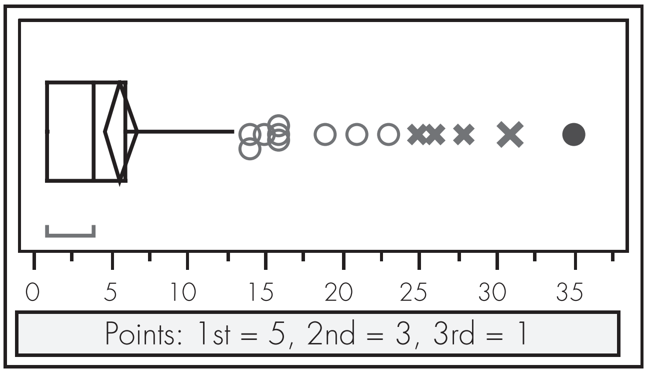


Figure 11. Distribution of cumulative "points" on the Tour de France

three points for second-place, and one point for finishing third, Armstrong's performance stands out distinctly from the rest of the riders.

So, is he the greatest cyclist ever to have ridden in the Tour de France? It can be argued, on the strength of the graph in Figure 11, that he is certainly the most successful rider to ever have competed. But the best cyclist ever? That depends on your measure of 'best,' and it is a hard question to answer. However, getting at hard questions through graphical data exploration can be fun and engaging. So do some exploration on data from your favorite sport and see what you find.

Data Sources

Tour de France Web Site: *http://www.letour.fr/2005/TDF/LIVE/us/2100/index.html*

Cycling News Web Site: *http://www.cyclingnews.com/road/2005/tour05/05index.php*

Torelli Web Site: *http://www.torelli.com/raceinfo/tdf/tdfindex.shtml* ■

STATS Puzzler's Answers

If you are like most people, you might have thought Mo (who went five-for-five in his final game) had the best end-of-season batting average, as he did the best in the last game. Surprisingly, Mo came in third place! Actually, it was Bo who won the family's batting-average bragging rights.

To understand why Bo ended up with the highest batting average, we need to know the specifics of their batting performance during the entire season. Table 1 provides the necessary data.

If you use these data to compute each player's full-season batting average (rounded to four decimal places), you get .3061 for Bo, .3059 for Jo, and .3058 for Mo. The fact that the order of these averages is exactly opposite the order of the averages in the final game is due to the different amounts of data used to compute the brothers' batting averages going into the final game.

The moral of this little puzzle is to be careful when you try to 'average' averages. It is important to compute a weighted average of what you wish to combine. For example, if you want to average Bo's performance in his last game with his performance in the earlier games, you need to compute [240 (.300) + 5 (.600)] / (240 + 5) = .3061. The result is a weighted average of the two numbers, .300 and .600, and it indicates how Bo did over the entire season. It would be incorrect to simply "average the averages" by calculating (.300 + .600) / 2 =.450.

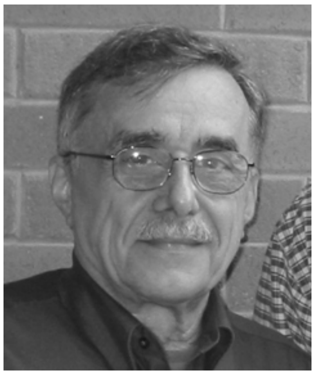
Before the Last Game			
Player	**Hits**	**Official At-Bats**	**Batting Average**
Bo	72	240	.300
Jo	126	420	.300
Mo	180	600	.300
During the Last Game			
Player	**Hits**	**Official At-Bats**	**Batting Average**
Bo	3	5	.600
Jo	4	5	.800
Mo	5	5	1.000

Table 1. Batting performance data for Bo, Jo, and Mo

Weighted averages are needed when percentages or proportions, each with a different 'basis' (denominator), are combined. In our example, the basis is 'at bats.' Failure to compute a weighted average can lead to silliness, such as thinking Bo batted .450 for the season.

R U SIMULATING?

Polls on Steroids



Bruce Trumbo

Suppose we want to conduct a poll to find out what people think about making laws against the use of performance-enhancing drugs in professional sports. First, we have to settle which 'people' we have in mind—our population of interest. (Do we want to know the opinions of everyone over a certain age? Of all registered voters? Just sports fans?) Then, we have to frame a question so the people we ask will know what we mean by 'laws' (What penalties?) and 'drugs' (Only illegal ones? All steroids?).

To be specific, let's use the question, "Do you think there should be a federal law imposing a heavy fine and a lifetime ban from the sport on any professional athlete who uses steroids?" Also, let's suppose our population of interest is all voters and that we will interview subjects selected at random from this population.

How Big a Sample Should We Take?

In this situation, how many people do we need to interview in order to get a sufficiently reliable reflection of the opinion in the population? Uneducated guesses vary widely. Some people think they can ask a few of their friends and get a good idea, and some think a pollster would need to ask nearly everybody in the population. We will see that both of these extreme views are wrong.

If you read the fine print in the report of a professionally conducted public opinion poll, you'll probably find information about the number of people who were interviewed (n) and the "margin of sampling error" (E) of the poll. The relationship between n and E has a solid foundation in probability theory, based on the binomial and normal distributions. But here we will take an intuitive approach to understanding the precision of a poll—an approach based on some simple simulations with the statistical software R.

In order to illustrate how simulation works, we need to deal with a specific population proportion. So let's suppose 53% of voters would answer *yes* to our question, and the rest would answer *no*. That is, a little more than half of our population favors the idea of fining and expelling professional athletes who get caught using steroids and no one is undecided. We hope our poll will give an estimate near 53%—say within ± 2%—thus accurately revealing the majority viewpoint.

Bruce Trumbo (bruce.trumbo@csueastbay.edu) is Professor of Statistics and Mathematics at California State University, East Bay (formerly CSU Hayward). He is a Fellow of ASA and holder of the ASA Founder's Award.

Are 25 Subjects Enough?

Let's begin by thinking about polls with only 25 people. What are the chances a 25-subject poll will accurately reflect that more than half the population would answer *yes*? In R, we can simulate such a poll with one statement:

```
> sample(c(0,1), 25, rep=T, prob=c(.47, .53))
```

Here is how the `sample` function works. In its 'console' window, R supplies the prompt >, and the user types the rest. The four items in parentheses are called 'arguments.' The first argument of the sample function lists the possible responses, here 0 and 1. We take 0 to stand for *no* and 1 for *yes*. (The symbol `c` is used because lists in R usually are considered to be arranged in a column, even though they sometimes are printed out as a row to save space.) The second argument is the sample size, here 25. With the argument `rep`, we say whether repetition is allowed. Here, the answer is `T`, for true. The argument `prob` lists probabilities 0.47 and 0.53 that correspond to 0 (*no*) and 1 (*yes*), respectively.

The first time we tried it, R returned the result below, where the numbers in brackets give the item number of the first item in each row of output.

```
 [1] 1 0 1 1 0 0 1 0 0 1 0 0 0
[14] 0 1 0 1 0 1 1 1 0 1 1 0
```

We interpret this as a sequence of responses to the poll by 25 subjects, where 1 stands for *yes* and 0 for *no*.

```
     Y N Y Y N N Y N N Y N N N
     N Y N Y N Y Y Y N Y Y N
```

Of our 25 simulated responses, 12 were *yes*. This is a disappointing and misleading result: The proportion of *yes* answers in the sample is $p = 12/25 < 1/2$, while the population proportion of *yes* answers is 53% > 1/2.

However, this is just one way the poll might have turned out. To fairly judge the usefulness of 25-subject polls, we must look at many of them. And before we can do that, we need to consider how to do the looking.

The Trace of a Poll

In order to visualize what happened in a simulated poll, it helps to plot its *trace*. At each step or *trial* 1 through 25, we determine the proportion of successes so far. This is done by summing the number of *yes* answers so far and dividing by the number of subjects interviewed so far. Thus, the proportion after the first trial is 1/1 = 1 (one *yes* for one subject), and 1/2 = 0.5 (one *yes* for two subjects) after the second trial. You can verify easily that the next values are 0.33, 0.75, 0.60, and so on.

Summarizing a trace in this way is easy to automate in R with the following code:

```
> x <- sample(c(0,1),25,rep=T, prob=c(.47,.53))
> r.tot <- cumsum(x);  Trial <- 1:25
> Proportion <- r.tot/Trial
> Proportion[25]
> cbind(Trial, x, r.tot, Proportion)[1:7,]
> plot(Trial, Proportion)
```

Here, `x` is the list of `0`s and `1`s, `cumsum` makes a list of running totals (`1`s so far), and `1:25` is a list of the integers 1 through 25 (subjects so far). The third line prints the endpoint 0.48, the fourth binds the relevant columns together and prints summary results for the first seven trials (see Figure 1). The last line makes the plot of the entire trace shown in Figure 2 (except for the labels).

```
       Trial   x   r.tot  Proportion
  [1,]     1   1       1       1.000
  [2,]     2   0       1       0.500
  [3,]     3   1       2       0.667
  [4,]     4   1       3       0.750
  [5,]     5   0       3       0.600
  [6,]     6   0       3       0.500
  [7,]     7   1       4       0.571
```

Figure 1. Summary results for the first seven subjects

Because this is a simulation of a random process, the result will be different each time we execute the same commands in the R console window. In Figure 3, we illustrate the variability of such results by overlaying traces of five simulated polls. Here, it happens that three of the five traces end above 1/2, but two simulated polls have misleading endpoints below 1/2.

More comprehensively, in Figure 4, we show the histogram of the endpoints of 10,000 25-subject polls. A normal curve centered at 53% with a standard deviation of 10% fits the results fairly accurately. Specifically, you can see that on average the polls tend to give results slightly above 50%. However, 3,847 of these 10,000 polls (38.5%)

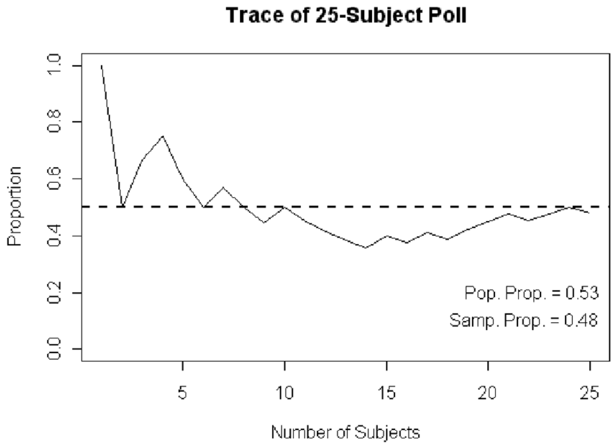


Figure 2. Trace of our first simulated 25-subject poll

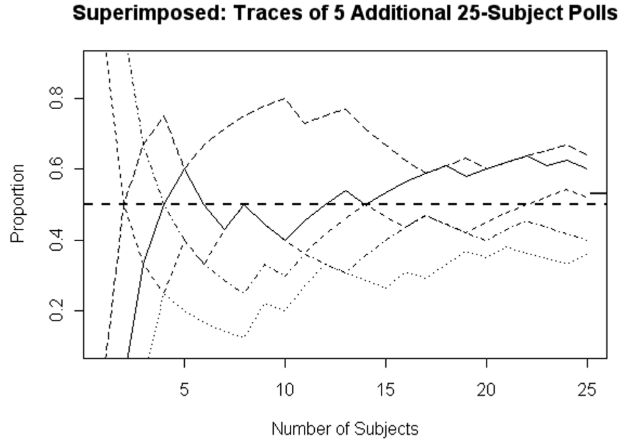


Figure 3. Each trace is plotted with a different style of line. Polls with 25 subjects can give widely varying results.

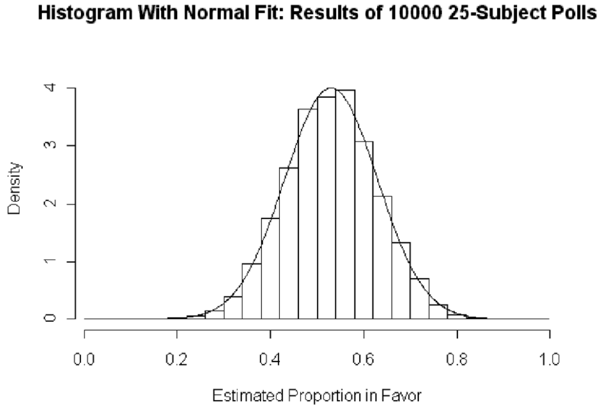


Figure 4. Sample proportions in 25-subject polls tend to lie in the interval 53% ± 20%

R and Binomial Distribution

If you know about binomial distributions, you can compute the exact proportion of 25-subject polls that give misleading tallies below 1/2. Suppose X has a binomial distribution with 25 trials and probability 0.53 of *yes* on any one trial, then P{X 12} = 0.3808. R does such computations easily. The required statement is `pbinom(12, 25, .53)`.

gave misleading results below 1/2. Any way you look at it, 25 subjects *clearly are not enough to get reliable results from a poll*. The results are just too variable to be useful.

Using 2,500 Subjects

If we simulate polls in which the number of subjects is n = 2500, instead of n = 25, we see the results are much more consistent. Figure 5 shows the traces of two simulated polls based on 2,500 subjects. In interpreting such random processes, we must try to distinguish the 'message' from the 'noise.'

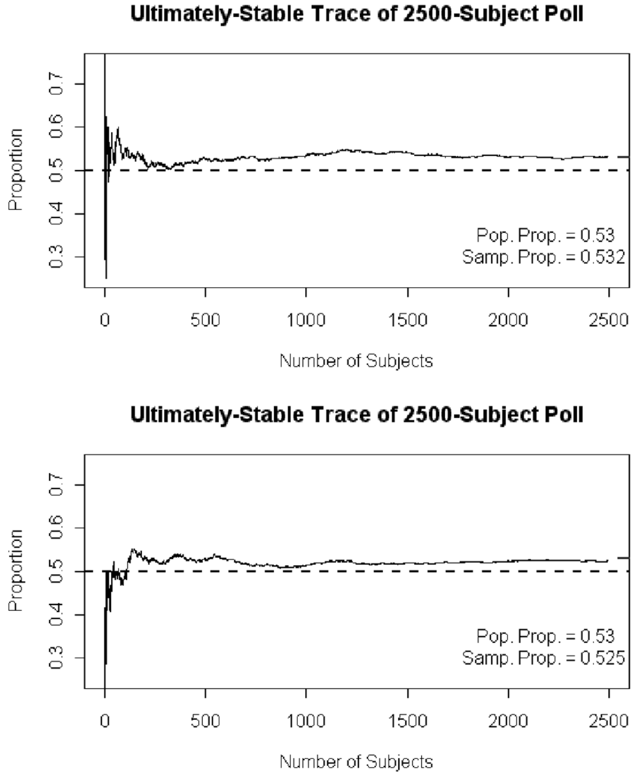


Figure 5. Traces of two 2,500-subject polls begin very differently, but have nearly equal endpoints

In Figure 6, we overlay the traces of 20 2,500-subject polls. The beginning (left side) of each trace can be quite erratic, about as likely to be much too high as to be much too low (the noise). But as a typical trace gets to larger numbers of subjects, it starts to become stable. And when it reaches its endpoint, it is very likely to be between 51% and 55%, and almost always above 1/2 (the message).

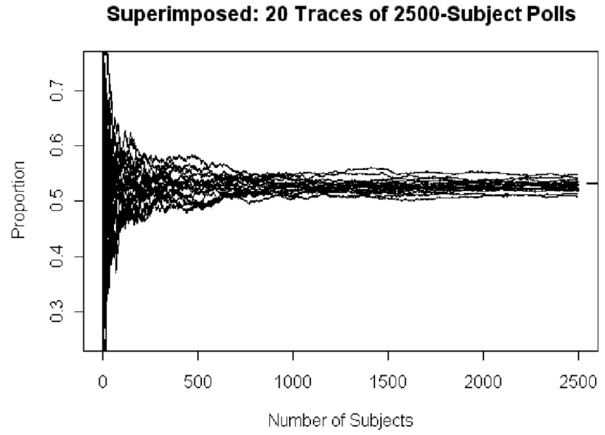


Figure 6. Twenty simulated polls illustrate the relatively small variability when 2,500 subjects are used

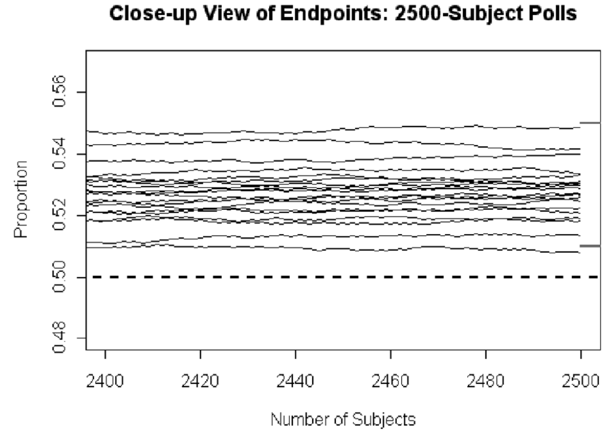


Figure 7. A close-up view of the right-hand side of Figure 6 shows 19 of the 20 simulated polls ending in the interval 53% ± 2% (tick marks) and all ending above 50% (dotted line)

Figure 7 shows a magnified view of the right end of Figure 6. In our particular run of 20 polls, 19 ended between 51% and 55%. This happens to be an 'average' result: 19 out of 20 is 95%. And one can show that the probability is 95% the trace of a 2,500-subject poll will end in the interval 53% ± 2%.

For sample sizes as large as 2,500 from a population with 53% in favor, the endpoint of a poll is very nearly the same as a random variable Y_{2500} that has a normal distribution with mean 0.53 and standard deviation 0.01. So P{0.51 < Y_{2500} < 0.55} = 95% and P{Y_{2500} < 1/2} = 0.00135. Figure 8 shows a histogram of 10,000 2,500-subject polls. Of our 10,000 simulated polls, only eight had endpoints below 1/2, which happens to be slightly fewer than the expected 13.5. Different simulation runs of 10,000 polls would give slightly different results. But they consistently show that about 95% of the endpoints are in the interval 53% ± 2%.

You may be thinking the histograms in Figures 4 and 8 look a lot alike. They are alike in terms of their approximately normal shape, and it is useful to know that polling distributions are often nearly normal. But the

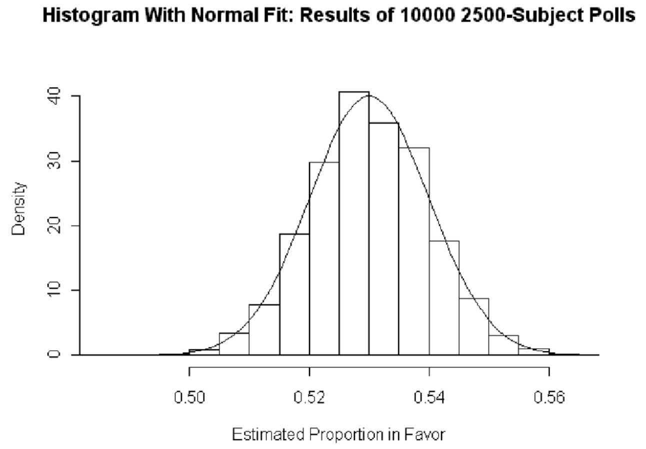


Figure 8. Sample proportions in 2,500-subject polls tend to lie in the interval 53% ± 2%

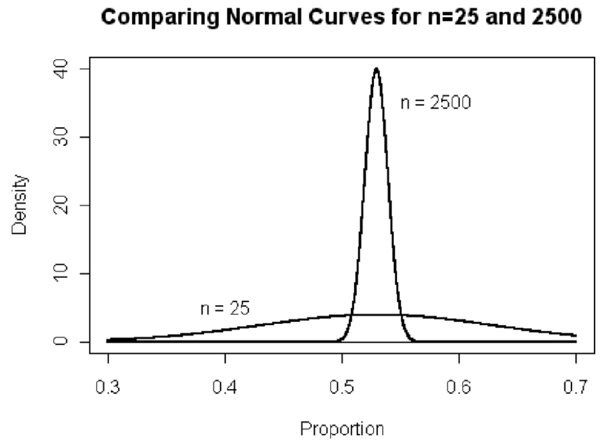


Figure 9. When plotted on the same scale, the normal curves approximating the histograms in Figures 4 and 8 show the much smaller variability of 2,500-subject polls

crucial point here is that Figures 4 and 8 are plotted on *very different scales*. In Figure 9, we use the same scale to plot the normal curves that approximate these two histograms. The tall, peaked curve shows the distribution when the sample size is 2,500. From these results, it is clear that a properly conducted poll with 2,500 subjects would be big enough to accurately answer our question about voters' opinions about a law against using steroids in professional sports.

Confidence Intervals

We have seen that if we repeatedly poll an imaginary population where we know 53% of the subjects will answer *yes*, then 95% of the polls get answers within 2% of the correct value. Now, taking a more practical point of view, suppose we sample from a population where the proportion in favor is unknown and we see 56% in favor in our sample. Then, we can be 95% sure that our poll result will be within 2% of the unknown population proportion. We conclude that the population proportion must lie in the interval 56% ± 2%.

Most professional polling organizations report such margins of sampling error along with their published results. For polls with population proportions in the range 35% to 65% and based on a simple random sample of n > 500 subjects, the margin of error is roughly 1/÷n. This rule implies that the margin of error for a poll with n = 1,100 is about ±3%, and for n = 2,500, the margin of error is about ±2%.

These margins of error are supposed to be valid for 95% of polls. However, they only allow for *sampling* error, which can be modeled by probability rules or can be simulated as we have done here. A conclusive analysis of such factors as nonrandom sampling, nonresponse, misunderstood questions, and dishonest answers is largely beyond the reach of probability or simulation. These factors may lead to an unknowable percentage of results that fall outside the claimed margin of error.

Figure 10 shows 95% confidence intervals for the simulation of 20 polls with 2,500 subjects that we illustrated in Figures 5 and 6. The confidence intervals are based on the standard formula $p \pm 1.96[p(1 - p)/n]^{1/2}$, where p denotes the sample proportion. When p is in the vicinity of 50%, the margin of error is about ± 2%. (For small n, it is best to "add two successes and two failures" to the counts before using this formula, but for n = 2,500, no adjustment is needed.)

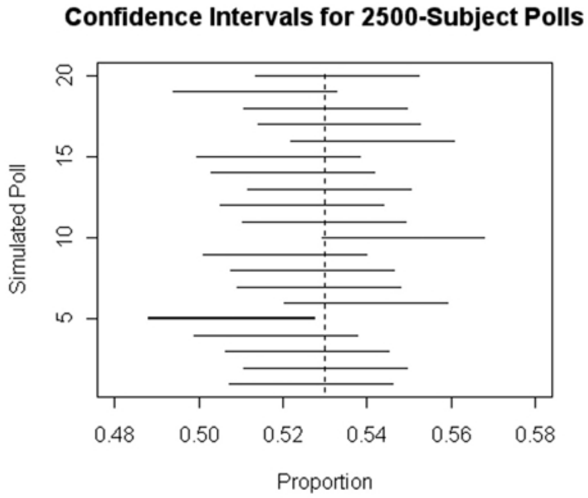


Figure 10. In this run of 20 simulated 2,500-subject polls, one out of 20 yielded a confidence interval that did not cover the population proportion in favor. The margin of sampling error for the confidence intervals is about 2%.

Exploring on Your Own

Simulation should not be entirely a spectator sport! We hope you will try the simulations in this column on your own. You can get the full R code and other related materials on the web at *www.sci.csueastbay.edu/~btrumbo/STATSSIM* or *www.amstat.org/publications/STATS/data.html*.

Here are three challenges about polling for you to try. Visit either of the web sites for clues to help you tackle these challenges. Let us know what you discover.

1. Elementary. If the population proportion in favor is very near 0 or 1, the "reciprocal root n" rule doesn't apply and the standard formula for confidence intervals may not be valid. To look at such a case, suppose the polling question called for a 10-year prison term for any professional athlete caught using steroids. Then, the population proportion of *yes* answers might be very low. On the assumption that 10% is in favor, make figures similar to some of those in this column to illustrate the results of simulations with n = 25 and 1,600. Does the normal distribution provide a good fit when n = 25? When n = 1,600? Illustrate the margin of error for a poll with sample size 1,600. You can do this by making a few minor changes in the code provided online.

2. Intermediate. Of course, it is unrealistic to assume everyone in the population has an opinion about a law against steroid use in professional sports. Suppose the population has 53% with opinion *yes* (represented as 1), 42% *no* (–1), and 5% *undecided* (0). Do a simulation with n = 2,500 to illustrate the margin of error for the lead of *yes* over *no*—that is, the proportion of *yes* responses minus the proportion of *no* responses. Give an intuitive explanation why it's a lot larger than 2%.

3. Advanced. Many people suppose that if the population proportion of *yes* answers is, say, 50%, the sample proportion will fluctuate continually above and below the 50% point as the poll progresses. But one can show that traces of such polls tend to be on one side of their 50% target value, eventually approaching the target without exactly touching it very often.

The average number of times the trace of a 2,500-subject poll touches 50% is only about $[2n/\pi]^{1/2}$ 40, and the most likely number is 0. Suggestions on how to explore the approximately 'half normal' distribution of the number of 'visits' to 50% (and related results) are shown on the web sites.

Acknowledgment

Thanks to my colleague Eric A. Suess at California State University, East Bay (formerly CSU Hayward), and our student Shuhei Okumura, a BS major in statistics and political science, for help with this column and the associated web materials. The three of us contributed a paper on closely related topics for JSM 2005 in Minneapolis (Section on Statistics Education, poster presentation by Okumura). ■

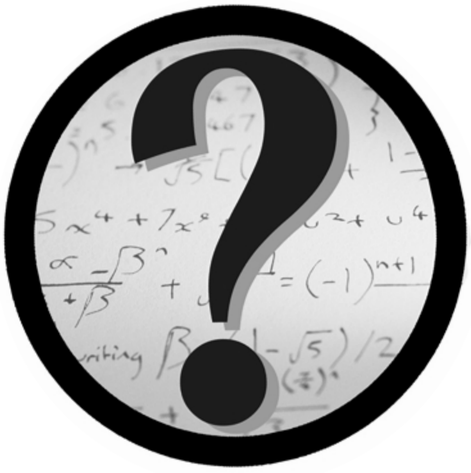
THE MAGAZINE FOR STUDENTS OF STATISTICS
WINTER 2005 • ISSUE # 42
SPRING 2005 • ISSUE # 43
FALL/WINTER 2005 • ISSUE # 44
STATS
TUESDAY
WEDNESDAY
THURSDAY
FRIDAY
STATISTICS in sports
Does Batting Average Measure Ability or Luck?
An NFL Cookbook: Qualitative Recipes for Winning
Is Lance Armstrong the Greatest Cyclist Ever?

CALL FOR PAPERS

STATS: The Magazine for Students of Statistics is interested in publishing articles that illustrate the many uses of statistics to enhance our understanding of the world around us. We are looking for engaging topics that inform, enlighten, and motivate readers, such as:

- statistics in everything from sports to medicine to engineering
- "statistics in the news," discussing current events that involve statistics and statistical analyses
- statistics on the internet, covering new web sites with statistical resources, such as datasets, programs, and examples
- interviews with practicing statisticians working on intriguing and fascinating problems
- famous statisticians in history and the classic problems they studied
- the "statistics almanac" that tells us what happened during this month in statistics history
- using particular probability distributions in statistical analyses
- examinations of surprising events and questions, such as "What are the chances?"
- reviews of books about statistics that are not textbooks
- student projects using statistics to answer interesting research questions in creative ways

So think of some great ideas and send a description of your concepts for feature articles to Paul J. Fields, editor, *pjfields@byu.edu*.



ASK STATS

Jackie Miller

Hey Coach...

Question 1: I've often heard the cliché "Defense wins championships." Is there any statistical evidence to back up this claim?

The implicit assumption in this claim is that defense is more important than offense in achieving the ultimate goal of winning a championship. The strong defense theory often is used to counterbalance the typical fan's attraction to the flashy offensive aspects of scoring in sports such as baseball, football, basketball, and hockey. A team that trades an offensive star for a defensive player or adopts a more conservative offensive philosophy often is criticized by its fans. The standard response is that having a strong defense is the key to winning at playoff time when championships are decided. Given the multitude of sports data available from past seasons, it should be relatively straightforward to determine whether defensive ability tends to be a better indicator of championship potential than offensive ability.

To investigate this question, consider the National Basketball Association (NBA). Since the 1970–71 season, the NBA has crowned 34 champions, including the Los Angeles Lakers nine times, the Chicago Bulls six times, and the Boston Celtics five times. Let us look at each of the 34 championship teams and see how each team's offense and defense stacked up against the other teams' in the league that year. We'll use average points scored and average points allowed per game during the regular season as the measures of offensive and defensive ability. Because changes in the rules and styles of play have affected scoring rates over the years, we'll rank each team versus the other teams in the league for each year. We'll give the rank of one to the team scoring the most points (offense) and the rank of one to the team allowing the fewest points (defense). These ranks are summarized in Table 1 (page 26).

How do the offensive and defensive ranks compare for the teams that won championships? In 16 of the 34 years, the eventual champion's defense was ranked better than its offense. But in 16 years, it went the other way, and in two years, the rankings were tied. So no preference for the defense is evident in that comparison.

What if we figure in the magnitude of the ranks? If we subtract the defensive rank from the offensive rank (so a positive difference means a stronger defense), the average difference is 0.71, which is in the right direction to support the strong defense theory, but not at all significantly different from zero.

Do the rankings of the opponent in the NBA finals matter? Table 2 (page 26) compares the head-to-head offensive and defensive ranks of the participants in the NBA's best-of-seven championship series.

The team with a better defensive ranking has won the championship 20 times in 34 years, while the better offensive team has won 21 times. Note that a team has come into the final series with the better offensive and the better defensive ranking just seven times, but has won the championship every time. There has never been a true 'underdog' with both the worse offensive and the worse defensive ranking that has prevailed in the final series. So it looks like a combination of high scoring offense and stingy defense is the surest way to win in the NBA.

Want to try this for a different sport? A good source for data in professional football (NFL), baseball (MLB), basketball (NBA), and hockey (NHL) can be found at *www.bballsports.com*, which has a web interface to a database of historical data in each of these sports.

Jackie Miller (miller.203@osu.edu) is a Statistics Education Specialist and auxiliary faculty member in the Department of Statistics at The Ohio State University. She earned both a BA and BS in mathematics and statistics at Miami University, along with an MS in statistics and a PhD in statistics education from The Ohio State University. She is very involved in the statistics education community. When not at school, Miller enjoys a regular life (despite what her students might think), including keeping up with her many dogs!

Question 2: What about a similar claim in golf that "you drive for show, but putt for dough"?

This theory has a similar origin in claiming that the least flashy aspect of golf (rolling putts on a green) is more important to scoring well than the impressive full swings and booming 300+ yard tee shots. For a nice statistical analysis of this question, see Scott Berry's A Statistician Reads the Sports Pages column titled "Drive for Show and Putt for Dough" in *CHANCE* magazine (Vol. 12, No. 4, 1999). Past issues of *CHANCE*, published jointly by the American Statistical Association (ASA) and Springer, are a particularly good source for readable articles of interest to the statistically-minded sports fan (or the sports-minded statistics fan).

Year	Off	Def	Champion	Opponent	Off	Def
2004	24	1	Detroit Pistons	Los Angeles Lakers	3	14
2003	12	3	San Antonio Spurs	New Jersey Nets	14	2
2002	3	10	Los Angeles Lakers	New Jersey Nets	13	5
2001	3	23	Los Angeles Lakers	Philadelphia 76ers	15	5
2000	6	5	Los Angeles Lakers	Indiana Pacers	4	11
1999	13	3	San Antonio Spurs	New York Knicks	27	4
1998	9	3	Chicago Bulls	Utah Jazz	3	12
1997	1	6	Chicago Bulls	Utah Jazz	2	8
1996	1	2	Chicago Bulls	Seattle Supersonics	3	8
1995	8	14	Houston Rockets	Orlando Magic	1	19
1994	13	5	Houston Rockets	New York Knicks	21	1
1993	14	2	Chicago Bulls	Phoenix Suns	1	18
1992	5	3	Chicago Bulls	Portland Trail Blazers	4	12
1991	7	4	Chicago Bulls	Los Angeles Lakers	13	2
1990	19	1	Detroit Pistons	Portland Trail Blazers	4	19
1989	16	2	Detroit Pistons	Los Angeles Lakers	5	8
1988	5	11	Los Angeles Lakers	Detroit Pistons	8	3
1987	2	12	Los Angeles Lakers	Boston Celtics	6	4
1986	7	5	Boston Celtics	Houston Rockets	6	13
1985	2	14	Los Angeles Lakers	Boston Celtics	5	5
1984	7	5	Boston Celtics	Los Angeles Lakers	4	16
1983	7	7	Philadelphia 76ers	Los Angeles Lakers	2	12
1982	2	15	Los Angeles Lakers	Philadelphia 76ers	10	5
1981	8	2	Boston Celtics	Houston Rockets	11	14
1980	2	11	Los Angeles Lakers	Philadelphia 76ers	5	10
1979	19	1	Seattle Supersonics	Washington Bullets	3	9
1978	7	12	Washington Bullets	Seattle Supesonics	18	2
1977	3	9	Portland Trail Blazers	Philadelphia 76ers	5	10
1976	7	7	Boston Celtics	Phoenix Suns	9	8
1975	1	14	Golden State	Washington Bullets	5	2
1974	4	7	Boston Celtics	Milwaukee Bucks	7	3
1973	11	1	New York Knicks	Los Angeles Lakers	4	5
1972	1	6	Los Angeles Lakers	New York Knicks	14	3
1971	1	3	Milwaukee Bucks	Baltimore Bullets	11	9

Table 1. Offensive and defensive ranks for the NBA finalists from 1971 though 2004

		Better Offense		
		Wins	Losses	Total
Better Defense	Wins	7	13	20
	Losses	14	0	14
	Total	21	13	

Table 2. Comparison of offensive and defensive ranks of NBA finalists from 1971 through 2004

Question 3: I know earned run average (ERA) is a common measure of the effectiveness of a baseball pitcher. I've developed my own statistic for measuring pitching effectiveness. How can I publicize my statistic and get others to adopt it or give me feedback on improving it?

There are two good avenues. Within the ASA, there is a Section on Statistics in Sports (SIS) that sponsors several sessions at the Joint Statistical Meetings each summer devoted to applications of statistics in sports. SIS also maintains an email list for sports-related queries and discussions in addition to a web site (*www.amstat.org/sections/sis*). If you are interested in sports, you definitely should join SIS.

For baseball statistics in particular, there is also the Society for the Advancement of Baseball Research (SABR). They produce various publications, sponsor conferences, host an e-discussion list, and maintain a web site (*www.sabr.org*) related to the field now known as "sabermetrics."

Question 4: Are there career opportunities as a sports statistician?

Yes, there are positions within sports teams, leagues, and the sports media that involve compiling, distributing, and analyzing statistics. Many of these are part-time positions, but some can turn into a full-time career. Experienced sports statisticians recommend volunteering your time with a local team as a way to get started and gain experience. Additional advice can be found at the ASA's Section on Statistics in Sports web site, *www.amstat.org/sections/sis/career/index.html*.

Robin Lock

Many thanks to Robin Lock for the answers to our Ask STATS *questions. Lock is the Burry Professor of Statistics at St. Lawrence University in Canton, New York, and is currently the chair of the Section on Statistical Education of the American Statistical Association. He also has been involved with the Section on Statistics in Sports. As you can tell from his picture, Lock is a hockey fan—but his sports interests are vast.*

Remember, this is a forum in which you may ask any question you have about statistics. To have your question answered, please email it to Jackie Miller at miller.203@osu.edu. *In the subject line, please put "Ask* STATS.*" In the body of the email, ask your question and tell us your school name and where it is located. Also, please let us know if we can use your name in the column. If we choose your question for publication, you will receive an ASA T-shirt!* ■

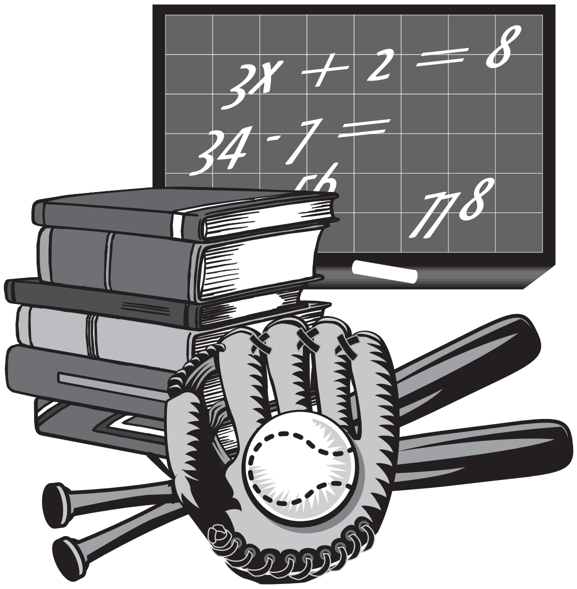
What Is a Fair Comparison?

Lessons from Baseball for Public Schools

Chris Olsen

Let me begin with some background to bring us all up to date. In the early 1980s, parents became increasingly concerned about the performance of their children's schools. Subsequently, there has been a dramatic increase in testing required by state law and federal legislation in the form of the "No Child Left Behind" Act. The provisions of these various laws generally require reporting of results so the public can be 'informed' about the performance of their schools.

The response by the lion's share of educators has been fairly predictable. In their view, reporting and comparing schools based on student achievement is inherently unfair because of social and economic differences and circumstances totally beyond the teachers' and administrators' control. One might as well attempt to rank the best batters in baseball history, even though they played in different parks, with different sets of teammates, in different eras! Crazy is what that would be! Who would attempt such an undergoing?



Well, as it turns out, Michael J. Schell, professor of biostatistics at The University of North Carolina at Chapel Hill attempted such an undergoing in *Baseball's All-time Best Sluggers: Adjusted Batting Performance from Strikeouts to Home Runs (BATS)*.

As an on-again, off-again follower of the fortunes—well OK, famines—of the Chicago Cubs, I cannot claim to be knowledgeable about our nation's pastime. In any baseball discussion, I am the natural prey of anyone whose command of baseball fact and lore rises above roughly zero. I didn't even know until recently that Tinker, Evers, and Chance were not the complete Cubs infield in their day. Fortunately for me, Schell uses sensible statistical analyses, artfully explained, and I only need to know a small amount about baseball to follow his arguments. Also fortunately for me, the statistical analyses are presented clearly with many graphic and elementary algebraic assists.

It cannot be claimed that all the statistical analyses in *BATS* are justified fully internally—one would have hoped for a bit more bibliography for the statistically semi-literate—but each statistical technique is presented clearly in the text and outlined in more detail in the appendixes. A year, or even a semester, of background in elementary statistics should suffice for appreciation of the statistical methodology. In fact, one suspects a high school student with a year of AP statistics under his or her belt could follow the exposition relatively easily! I will present a bit of detail about Schell's methods so the potential reader of *BATS* may gauge the statistical variety and level of the methods.

One immediately notices the kind of rational approach one would expect from a statistician. He identifies and considers a set of 'offensive' categories related to batting (e.g., triples, doubles, RBIs, etc.) and ranks each player according to a seven-step method.

Step 1: Establish guidelines for deciding which players are eligible for the single-season and career lists. Statistical aficionados everywhere will recognize this as "defining the population."

Chris Olsen (colsen@cr.k12.ia.us) teaches mathematics and statistics at George Washington High School in Cedar Rapids, Iowa. He has been teaching statistics in high school for 25 years and has taught AP statistics since its inception.

Step 2: Calculate the "mean-adjusted averages" using standardizing averages. (In this step, Schell establishes the prevalence of an offensive event relative to the particular season, thus 'standardizing' across seasons.)

Step 3: Obtain good estimates of the park effects for all offensive events and calculate "park-adjusted averages" for all players. What? You mean all ball parks aren't created equal? The folks at Coors Field (hitter heaven) and Fenway Park (Jolly Green Giant) surely will be surprised at this piece of news! The park effect apparently is a significant item, and Schell uses some time-series methods on this one—would you believe a "Multiple Changepoint Regression with Backward Elimination"? I'm not sure, but I think Casey Stengel used that sort of method when he managed the New York Mets.

Steps 4 & 5: For each season, assess the performance spread, obtain means and standard deviations from transformed distributions, and stabilize them using five-year moving averages. As we all know, sometimes those distributions are peskily non-normal. And not only that, they vary from year to year. Here, power transformations and what appears to be something like z-scores are utilized.

Step 6: Calculate the fully adjusted average. After all those adjustments for years, parks, teammates, etc., we finally get an old-fashioned z-score!

Step 7: Adjust for late career declines.

So there you have it, as Schell says, his method "in a nutshell."

Now we get to a sticky issue: Who should read *BATS*? I'm not sure most baseball fans should pick this up, especially as Tinker, Evers, and Chance are ranked only 45th, 46th, and 53rd in their respective positions. However, I am sure students of statistics would learn a great deal about applying fairly elementary statistics with pleasure derived from reading an interesting and illuminating presentation.

Finally, I suggest those on any side of the "No Child Left Behind" discussion should most certainly read *BATS*. If the public is to be well-informed about its public schools at a level equal to Schell's remarkable presentation about baseball's best sluggers, we should be honest about how to make comparisons among schools fairly and about how to inform the public. Were that to happen, Schell's most important contribution would not be to baseball or to statistics, but to the public good!

Reference

Schell, Michael J. 2005. *Baseball's All-time Best Sluggers: Adjusted Batting Performance from Strikeouts to Home Runs.* Princeton University Press, Princeton, New Jersey. ■

Statistical Computing and Statistical Graphics Sections, American Statistical Association

Student Paper Competition 2006

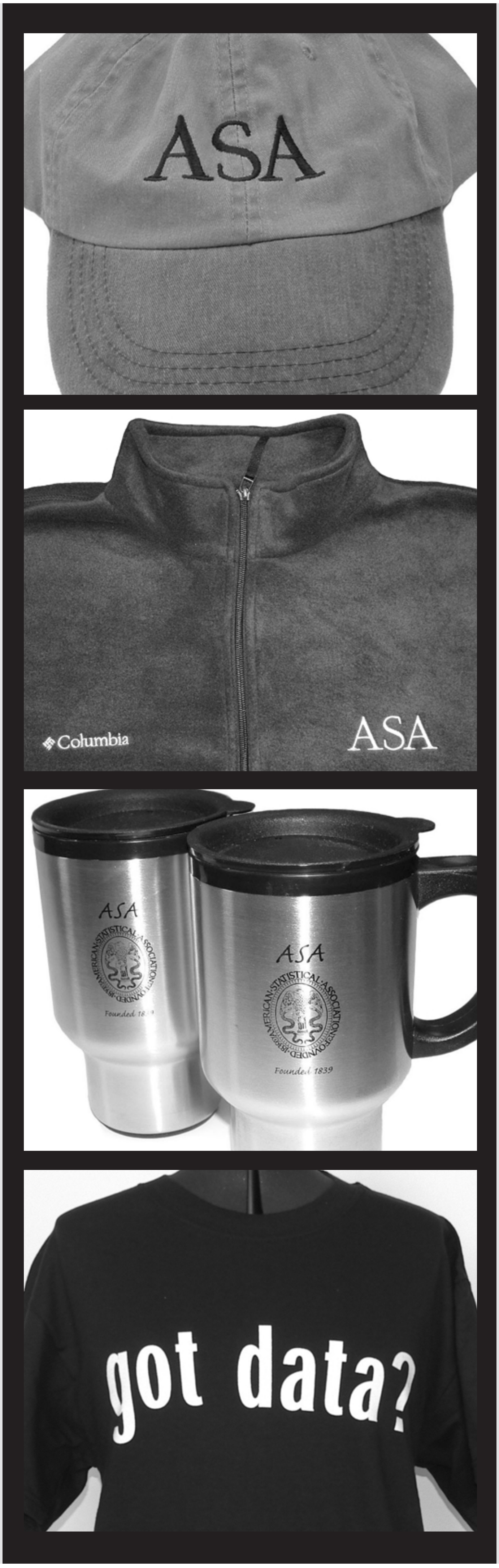
The Statistical Computing and Statistical Graphics Sections of the ASA are cosponsoring a student paper competition on the topics of Statistical Computing and Statistical Graphics. Students are encouraged to submit a paper in one of these areas, which might be original methodological research, some novel computing or graphical application in statistics, or any other suitable contribution (for example, a software-related project). The selected winners will present their papers in a topic-contributed session at the 2006 Joint Statistical Meetings. The Sections will pay registration fees for the winners as well as a substantial allowance for transportation to the meetings and lodging (which in most cases covers these expenses completely).

Anyone who is a student (graduate or undergraduate) on or after September 1, 2005, is eligible to participate. An entry must include an abstract, a six-page manuscript (including figures, tables, and references), a CV, and a letter from a faculty member familiar with the student's work. The applicant must be the first author of the paper. The faculty letter must include a verification of the applicant's student status and, in the case of joint authorship, should indicate what fraction of the contribution is attributable to the applicant. It is preferred that electronic submissions of papers be in Postscript or PDF. All materials must be in English.

All application materials MUST BE RECEIVED by 5 p.m. EST, Monday, December 19, 2005, at the address below. They will be reviewed by the Student Paper Competition Award Committee of the Statistical Computing and Graphics Sections. The selection criteria used by the Committee will include innovation and significance of the contribution. Award announcements will be made in late January 2006.

Additional important information on the competition can be accessed on the web site of the Statistical Computing Section, *www.statcomputing.org*. A current pointer to the web site is available at *www.amstat.org*. Inquiries and application materials should be emailed or mailed to:

Student Paper Competition
c/o Dr. José Pinheiro
Biostatistics, Novartis Pharmaceuticals
One Health Plaza, Room 419/2115
East Hanover, NJ 07936
jose.pinheiro@novartis.com ■



SHOP ASA

The ASA's new online marketplace is now open at ***www.amstat.org/asastore.***

Purchase T-shirts, books, ASA Proceedings, and gift items—such as the ASA statistical words magnet!

SAVE 10%

on your first purchase by entering **ASASTORE** at checkout.